VC Dimension : یک بررسی جامع

فاطمه بناءهمزایی

گروه مهندسی کامپیوتر ،مهندسی هوش مصنوعی و رباتیک ، دانشگاه آزاد اسلامی ، تهران ، ایران

از زمان تکامل بشریت، انسان ها از انواع مختلفی از ابزارها برای انجام وظایف مختلف به روشی ساده تر استفاده کرده اند. خلاقیت مغز انسان منجر به اختراع ماشین های مختلف شد. این ماشین‌ها با قادر ساختن افراد برای برآوردن نیازهای مختلف زندگی از جمله سفر، صنایع و محاسبات، زندگی انسان را آسان کردند. و یادگیری ماشینی یکی از آنهاست. به گفته آرتور ساموئل، یادگیری ماشینی به عنوان رشته‌ای تعریف می‌شود که به رایانه‌ها توانایی یادگیری بدون برنامه‌ریزی صریح را می دهند و به هشت روش یادگیری کلی تقسیم می شود: یادگیری تحت نظارت شده، یادگیری نظارت نشده، یادگیری نیمه نظارت، یادگیری تقویتی، یادگیری چند وظیفه ای، یادگیری گروهی، شبکه عصبی و یادگیری مبتنی بر نمونه. هیچ نوع الگوریتم واحدی وجود ندارد که برای حل یک مسئله بهترین باشد. نوع الگوریتم به کار گرفته شده به نوع مسئله ای که می خواهید حل کنید، تعداد متغیرها، نوع مدلی که برای آن مناسب است و غیره بستگی دارد.

(VC Dimension) یکی از مفاهیم بنیادین در نظریه یادگیری ماشین است که برای سنجش پیچیدگی و توانایی تعمیم مدل‌های یادگیری استفاده می‌شود. اگرچه به طور مستقیم یک الگوریتم یا مدل یادگیری ماشین نیست، اما یک ابزار بسیار مهم برای تحلیل و درک عملکرد مدل‌های یادگیری ماشین محسوب می‌شود.

این مفهوم، ارتباط مستقیمی با احتمال بیش‌برازش (Overfitting) مدل دارد و به ما کمک می‌کند تا مدل‌هایی را انتخاب کنیم که نه تنها داده‌های آموزشی را به خوبی یاد می‌گیرند، بلکه توانایی پیش‌بینی دقیق داده‌های جدید را نیز داشته باشند.

بنابراین مطالعه خود را در زمینه VC Dimension انجام دادیم و مثال های تئوری و عددی آن ها را تحلیل کردیم. و VC Dimension را از جنبه های مختلف از جمله اهمیت و چالش ها و کاربرد های کلی مورد بررسی قرار دادیم.علاوه بر این ارتباط آن را با Overfitting تحلیل و بررسی نموده. و همچنین بررسی نامساوی VC و محاسبات آن را شرح دادیم.سپس استفاده از VC Dimension در انتخاب مدل ها را عنوان کردیم.سرانجام VC Dimension را با مثال هایی در دنیای واقعی ارزیابی کردیم. و همچنین مشاهده کردیم علم یادگیری ماشین پاسخ گو بسیاری از چالش های مطرح شده می باشد. در این راستا مقالات مروری و مقالات تحقیقاتی را از پایگاه های داده معتبر علمی دنیا دریافت و بطور کامل بررسی و تحلیل کردیم. تا بتواند مورد استفاده دقیق و کاربردی محققان این عرصه قرار گیرد.

کلمات کلیدی: مدل های یادگیری ماشین، نظریه یادگیری، عمق تعمیم، تعمیم پذیری، تحلیل پیچیدگی، پیش برازش، مدل های پیچیده، VC Dimension ، سرعت یادگیری.

۱- مقدمه:

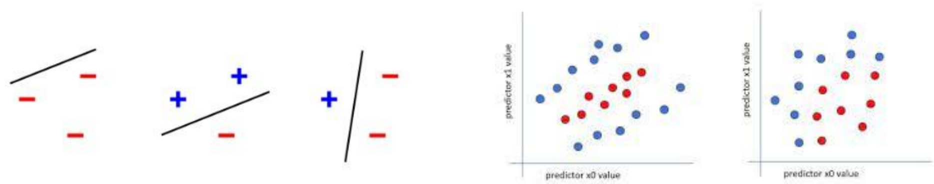
VC Dimension یک مفهوم کلیدی در یادگیری ماشین است که به ما کمک می‌کند تا پیچیدگی یک مدل یادگیری را درک کنیم. به زبان ساده، VC Dimensionنشان می‌دهد که یک مدل تا چه اندازه می‌تواند داده‌های مختلف را از هم تفکیک کند و الگوهای پیچیده را یاد بگیرد.

۱-۱ مثال: تقسیم نقاط با یک خط:

تصور کنید مجموعه‌ای از نقاط در یک صفحه داریم. می‌خواهیم با کشیدن یک خط، این نقاط را به دو گروه تقسیم کنیم.

اگر نقاط به صورت تصادفی پخش شده باشند، یک خط ساده می‌تواند به راحتی آن‌ها را جدا کند. در این حالت، مدل ما (یعنی خط) پیچیدگی کمی دارد و VC Dimensionآن پایین است.

اگر نقاط به شکل پیچیده‌تری قرار گرفته باشند، ممکن است به یک خط منحنی یا حتی چندین خط نیاز داشته باشیم تا آن‌ها را به طور کامل از هم جدا کنیم. در این حالت، مدل ما پیچیده‌تر است و VC Dimension آن بالاتر است.



۱-۲ مثال:

تصور کنید می‌خواهید یک ربات را آموزش دهید تا اشیاء مختلف را در یک محیط شلوغ تشخیص دهد. این ربات باید بتواند بین هزاران شیء مختلف، مانند انواع مختلف حیوانات، وسایل خانگی، وسایل نقلیه و... تمایز قائل شود.

VC Dimension پایین: اگر این ربات فقط بتواند بین دو دسته کلی مانند "حیوان" و "غیر حیوان" تمایز قائل شود، مدل آن ساده‌تر است و VC Dimension پایین‌تری دارد. این مدل ممکن است در تشخیص اشیاء جدید دچار مشکل شود.

VC Dimension بالا: اگر این ربات بتواند بین انواع مختلف نژادهای سگ، مدل‌های مختلف خودرو و حتی اجسام بسیار ریز و جزئی تفاوت قائل شود، مدل آن بسیار پیچیده‌تر است VC Dimension بالاتری دارد. این مدل ممکن است در داده‌های آموزشی بسیار خوب عمل کند، اما در مواجهه با داده‌های جدید که در آموزش ندیده است، دچار بیش‌برازش شود و اشتباهات زیادی مرتکب شود.

۱-۳ مثال، پیش‌بینی قیمت خانه:

VC Dimension پایین: اگر برای پیش‌بینی قیمت خانه فقط از متغیرهایی مانند متراژ و تعداد اتاق استفاده کنیم، مدل ما ساده‌تر است و VC Dimension پایین‌تری دارد. این مدل ممکن است نتواند عوامل پیچیده‌تری مانند موقعیت جغرافیایی، نوع ساختمان و امکانات اطراف را در نظر بگیرد.

VC Dimension بالا: اگر برای پیش‌بینی قیمت خانه از تعداد بسیار زیادی متغیر استفاده کنیم، مانند نوع کفپوش، جنس پنجره‌ها، نوع سیستم گرمایشی و...، مدل ما پیچیده‌تر است و VC Dimension بالاتری دارد. این مدل ممکن است در داده‌های آموزشی بسیار دقیق عمل کند، اما در پیش‌بینی قیمت خانه‌هایی که با ویژگی‌های خاص و نادر هستند، دچار مشکل شود.

۲- اهمیت VC Dimension:

۲-۱ چرا ابعاد VCمهم است؟

جلوگیری از بیش‌برازش (Overfitting): مدل‌هایی با VC Dimension بالا، توانایی یادگیری الگوهای بسیار پیچیده، حتی نویز موجود در داده‌ها را دارند. این امر منجر به بیش‌برازش می‌شود و مدل در برابر داده‌های جدید عملکرد ضعیفی خواهد داشت. VC Dimension به ما کمک می‌کند تا پیچیدگی مدل را کنترل کرده و از بیش‌برازش جلوگیری کنیم.

انتخاب مدل مناسب: با دانستن VC Dimension ، می‌توانیم بین مدل‌های ساده و پیچیده انتخاب کنیم. مدل‌های ساده‌تر با VC Dimension کمتر، احتمال بیش‌برازش کمتری دارند اما ممکن است نتوانند الگوهای پیچیده را به خوبی یاد بگیرند. از سوی دیگر، مدل‌های پیچیده‌تر با VC Dimension بالاتر، توانایی یادگیری الگوهای پیچیده را دارند اما احتمال بیش‌برازش آن‌ها بیشتر است.

درک تعمیم‌پذیری: VC Dimensionبه ما کمک می‌کند تا درک بهتری از توانایی تعمیم مدل داشته باشیم. مدل‌هایی با VC Dimension پایین‌تر، معمولاً توانایی تعمیم بهتری دارند زیرا کمتر به جزئیات خاص داده‌های آموزشی وابسته هستند.

۲-۲ مثال: تشخیص دستخط

تصور کنید می‌خواهیم یک مدل یادگیری ماشین طراحی کنیم که بتواند دست‌خط افراد مختلف را تشخیص دهد.

VC Dimension پایین: اگر مدل ما فقط بتواند بین دو نوع دست‌خط (مثلاً دست‌خط کودک و بزرگسال) تفاوت قائل شود، مدل ما ساده‌تر است و VC Dimension پایین‌تری دارد. این مدل ممکن است در تشخیص جزئیات ظریف دست‌خط افراد مختلف ناتوان باشد.

VC Dimension بالا: اگر مدل ما بتواند بین دست‌خط‌های افراد مختلف با جزئیات بسیار زیاد (مثلاً تفاوت در شیب خطوط، اندازه حروف، نوع قلم و...) تفاوت قائل شود، مدل ما پیچیده‌تر است و VC Dimension بالاتری دارد. این مدل ممکن است در تشخیص دست‌خط‌های جدید که در داده‌های آموزشی وجود نداشته است، دچار مشکل شود.

۲-۳ مثال:

فرض کنید داریم یک مدل برای تشخیص اعداد دست‌نوشته طراحی می‌کنیم. هر عدد می‌تواند یکی از ۱۰ رقم ۰ تا ۹ باشد.

VC Dimension پایین: اگر مدل ما فقط بتواند بین اعداد زوج و فرد تفاوت قائل شود، VC Dimension آن بسیار پایین است.

VC Dimension بالا: اگر مدل ما بتواند بین تمام اعداد ۰ تا ۹ و حتی انواع مختلف نوشتن هر عدد (مثلاً اعداد با خط شکسته یا مدور) تفاوت قائل شود، ابعاد VCآن بسیار بالا است.

چرا این مثال‌ها مهم هستند؟

تشخیص بیش‌برازش: اگر مدل ما بتواند تمام نمونه‌های آموزشی را به درستی طبقه‌بندی کند، اما در تشخیص نمونه‌های جدید دچار مشکل شود، احتمالاً دچار بیش‌برازش شده است. این نشان می‌دهد که مدل ما پیچیدگی بیش از حدی دارد (VC Dimension بالا).

انتخاب مدل مناسب: برای هر مسئله، باید مدلی با VC Dimension مناسب انتخاب کنیم. اگر مسئله ما ساده است، مدلی با VC Dimension پایین کافی است. اما اگر مسئله پیچیده است، ممکن است به مدلی با VC Dimension بالاتر نیاز داشته باشیم.

تعمیم‌پذیری: مدلی با VC Dimension مناسب، توانایی تعمیم به داده‌های جدید را بهتر دارد. یعنی می‌تواند الگوهایی را که در داده‌های آموزشی یاد گرفته است، به داده‌های ندیده نیز تعمیم دهد.

۳- ارتباط VC Dimensionبا پدیده بیش‌برازسازی (Overfitting):

VC Dimension و بیش‌برازسازی دو مفهوم کلیدی در یادگیری ماشین هستند که ارتباط تنگاتنگی با هم دارند.

VC Dimension ، همانطور که قبلاً توضیح داده شد، به ما نشان می‌دهد که یک مدل چقدر پیچیده است و توانایی آن در یادگیری الگوهای پیچیده را اندازه می‌گیرد. هرچه VC Dimension بالاتر باشد، مدل پیچیده‌تر است و می‌تواند الگوهای پیچیده‌تری را یاد بگیرد.

بیش‌برازسازی زمانی اتفاق می‌افتد که یک مدل به اندازه‌ای پیچیده باشد که نه تنها الگوهای اساسی داده‌ها را یاد بگیرد، بلکه نویز و جزئیات تصادفی داده‌های آموزشی را نیز یاد بگیرد. در نتیجه، مدل در داده‌های آموزشی عملکرد بسیار خوبی دارد، اما در داده‌های جدید (که مدل قبلاً ندیده است) عملکرد ضعیفی نشان می‌دهد.

۳-۱ ارتباط بین VC Dimensionو بیش‌برازسازی:

VC Dimension بالا، خطر بیش‌برازسازی بیشتر: هرچه VC Dimension یک مدل بالاتر باشد، احتمال بیش‌برازسازی بیشتر است. زیرا مدل‌های پیچیده‌تر با VC Dimension بالا، توانایی تطبیق با هر نوع الگویی، حتی الگوهای تصادفی در داده‌های آموزشی را دارند.

تعادل بین پیچیدگی و تعمیم‌پذیری: VC Dimension به ما کمک می‌کند تا تعادل بین پیچیدگی مدل و توانایی آن در تعمیم به داده‌های جدید را برقرار کنیم. یک مدل با VC Dimensionبسیار پایین ممکن است نتواند الگوهای پیچیده داده‌ها را یاد بگیرد و به (underfitting) منجر شود. از سوی دیگر، یک مدل با VC Dimension بسیار بالا ممکن است به بیش‌برازسازی منجر شود.

نقش VC Dimension در انتخاب مدل: با استفاده از مفهوم VC Dimension ، می‌توانیم مدلی را انتخاب کنیم که پیچیدگی آن با تعداد داده‌های آموزشی متناسب باشد و خطر بیش‌برازسازی را کاهش دهیم.

۳-۲ مثال:

فرض کنید می‌خواهیم داده‌های یک مسئله طبقه‌بندی را با استفاده از یک مدل خطی و یک مدل چندجمله‌ای درجه ده مدل‌سازی کنیم. مدل چندجمله‌ای درجه ده پیچیده‌تر است و VC Dimensionبالاتری دارد. اگر تعداد داده‌های آموزشی کم باشد، مدل چندجمله‌ای درجه ده ممکن است به داده‌های آموزشی بیش از حد تطابق پیدا کند و در نتیجه در پیش‌بینی داده‌های جدید عملکرد ضعیفی داشته باشد. در حالی که مدل خطی، با وجود اینکه پیچیدگی کمتری دارد، ممکن است بتواند الگوهای اصلی داده‌ها را به خوبی یاد بگیرد و در نتیجه تعمیم‌پذیری بهتری داشته باشد.

۳-۳ راهکارهای کاهش بیش‌برازسازی مرتبط با VC Dimension:

کاهش پیچیدگی مدل: می‌توانیم با کاهش تعداد پارامترهای مدل یا استفاده از مدل‌های ساده‌تر، پیچیدگی مدل را کاهش دهیم و خطر بیش‌برازسازی را کاهش دهیم.

افزایش تعداد داده‌های آموزشی: با افزایش تعداد داده‌های آموزشی، می‌توانیم به مدل کمک کنیم تا الگوهای کلی داده‌ها را بهتر یاد بگیرد و از تأثیر نویز و جزئیات تصادفی داده‌های آموزشی بکاهد.

استفاده از روش‌های منظم‌سازی: روش‌های منظم‌سازی مانند L1 وL2 regularizationبه طور مستقیم با مفهوم VC Dimension مرتبط هستند و برای کاهش پیچیدگی مدل و جلوگیری از بیش‌برازش استفاده می‌شوند.

اعتبارسنجی متقاطع: با استفاده از اعتبارسنجی متقاطع می‌توانیم عملکرد مدل را بر روی داده‌های دیده نشده ارزیابی کنیم و از بیش‌برازش جلوگیری کنیم.

۳-۴ تاریخچه:

در اینجا، تاریخچه‌ای مختصر از این حوزه و پیشرفت‌های اخیر در زمینه مقابله با بیش‌برازسازی را بررسی می‌کنیم.

دهه ۱۹۵۰ و ۱۹۶۰: ریشه‌های یادگیری ماشین به این دهه‌ها برمی‌گردد. پژوهشگران اولیه بر روی شبکه‌های عصبی مصنوعی متمرکز بودند، اما به دلیل محدودیت‌های محاسباتی و عدم وجود داده‌های کافی، پیشرفت قابل توجهی حاصل نشد.

دهه ۱۹۸۰ و ۱۹۹۰: با پیشرفت رایانه‌ها و الگوریتم‌های جدید، یادگیری ماشین دوباره مورد توجه قرار گرفت. شبکه‌های عصبی مصنوعی دوباره احیا شدند و الگوریتم‌های یادگیری ماشین مانند SVMماشین بردار پشتیبان معرفی شدند. در این دوره، مشکل بیش‌برازسازی به عنوان یک چالش اساسی شناخته شد.

دهه ۲۰۰۰ تاکنون: با افزایش حجم داده‌ها و قدرت محاسباتی، یادگیری ماشین به یک حوزه تحقیقاتی بسیار فعال تبدیل شد. الگوریتم‌های یادگیری عمیق(Deep Learning) بر پایه شبکه‌های عصبی با لایه‌های متعدد، تحولی عظیم در این حوزه ایجاد کردند. با این حال، مشکل بیش‌برازسازی همچنان یکی از چالش‌های اصلی در یادگیری عمیق است.

۳-۵ پیشرفت‌های اخیر در مقابله با بیش‌برازسازی

۳-۵-۱ منظم‌سازی Regularization:

۳-۵-۱-۱ L1 وL2 regularization: این روش‌ها با افزودن یک جریمه به تابع هزینه، از پیچیدگی بیش از حد مدل جلوگیری می‌کنند.

۳-۵-۱-۲ Dropout: در این روش، به طور تصادفی برخی از نورون‌ها در طول آموزش غیرفعال می‌شوند تا از وابستگی بیش از حد به برخی نورون‌ها جلوگیری شود.

۳-۵-۱-۳ Early stopping: این روش با متوقف کردن آموزش مدل در زمانی که خطای اعتبارسنجی شروع به افزایش می‌کند، از بیش‌برازسازی جلوگیری می‌کند.

۳-۵-۲ افزایش حجم داده‌ها:با افزایش حجم داده‌های آموزشی، مدل می‌تواند الگوهای کلی داده‌ها را بهتر یاد بگیرد و از تأثیر نویز و جزئیات تصادفی داده‌های آموزشی بکاهد.

۳-۵-۳ اعتبارسنجی متقاطع Cross-validation:این روش برای ارزیابی عملکرد یک مدل بر روی داده‌های دیده نشده استفاده می‌شود. با تقسیم داده‌ها به چندین بخش، می‌توان عملکرد مدل را بر روی هر بخش ارزیابی کرد و از بیش‌برازسازی جلوگیری کرد.

۳-۵-٤ روش‌های پیشرفته در یادگیری عمیق:

۳-۵-۴-۱ Batch normalization : این روش با نرمال‌سازی ورودی هر لایه، به آموزش سریع‌تر و پایدارتر شبکه‌های عصبی کمک می‌کند و از بیش‌برازسازی جلوگیری می‌کند .

۳-۵-۴-۲ Attention mechanisms : این روش‌ها به مدل اجازه می‌دهند تا به قسمت‌های مهم ورودی توجه کند و از توجه به اطلاعات غیرضروری جلوگیری کند .

۳-۵-۵ روش‌های بیزی:

روش‌های بیزی با در نظر گرفتن عدم قطعیت در پارامترهای مدل، به کاهش بیش‌برازسازی کمک می‌کنند.

۴- مفاهیم بنیادی:

۴-۱ شکستن (Shattering) و تابع رشد در VC Dimension:

شکستن (Shattering) و تابع رشد دو مفهوم اساسی در نظریه یادگیری ماشین هستند که برای درک بهتر پیچیدگی مدل‌ها و توانایی آن‌ها در یادگیری الگوها استفاده می‌شوند. این مفاهیم به طور مستقیم با Dimension VC (Vapnik-Chervonenkis Dimension) در ارتباط هستند.

۴-۱-۱ شکستن (Shattering) چیست؟

تصور کنید مجموعه‌ای از نقاط داده داریم. یک مدل (مثلاً یک خط یا یک چندجمله‌ای) می‌تواند این نقاط را به دو دسته تقسیم کند. اگر یک مدل بتواند هر تقسیم‌بندی ممکن از این نقاط را ایجاد کند، گفته می‌شود که آن مدل مجموعه نقاط را شکسته است.

به عبارت دیگر، شکستن به معنای توانایی یک مدل در یادگیری هر الگوی ممکن در داده‌ها است.

۴-۱-۲ تابع رشد چیست؟

تابع رشد، تابعی است که تعداد حداکثری تقسیم‌بندی‌های مختلفی که یک مدل می‌تواند برای یک مجموعه با اندازه مشخص ایجاد کند را نشان می‌دهد. به عبارت دیگر، تابع رشد نشان می‌دهد که یک مدل چقدر انعطاف‌پذیر است و می‌تواند الگوهای پیچیده‌ای را یاد بگیرد.و به ما نشان می‌دهد یک مدل چقدر قادر است داده‌های مختلف را به دسته‌های مختلف تقسیم کند. به عبارت دیگر، تابع رشد نشان‌دهنده پیچیدگی یک مدل است. هرچه تابع رشد بالاتر باشد، مدل پیچیده‌تر است و می‌تواند الگوهای پیچیده‌تری را یاد بگیرد.

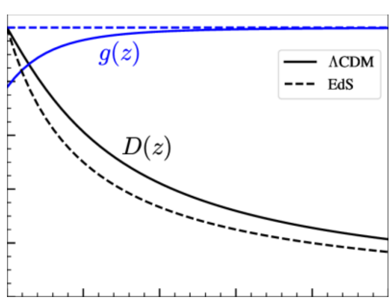
۴-۱-۲-۱ نمودار تابع رشد:

نموداری است که تعداد تقسیم‌بندی‌های ممکن یک مدل را بر حسب تعداد نقاط داده نشان می‌دهد. این نمودار به ما یک تصویر بصری از اینکه چگونه پیچیدگی یک مدل با افزایش تعداد داده‌ها تغییر می‌کند، ارائه می‌دهد.

از انواع نمودارهای تابع رشد

نمودارهای تابع رشد برای مدل‌های مختلف، شکل‌های متفاوتی دارند. برخی از نمونه‌های رایج عبارتند از:

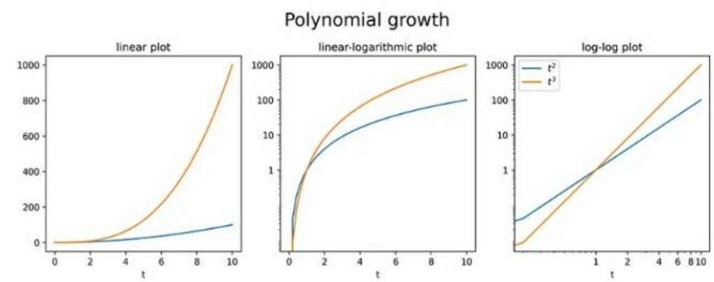
۴-۱-۲-۲ مدل‌های خطی: برای مدل‌های خطی، تابع رشد به صورت نمایی افزایش می‌یابد. این بدان معنی است که با افزایش تعداد نقاط داده، تعداد تقسیم‌بندی‌های ممکن به سرعت افزایش می‌یابد.



در این نمودار، محور افقی تعداد نقاط داده و محور عمودی تعداد تقسیم‌بندی‌های ممکن را نشان می‌دهد. همانطور که مشاهده می‌کنید، تابع رشد مدل‌های خطی به صورت نمایی افزایش می‌یابد. این بدان معنی است که با افزایش تعداد نقاط داده، تعداد تقسیم‌بندی‌های ممکن به سرعت افزایش می‌یابد.

۴-۱-۲-۳ مدل‌های چندجمله‌ای:

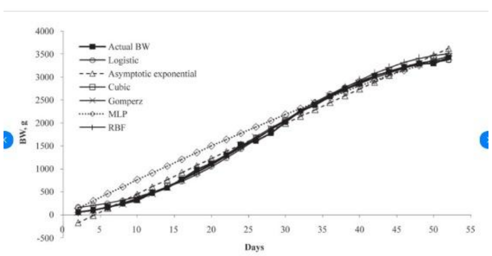
برای مدل‌های چندجمله‌ای، تابع رشد سریع‌تر از مدل‌های خطی افزایش می‌یابد. هرچه درجه چندجمله‌ای بالاتر باشد، تابع رشد نیز سریع‌تر رشد می‌کند.



در این نمودار، تابع رشد مدل‌های چندجمله‌ای سریع‌تر از مدل‌های خطی افزایش می‌یابد. هرچه درجه چندجمله‌ای بالاتر باشد، تابع رشد نیز سریع‌تر رشد می‌کند.

۴-۱-۲-۴ شبکه‌های عصبی:

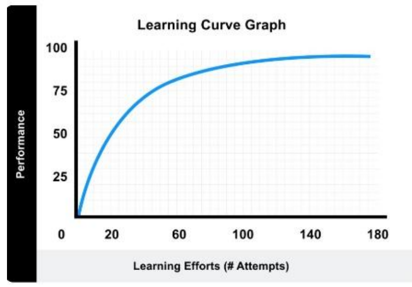
برای شبکه‌های عصبی، تابع رشد به شدت به معماری شبکه و تعداد پارامترها بستگی دارد. به طور کلی، شبکه‌های عصبی با تعداد لایه‌ها و پارامترهای بیشتر، تابع رشد بالاتری دارند.



تابع رشد شبکه‌های عصبی به شدت به معماری شبکه و تعداد پارامترها بستگی دارد. به طور کلی، شبکه‌های عصبی با تعداد لایه‌ها و پارامترهای بیشتر، تابع رشد بالاتری دارند.

۴-۱-۲-۵ نمودار منحنی یادگیری :

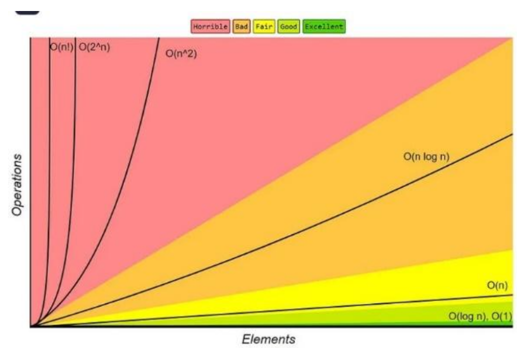
منحنی یادگیری، نموداری است که رابطه بین میزان تلاش و زمان صرف شده برای یادگیری یک مهارت جدید و افزایش توانایی در انجام آن را نشان می‌دهد. این نمودار به ما کمک می‌کند تا پیشرفت خود را در یادگیری یک مهارت جدید بهتر درک کنیم و همچنین به ما نشان می‌دهد که در کدام مراحل یادگیری، سرعت پیشرفت بیشتر یا کمتر است.



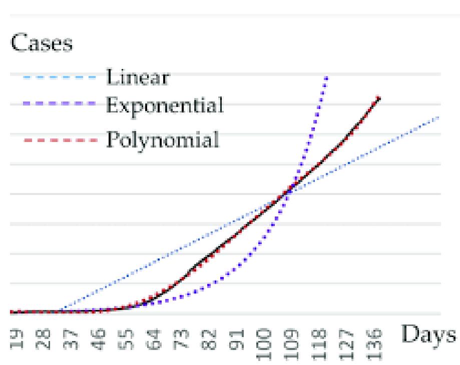
منحنی یادگیری، نوع دیگری از نمودار است که به ما کمک می‌کند تا پیچیدگی مدل را ارزیابی کنیم. این نمودار نشان می‌دهد که چگونه خطای مدل بر روی داده‌های آموزشی و اعتبارسنجی با افزایش تعداد دورهای آموزش تغییر می‌کند.

۴-۱-۲-۶ نمودار پیچیدگی محاسباتی:

نمودار پیچیدگی محاسباتی، ابزاری قدرتمند برای تحلیل و مقایسه کارایی الگوریتم‌های مختلف است. این نمودار نشان می‌دهد که با افزایش حجم ورودی، زمان اجرا یا تعداد عملیات یک الگوریتم چگونه تغییر می‌کند. به عبارت ساده‌تر، این نمودار به ما می‌گوید که یک الگوریتم با بزرگ‌تر شدن مسئله، چقدر کندتر می‌شود.



۴-۱-۳ مثال، مقایسه تابع رشد مدل‌های خطی و چندجمله‌ای:



در این نمودار، محور x تعداد نقاط داده و محور y تعداد تقسیم‌بندی‌های ممکن را نشان می‌دهد. خط آبی نشان‌دهنده تابع رشد یک مدل خطی و خط قرمز نشان‌دهنده تابع رشد یک مدل چندجمله‌ای درجه دو است.

همانطور که مشاهده می‌شود، تابع رشد مدل چندجمله‌ای بسیار سریع‌تر از مدل خطی افزایش می‌یابد. این بدان معنی است که مدل چندجمله‌ای پیچیده‌تر است و می‌تواند الگوهای پیچیده‌تری را یاد بگیرد، اما در عین حال بیشتر مستعد بیش‌برازسازی است.

۴-۲ رابطه بین شکستن، تابع رشد و VC Dimension

۴-۲-۱ VC Dimension:

VC Dimension به عنوان کوچک‌ترین اندازه یک مجموعه از نقاط است که توسط یک مدل می‌تواند شکسته شود. به عبارت دیگر، VC Dimension نشان‌دهنده پیچیدگی حداکثری مدلی است که می‌تواند تمام تقسیم‌بندی‌های ممکن یک مجموعه را ایجاد کند.

۴-۲-۲ تابع رشد و VC Dimension:

تابع رشد و VC Dimension به هم مرتبط هستند. هرچه VC Dimension یک مدل بالاتر باشد، تابع رشد آن نیز سریع‌تر رشد می‌کند، به این معنی که مدل می‌تواند الگوهای پیچیده‌تری را یاد بگیرد.

۴-۳ اهمیت شکستن و تابع رشد در یادگیری ماشین:

درک پیچیدگی مدل: شکستن و تابع رشد به ما کمک می‌کنند تا پیچیدگی یک مدل را بهتر درک کنیم.

جلوگیری از بیش‌برازش: مدل‌هایی با VC Dimension بالا و تابع رشد سریع، مستعد بیش‌برازش هستند. این بدان معنی است که آن‌ها ممکن است الگوهای تصادفی در داده‌های آموزشی را یاد بگیرند و در نتیجه در پیش‌بینی داده‌های جدید عملکرد ضعیفی داشته باشند.

انتخاب مدل مناسب: با درک شکستن و تابع رشد، می‌توانیم مدل مناسبی را برای مسئله خود انتخاب کنیم. مدل‌هایی با VC Dimension پایین‌تر، احتمال بیش‌برازش کمتری دارند اما ممکن است نتوانند الگوهای پیچیده را یاد بگیرند. از سوی دیگر، مدل‌های با VC Dimension بالاتر، توانایی یادگیری الگوهای پیچیده را دارند اما احتمال بیش‌برازش آن‌ها بیشتر است.

۴-۳-۱ مثال:

بیایید یک مثال ساده‌تر را در نظر بگیریم: خطوط در صفحه دو بعدی. یک خط می‌تواند صفحه را به دو نیم‌فضا تقسیم کند.

VC Dimension:برای شکستن سه نقطه غیر همخط، همیشه می‌توان یک خط یافت که آن‌ها را به دو دسته تقسیم کند.

اما برای شکستن چهار نقطه، همیشه نمی‌توان یک خط یافت که آن‌ها را به دو دسته تقسیم کند. مثلاً اگر چهار نقطه یک مربع را تشکیل دهند، هیچ خطی نمی‌تواند آن‌ها را به دو دسته تقسیم کند که هر دسته دقیقا دو نقطه داشته باشد.بنابراین، VC Dimensionبرای خطوط در صفحه دو بعدی برابر با ۳ است.

تابع رشد:برای یک مجموعه شامل m نقطه، حداکثر تعداد تقسیم‌بندی‌های ممکن توسط یک خط برابر با $m^۲$ است. زیرا هر نقطه می‌تواند در یکی از دو طرف خط قرار بگیرد.

بنابراین، تابع رشد برای خطوط در صفحه دو بعدی برابر با ۲= $m^۲$ است.

۲-۳-۴ مثال:

چندجمله‌ای‌های درجه دو در صفحه

در مثال قبلی، ما به سادگی خطوط مستقیم را در نظر گرفتیم. حال بیایید به سراغ یک مدل پیچیده‌تر برویم: چندجمله‌ای‌های درجه دو در صفحه. یک چندجمله‌ای درجه دو به شکل زیر است:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

چالش‌ها:

فضای پارامتر بزرگ‌تر: در مقایسه با خطوط مستقیم، تعداد پارامترهای بیشتری درگیر هستند (a, b, c, d, e, f).

انواع مختلف منحنی‌ها :چندجمله‌ای‌های درجه دو می‌توانند انواع مختلفی از منحنی‌ها را ایجاد کنند، از دایره‌ها و بیضی‌ها گرفته تا سهمی‌ها و.....

پیچیدگی هندسی :تعیین اینکه آیا یک مجموعه از نقاط می‌تواند توسط یک چندجمله‌ای درجه دو به دو دسته تقسیم شود، بسیار پیچیده‌تر از خطوط مستقیم است .

محاسبه دقیق (تقریبی): محاسبه دقیق VC Dimensionبرای چندجمله‌ای‌های درجه دو بسیار دشوار است. با این حال، می‌توانیم از کران‌های بالا و پایین برای تخمین آن استفاده کنیم.

کران بالا :با استفاده از قضایای موجود در نظریه یادگیری ماشین، می‌توان نشان داد که VC Dimension برای چندجمله‌ای‌های درجه دو در صفحه، حداقل به اندازه تعداد پارامترهای آزاد است. در این مورد، تعداد پارامترهای آزاد ٦ است. بنابراین، VC Dimensionحداقل ٦ است .

کران پایین :پیدا کردن یک کران پایین دقیق‌تر نیازمند تحلیل‌های هندسی پیچیده‌تری است. به طور کلی، می‌توانیم بگوییم که VC Dimensionبزرگتر از ۳ است (زیرا هر خط مستقیم یک مورد خاص از یک چندجمله‌ای درجه دو است) .

تابع رشد: محاسبه دقیق تابع رشد برای چندجمله‌ای‌های درجه دو حتی دشوارتر است. با این حال، می‌توانیم بگوییم که تابع رشد به طور نمایی با تعداد نقاط افزایش می‌یابد.

چرا این مثال پیچیده‌تر است؟

تنوع هندسی: چندجمله‌ای‌های درجه دو می‌توانند انواع مختلفی از منحنی‌ها را ایجاد کنند که منجر به پیچیدگی بیشتر در تحلیل می‌شود.

تعامل پارامترها: پارامترهای مختلف در معادله چندجمله‌ای با هم تعامل دارند و این تعامل باعث پیچیدگی محاسبات می‌شود.

روش‌های تقریبی: برای محاسبه تقریبی VC Dimension و تابع رشد برای مدل‌های پیچیده، از روش‌های مختلفی استفاده می‌شود:

روش‌های نمونه‌گیری: با انتخاب تصادفی مجموعه‌های مختلف از نقاط و بررسی اینکه آیا مدل می‌تواند آن‌ها را شکسته شود، می‌توان تخمین‌هایی از VC Dimension و تابع رشد بدست آورد.

روش‌های آماری: با استفاده از روش‌های آماری، می‌توان کران‌های بالا و پایین دقیق‌تری برای VC Dimension و تابع رشد بدست آورد.

شبکه‌های عصبی مصنوعی: برای مدل‌های بسیار پیچیده مانند شبکه‌های عصبی، از روش‌های خاصی برای تخمین VC Dimension استفاده می‌شود.

۵- VC Dimension و VC :

۵-۱ انواع نامساوی‌های VC و انواع VC dimension از نظر مفهوم و کاربرد متفاوت هستند:

۵-۱-۱ نامساوی‌های VC:

نامساوی‌های VC ابزارهایی برای تعیین ظرفیت تعمیم یک کلاس از توابع در یادگیری ماشین هستند. این نامساوی‌ها معمولاً به بررسی ارتباطات بین تعداد نقاط، ظرفیت مدل و خطای تعمیم می‌پردازند. مهم‌ترین نامساوی‌ها عبارتند از

۵-۱-۲ نامساوی Sauer:

حداکثر تعداد توابعی که می‌توانند n نقطه را به صورت کامل تفکیک کنند را بیان می‌کند.

فرمول:

$$|H(X)|\leq\sum_{i=0}^{d}\binom{n}{i}$$

۵-۱-۳ نامساوی Vapnik-Chervonenkis:

این نامساوی به ارتباط بین خطای تعمیم، خطای یادگیری و VC dimension اشاره دارد.

فرمول:

$$Error(H)\leq Training\ Error(H)+\sqrt{\frac{d\ln(n/d)+\ln(1/\delta)}{n}}\ Type$$

۵-۱-۴ نامساوی Talagrand:

این نامساوی به تعمیم نظریه احتمال‌ها و ظرفیت بر اساس VC dimension می‌پردازد و روابط متفاوتی را ارائه می‌دهد

۵-۲ VC Dimension:

از سوی دیگر، VC dimension معیاری است برای سنجش ظرفیت تفکیکی یک کلاس از توابع. این بعد به حداکثر تعدادی از نقاطی اشاره دارد که می‌توانند به طور کامل تفکیک شوند. انواع VC dimension عبارتند از:

۵-۲-۱ VC Dimension:

همان‌طور که اشاره شد، حداکثر تعداد نقاطی است که می‌توانند به شکلی از یکدیگر جدا شوند توسط یک کلاس تابع.به عنوان مثال، VC dimension یک کلاس خطی در R2 برابر ۳ است، زیرا می‌توان ۳ نقطه را به‌طور کامل تفکیک کرد.

۵-۲-۲ Ideal VC Dimension:

در برخی موارد، یک VC dimension ایده‌آل می‌تواند به عنوان یک هدف مورد نظر برای مدل‌های یادگیری تعریف شود، به‌طوری‌که به تعادل ایده‌آلی از پیچیدگی و توانایی تعمیم دست یابند.

۵-۲-۳ Empirical VC Dimension:

به VC dimension براساس داده‌های واقعی و تجربه‌های یادگیری اشاره دارد و ممکن است با VC dimension تئوری متفاوت باشد

۵-۳ تفاوت‌ها:

۵-۳-۱ نوع استفاده: نامساوی‌های VC برای تحلیل و پیش‌بینی عملکرد مدل‌ها و ارزیابی ظرفیت تعمیم استفاده می‌شوند، در حالی که VC dimension به خودی خود به ظرفیت تفکیک‌پذیری مدل‌ها اشاره دارد.

۵-۳-۲ تعاریف: VC dimension یک مفهوم خاص و هدفمند است که یک عدد مشخص برای هر کلاس تابع را بیان می‌کند، اما نامساوی‌ها به روابط بین جنبه‌های مختلف یادگیری ماشین می‌پردازند.

۵-۳-۴ کاربرد: نامساوی‌ها در فرآیندهای تحلیلی و تضمین‌های نظری استفاده می‌شوند، در حالی که VC dimension به توصیف ساختار و قابلیت مدل‌ها در تفکیک داده‌ها کمک می‌کند.

بنابراین، نامساوی‌های VC و VC dimension دو مفهوم متفاوت هستند که هر کدام نقش مهمی در نظریه یادگیری ماشین و تحلیل ظرفیت مدل‌ها دارند.

۶- رابطه بین VC Dimension و پیچیدگی مدل و نامساوی VC و اهمیت آن در تئوری یادگیری:

نامساوی **VC** (Vapnik-Chervonenkis dimension) یکی از مفاهیم بنیادین در نظریه یادگیری ماشین است که به ما کمک می‌کند تا پیچیدگی و توانایی تعمیم‌پذیری یک مدل یادگیری را ارزیابی کنیم.

این نامساوی به نام ورنون چن (Vladimir Vapnik) و الکساندر چرووننکیس Alexandr (Chervonenkis) نامگذاری شده است که سهم بسزایی در توسعه این مفهوم داشته‌اند.

نامساوی VCیک رابطه ریاضی است که ارتباط بین بعد VCیک مدل، اندازه مجموعه داده و احتمال خطای تعمیم‌پذیری را نشان می‌دهد. این نامساوی به ما می‌گوید که با افزایش پیچیدگی یک مدل (یعنی افزایش VC Dimension)، احتمال اینکه مدل بیش از حد به داده‌های آموزشی تطبیق پیدا کند (بیش‌برازش) و در نتیجه روی داده‌های جدید عملکرد ضعیفی داشته باشد، بیشتر می‌شود.

به زبان ساده‌تر**:** اگر یک مدل خیلی پیچیده باشد، حتی می‌تواند نویز و تصادف‌های موجود در داده‌های آموزشی را نیز یاد بگیرد. در نتیجه، زمانی که این مدل با داده‌های جدیدی روبرو می‌شود که حاوی این نویزها نیست، عملکرد ضعیفی خواهد داشت.

نامساوی VC dimensionبه خودی خود یک فرمول خاص ندارد، بلکه به عنوان یک مفهوم در نظریه یادگیری ماشین و نظریه عمومی‌سازی (generalization theory) معرفی می‌شود. این مفهوم به حداکثر تعداد نقاطی اشاره دارد که می‌توانند توسط یک مجموعه از توابع طبقه‌بندی به‌طور کامل جدا شوند.

۶-۱ محدودیت‌های نامساوی VC:

۶-۱-۱ فرضیات ساده‌کننده: نامساوی VCبر اساس برخی فرضیات ساده‌کننده استوار است که ممکن است در دنیای واقعی همیشه برقرار نباشند.

۶-۲-۲ محاسبه پیچیده VC Dimension : محاسبه دقیق VC Dimension برای مدل‌های پیچیده بسیار دشوار است.

۶-۲ کاربردهای نامساوی VC:

۶-۲-۱انتخاب روش‌های تنظیم پارامتر: نامساوی VCمی‌تواند در انتخاب روش‌های تنظیم پارامتر (hyperparameter tuning) مانند regularizationاستفاده شود.

۶-۲-۲تخمین خطای تعمیم‌پذیری: با استفاده از نامساوی VCمی‌توان تخمین‌هایی از خطای تعمیم‌پذیری یک مدل ارائه داد.

۶-۲-۳طراحی الگوریتم‌های یادگیری جدید: در طراحی الگوریتم‌های یادگیری جدید، نامساوی VCمی‌تواند به عنوان یک ابزار نظری مفید باشد.

۶-۳ تکنیک‌های ریاضی مورد استفاده در اثبات نامساوی VC:

اثبات نامساوی VC، این تکنیک‌ها به ما کمک می‌کنند تا رابطه بین پیچیدگی یک مدل (که VC Dimension اندازه‌گیری می‌شود)، تعداد داده‌های آموزشی و احتمال خطای تعمیم‌پذیری را درک کنیم.و به برخی از مهم‌ترین تکنیک‌های ریاضی مورد استفاده در این اثبات اشاره می‌کنیم.که چند نمونه را بررسی خواهیم کرد.

۱-۳-۶تئوری احتمال:

۱-۱-۳-۶ نامساوی‌های احتمال:

۱-۱-۱-۳-۶نامساوی مک‌دیارمید (McDiarmid's inequality):

این نامساوی بسیار مهم به ما اجازه می‌دهد تا احتمال انحراف یک متغیر تصادفی از مقدار انتظار آن را محدود کنیم. در اثبات نامساوی VC، از این نامساوی برای محدود کردن احتمال تفاوت بین خطای تجربی و خطای واقعی استفاده می‌شود.

نامساوی مک دیارمید به زبان ساده:نامساوی مک دیارمید یکی از مفاهیم اساسی ریاضیات است که در حل مسائل به کار میرود.این نامساوی به صورت عمومی به شکل زیر است:

$$ax+b<c$$

در این نامساوی $a \neq 0$و $b, c \in R$ است.

مثال: برای حل نامعادله زیر و بدست آوردن مقادیر ممکن برای x :

$$2x+3<7$$

ابتدا از دو طرف نامعادله ۳ واحد کم می‌کنیم:

$$2x+3-3<7-3$$

$$2x<4$$

سپس برای حذف ضریب x، دو طرف نامعادله را در عبارت 1/2 ضرب می‌کنیم:

$$1/2 \times 2x < ½ \times 4$$

$$x<2$$

$$x \in (-\infty,2)$$

پس مقادیر ممکن برای x تمام اعداد حقیقی کوچکتر از ۲ می‌باشند.

۲-۱-۱-۳-۶ نامساوی چرنوف (Chernoff bound):

این نامساوی برای محدود کردن احتمال انحراف شدید یک مجموع از متغیر های تصادفی مستقل از مقدار انتظار آن استفاده می‌شود.

زمانی که تابع مولد متغیر تصادفی ط معلوم باشد ، میتوان کرانهای خوبی برای $P\{X<a\}$ بدست آوریم.

$$P\{X \geq a\} \leq e^{-ta} M(t) \qquad (t > 0)$$

$$P\{X<a\} \leq e^{ta} M(-t) \qquad (t > 0)$$

چون کران‌های چرنوف برای همه tها مثبت و منفی برقرار است، می‌توان بهترین کران برای $P(X\geq a)$ را با استفاده از e^{-ta} $M(t)$ را حداقل میکند به دست آورد.کران های چرنوف برای متغیر تصادفی نرمال استاندارد ،اگر Z متغیر تصادفی نرمال استاندارد باشد، آنگاه تابع مولد گشتاوری آن،

$M(t)= e^{t2/2}$ است.پس کران چرنوف برای P(Z≥a) به صورت زیر است:

$$P(Z\geq a)\leq e^{-ta}e^{\frac{t^2}{2}}$$

حالا مقدار t مثبتی که $e^{\left(\frac{t2}{2}\right)-ta}$ را حداقل میکنیم.

$$P(Z\geq a)\leq e^{\frac{-a^2}{2}}$$

$$P(Z\leq a)\leq e^{\frac{-a^2}{2}}$$

مثال:حال مثالی از کاربرد این کران را نشان می دهیم . فرض کنیم X_n و .... X_1 متغییر های تصادفی مستقل با توزیع برنولی باشند،که هر کدام دارای احتمال p> ½ باشند ، احتمال رخداد همزمان بیش از n/2 متغیر تصادفی $\{x_k=1\}$ برابر است با:

$$s = \sum_{i=\left[\frac{n}{2}\right]+1}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

کران چرنوف نشان می دهد که رابطه بالا دارای کران پایین زیر است

توجه کنید که برای تعدادی نمونه تصادفی با توزیع برنولی داریم $u = np$ با استفاده از نتایج کران چرنوف میتوان نشان داد که:

$$P\left[s \leq \frac{n}{2}\right] = p\left[s \leq (1-\frac{1}{2p})u\right] \leq e \wedge -\frac{u}{2p}\, n \left(p-\frac{1}{2p}\right)2 = e \wedge -\frac{1}{2p}\, n \left(p-\frac{1}{2}\right)2$$

باید توجه کرد که با توجه به خواسته های مسئله می توان از فرم های مختلفی از کران چرنوف استفاده کرد.

$$S \geq 1 - e \wedge \left(-\frac{1}{2p}\, n \left(p-\frac{1}{2}\right)2\right)$$

۲-۱-۳-۶ توزیع‌های احتمال:

۱-۲-۱-۳-۶ توزیع برنولی:

برای مدل‌سازی رخدادهای برنولی (مثلاً درست یا غلط بودن یک طبقه‌بندی) استفاده می‌شود. توزیع برنولی: توزیعی است که نام آن از نام دانشمند سوئیسی ژاکوب برنولی گرفته شده‌است. توزیع برنولی یک توزیع گسسته است که مقادیر یک (در صورت موفقیت آزمایش) و صفر را (در صورت شکست) می‌گیرد.

احتمال موفقیت آزمایش برابر p است و احتمال شکست آن برابر q=1-p است. بنابراین اگر X یک متغیر تصادفی با توزیع برنولی باشد، داریم:

$$1-q=p\cdot 1-\Pr(X=0)=\Pr(X=1)$$

و تابع توزیع(pmf) آن به صورت زیر خواهد بود:

$$\boxed{f(k;p)=}\begin{cases} if\ k=1 & P \\ 1-p & if\ k=0 \\ 0 & otherwise \end{cases}$$

امید ریاضی این توزیع برابر p و واریانس آن برابر p(1−p) است. کشیدگی این توزیع برای مقادیر p نزدیک به صفر یا یک، به سمت بی‌نهایت میل می‌کند و برای p=0.5 کمترین مقدار کشیدگی را خواهیم داشت.توزیع برنولی جز خانواده نمایی طبقه‌بندی می‌شود.

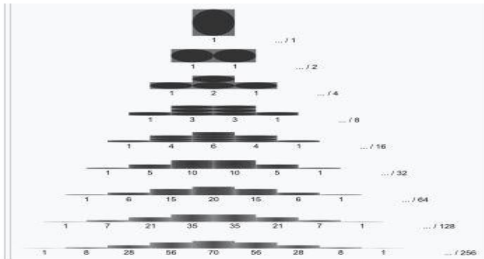
۲-۲-۱-۳-۶توزیع های مرتبط: اگر $X1,\ldots,Xn$ متغیرهای تصادفی با توزیع برنولی با پارامتر یکسان و مستقل باشند، آنگاه:متغیر تصادفی $\sim\text{Binomial}(n,p)$ $Y=\sum_{k=1}^{n}(xk)$ یک توزیع دوجمله ای خواهد بود.در واقع توزیع برنولی همان توزیع دو جمله‌ای با پارامتر n=1 یعنی $\text{Binomial}(n=1,p)$ خواهد بود.درواقع تابع جرم توزیع دوجمله ای به صورت زیر میباشد:

$$P(x)=\binom{n}{x}p^{x}(1-p)^{n-x}$$

برنولی	
پارامترها	شانس موفقیت P>0
تکیه گاه	k{={0,1}
تابع چگالی احتمال	q for k = 0 P for k = 1
تابع توزیع تجمعی	0 for k < 0 q for 0 ≤ k < 1 1 For k ≥ 1
میانگین	P
میانه	N/A
مد	0 if q > p 0,1 if q = p 1 If q < p
واریانس	Pq
چولگی	$\frac{q-p}{\sqrt{pq}}$
کشیدگی	$\frac{6p^2-6p+1}{p(1-p)}$
آنتروپی	- Q ln (q) – p ln (p)
تابع مولد گشتاور	Q + pe^t

تابع مشخصه	$Q + pe^{it}$

۲-۲-۱-۳-۶ توزیع دو جمله‌ای: برای مدل‌سازی تعداد موفقیت‌ها در تعداد مشخصی آزمایش برنولی استفاده می‌شود.



توزیع دوجمله‌ای برای p=۰٫۵ , مثلث خیام _ پاسکال

مشخصه ها:

۱-۲-۲-۱-۳-۶ تابع جُرم احتمال:

در حالت کلی اگر X یک متغیر تصادفی دو جمله ای با پارامترهای n و p باشد، آن را به صورت X~B(n,p) نشان می‌دهیم. احتمال به دست آوردن k برابر است با تابع جرم احتمال زیر مشخص می‌شود:

$$P(k) = pr(X=k) = \binom{n}{k} p^{k}(1-p)^{\,n-k}$$

$$\text{For} \quad k = 0 , 1 , 2 , \ldots\ldots, n$$

حال می‌خواهیم بررسی کنیم که این فرمول چگونه بدست آمده است. توجه کنید که تعداد راه‌های ممکن در انجام n آزمایش برنولی که می‌تواند به k موفقیت منتهی شود برابر است با تعداد دنباله های مختلف به طول n از حروف a , b با k حرف a(موفقیت) و n-k حرف b(شکست). اما تعداد این دنباله برابر است با $\binom{n}{k}$. زیرا تعداد جایگشت‌های متمایز n حرف از دو نوع مختلف، k تا همانند از نوع اول و n−k تا همانند از نوع دوم وجود دارد برابر است با $\frac{n!}{k!(n-k)!}$.

با توجه به استقلال امتحان ها چون احتمال هر یک از این دنباله پیشامد ها برابر $p^{k}(1-p)^{n-k}$ است داریم:

$$Pr(x=k) = \binom{n}{k} p^{k} (1-p)^{n-k}$$

دلیل اینکه به این توزیع دو جمله ای میگویند این است که قضیه بسط دوجمله ای تضمین میکند که p(k) یک تابع جرم احتمال است:

$$\boxed{\sum_{k=0}^{d} p(k) = \sum_{k=0}^{d}\binom{n}{k}(1-p)^{n-k} p^{k} = [p + (1-p)]^{n} = 1^{n} = 1}$$

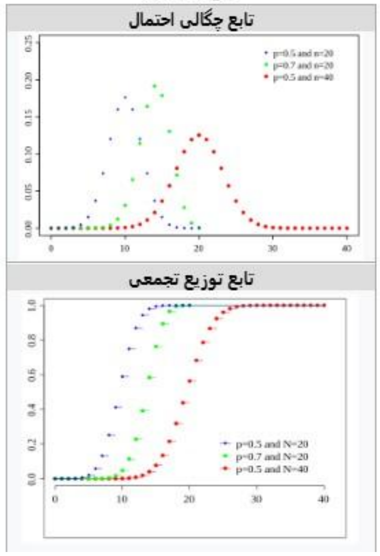
۲-۲-۲-۱-۳-۶تابع توزیع تجمیعی

تابع توزیع تجمعی متغیر تصادفی دو جمله ای به شکل زیر است:

$$F(k;n,p) = pr(X \le k) = \sum_{i=0}^{k}\binom{n}{i}\ p\ \hat{}\ i\ (1-p)n\ \hat{}-i$$

دو جمله ای	
پارامتر ها	n>0 تعداد تکرار ها شانس موفقیت$0 \le p \le 1$
تکیه گاه	kϵ{0,…….,n}
تابع چگالی احتمال	$\binom{n}{k}$ p ^{k} (1-p) $^{n-k}$
تابع توزیع تجمعی	I_{1-p}(n-[k],1+[k])
میانگین	np
میانه	یکی از [np]-1, [np],[np]+1}{
مد	[(n+1)p]
واریانس	Np(1-p)
چولگی	$\frac{1-2p}{\sqrt{np(1-p)}}$
کشیدگی	$\frac{1-6p(1-p)}{np(1\ -p)}$
آنتروپی	½ ln (2πnep(1-p))+O($\frac{1}{n}$)
تابع مولد گشتاور	(1-p+pe ^{t})^{n}
تابع مشخصه	(1-p+pe ^{it})^{n}

دوجمله‌ای



مثال:

اگر یک تیرانداز با احتمال 0.7 تیری را به هدف می‌زند و تعداد تیرهای او هدف خورده در ۵ شلیک باشد، ابتدا توزیع احتمال x را معلوم کنید و هر یک از احتمال‌ها را به دست آورید.

۱.دقیقاً ۳ تیر به هدف بزند

۲. حداکثر ۳ تیر به هدف بزند

۳.هیچ تیری به هدف نزند

متغیر تصادفی X دارای توزیع دو جمله‌ای با پارامترهای n=5 و p=0.7 است که تابع احتمال آن را به صورت زیر می‌نویسیم:

$$P(k)=\binom{n}{k}\,(0.7)^k(0.3)^{5-k}$$

برای به دست آوردن احتمالات در این مثال داریم:

$$P(x=3)=\binom{5}{3}\,(0.7)^3(0.3)^2=0.3087$$

$$P(x\leq 2)= p(0) + p\,(1) + p(2)=0.16308$$

$$P(x=0) = p(0) = 0.00243$$

۳-۲-۲-۱-۳-۶میانگین و واریانس متغیرهای تصادفی دو جمله‌ای

فرض کنید Xیک متغیر تصادفی دو جمله‌ای با پارامترهای nو pباشد. به طور شهودی انتظار داریم که میانگین Xبرابر np باشد.برای مثال اگر سکه سالمی را صد بار پرتاب کنیم، انتظار داریم به طور متوسط ۵۰ بار شیر مشاهده کنیم که برابر است با:

np=100*(1/2). فرمول E[X] = np را میتوان مستقیما از تعریف امید ریاضی بدست آورد. در زیر این فرمول های امید ریاضی و واریانس این متغیر را آورده ایم:

$$E[X] = np$$

$$Var[X] = np(1-p)$$

مثال:تاسی را ۳ بار پرتاب می کنیم. اگر متغیر تصادفی X تعداد تاسهایی که شش آمده باشد،تابع احتمال X را بنویسید و امید ریاضی و واریانس آن را محاسبه کنید.

چون پرتاب ها از یکدیگر مستقل اند و احتمال موفقیت (شش آمدن) در هر پرتاب ۶/۱ است و این آزمایش n=3 بار تکرار می شود ، بنابراین شرایط توزیع دوجمله ای با p=1/6 ، n=3 برقرار است و توزیع احتمال X را به صورت زیر می نویسیم:

For \ quad k=0 ,1,2,3

$$P(k)=\binom{3}{k}\ (1/6)^{k}(5/6)^{3-k}$$

$$E[X] = np=3*1/6$$

$$Var[X] = np(1-p)=3 * 1/6 * 5/6$$

۲-۳-۶ ترکیبیات:

۱-۲-۳-۶ شمارش حالت‌ها:

برای شمارش تعداد حالت‌های مختلفی که یک مدل می‌تواند داده‌ها را طبقه‌بندی کند، از تکنیک‌های شمارش استفاده می‌شود.

۱-۱-۲-۳-۶فرمول عمومی:

اگر مجموعه ای متشکل از n نقطه داشته باشیم ، تعداد روش های ممکن برای تقسیم این نقاط به دو گروه (به عنوان مثال: دو کلاس) برابر است با:

$$2^n$$

این به این معناست که برای هر نقطه می توان تصمیم گرفت که آیا در گروه اول باشد یا دوم.بنابراین ، با تمام نقاط ، تعداد تقسیمات ممکن 2^n خواهد بود.

۲-۱-۲-۳-۶ VC استفاده از نامساوی

به ما می گوید که اگر H یک خانواده از توابع باشد که n نقطه را می تواند شکل دهد و به آن می گویند VC نامساوی ، آنگاه VC که ظرفیت آن d است :

$$|H| \leq \binom{n}{d} \cdot (n \leq d)$$

که در آن $\binom{n}{d}$ که می تواند به عنوان تعداد حالت های انتخاب d عنصر از n عنصر شمرده شود می باشد. این فرمول در واقع بیان گر این است که یک مدل نمی تواند بیش از تعداد مشخصی از تقسیم ها برای هر مجموعه از داده ها ایجاد کند.

۲-۲-۳-۶ اصل شمول و استثنا: برای محاسبه تعداد عناصر در اجتماع چند مجموعه استفاده می‌شود.

در اثبات نامساوی VC، این اصل می‌تواند به ما کمک کند تا تعداد حالت‌های ممکن تقسیمات یک خانواده توابع را به درستی محاسبه کنیم.بیان فرمول شمول و استثنا:

برای kمجموعه $A_1,A_2,\ldots\ldots,A_k$، تعداد عناصر حداقل یکی از این مجموعه ها به صورت زیر محاسبه میشود:

$$|A_1 \cup A_2 \cup \ldots\ldots \cup A_K| = \sum_{i=1}^{K} |Ai| - \sum_{1\leq i<j<l\leq k} |Ai \cap Aj \cap Al| - \cdots +$$

توضیح اجزای فرمول:

<u>**ترم‌های مثبت**</u> $\sum_{i=1}^{K} |Ai|$مجموع اندازه‌های همه مجموعه‌ها را در نظر می‌گیرد.

<u>**ترم‌های منفی:**</u> $|Ai \cap| - \sum_{1\leq i<j<l\leq k} |Ai|$از تعداد مواردی که در تقاطع هر دو مجموعه وجود دارند، کم می‌کند. زیرا این موارد به دو بار در ترم‌های مثبت حساب شده‌اند.

<u>**ترم‌های بالا:**</u> $|Ai \cap| - \sum_{1\leq i<j<l\leq k} |Aj| A_i$ مجدداً موارد تقاطع‌های چندگانه را اضافه می‌کند، زیرا این‌ها از ترم‌های قبلی کاسته شدند.

این روند ادامه می‌یابد با تغییر علامت برای هر تعداد مجموعه.

مثال: فرض کنید که ما سه مجموعه A_1 , A_2 , A_3 داریم که شامل موارد زیر است:

$$A_1 = \{1,2,3\}$$

$$A_2 = \{2,3,4\}$$

$$A_3 = \{3,4,5\}$$

محاسبه اندازه مجموعه ها:

$$|A_1| = 3$$

$$|A_2| = 3$$

$|A_3| = 3$

محاسبه تقاطع ها:

$|A_1 \cap A_2| = 2$ (عناصر 3,2)

$|A_1 \cap A_3| = 1$ (عناصر 33)

$|A_2 \cap A_3| = 2$ (عناصر 4,3)

$|A_1 \cap A_2 \cap A_3| = 1$ (عناصر 3)

استفاده از فرمول شمول و استثنا: حالا میخواهیم تعداد عناصر در مجموعه:

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|$$

بنابراین:

$$|A_1 \cup A_2 \cup A_3| = 3 + 3 + 3 - 2 - 1 - 2 + 1 = 5$$

در نتیجه: مجموعه کلی $A_1 \cup A_2 \cup A_3$ شامل موارد {1,2,3,4,5} است که متوجه شدیم ۵ عنصر دارد.

۳-۳-۶ آنالیز:

۱-۳-۳-۶ محدودیت‌ها: برای یافتن محدوده‌های بالا و پایین برای مقادیر مختلف استفاده می‌شود.

۱-۱-۳-۳-۶ محدوده های بالایی:

با استفاده از فرمول‌های نامساوی VC ، ما می‌توانیم حداکثر تعداد نمونه‌هایی که برای ایجاد یک مدل قابل اعتماد لازم است را تخمین بزنیم. مثلا در فرمول ذکر شده، با مشخص کردن ظرفیت VC و پارامترهای دقت و احتمال خطا، می‌توان یک حداکثر (upper bound) برای تعداد نمونه‌ها به دست آورد.

۲-۱-۳-۳-۶ محدوده های پایینی:

در برخی موارد، می‌توان استدلال کرد که برای تعمیم موفق، تعداد بیشتری از نمونه‌ها نیاز است و نمی‌توان یک حداقل (حدود پایین) تعریف کرد. هرچند نامساوی VC به طور عمده برای پیدا کردن حد بالا استفاده می‌شود، اما برخی تحلیل‌ها می‌توانند حداقل‌های تقریبی را نیز ارائه دهند.

مثال: فرض کنید که می‌خواهید یک مدل جدید بسازید و می‌خواهید بدانید چه تعداد داده برای آموزش آن لازم است. با استفاده از نامساوی VC، می‌توان شما را در یادگیری اینکه این تعداد در محدوده‌ای خصوصی و قابل قبول است، راهنمایی کند.

در زمینه نامساوی VC، ظرفیت VC یک مجموعه از توابع H به‌صورت زیر تعریف می‌شود:

اگر H ظرفیت VC برابر با d داشته باشد و n نقاط وجود داشته باشد، آنگاه حداقل تعداد نمونه‌ها برای اینکه بتوانیم با دقت تعیین کنیم یک مدل چگونه عمل می‌کند، به این صورت تخمین زده می‌شود:

$$N_H(\epsilon,\delta) \leq \left(\frac{2}{\epsilon}\right)^d . \mathrm{Ln}\left(\frac{4}{\delta}\right)$$

که در آن:

$NH(\epsilon,\delta)$ تعداد نمونه‌های مورد نیاز برای تعمیم خوب، ϵ میزان خطای قابل قبول در تعمیم، δ احتمال خطای غیرقابل قبول

توضیح اجزای فرمول

$\left(\frac{2}{\epsilon}\right)^d$ این جزء نشان می‌دهد که ظرفیت مدل به d و دقت ϵ وابسته است. یعنی هر چه ظرفیت مدل بیشتر باشد، به تعداد بیشتری داده برای تعمیم درست نیاز داریم.

$\mathrm{Ln}\left(\frac{4}{\delta}\right)$ این عبارت نمایانگر رابطه بین احتمال خطا و تعداد نمونه‌هاست. با کاهش احتمال خطا (افزایش δ)، تعداد نمونه‌ها به‌طور نمایی افزایش می‌یابد.

مثال: برای درک بهتر، بیایید یک مثال ساده فرض کنیم.فرض کنید که ما یک مدل یادگیری داریم که ظرفیت VC آن d=2 است. ما می‌خواهیم بدانیم که چقدر نمونه برای رسیدن به دقت ϵ=0.1 نیاز داریم و می‌خواهیم اطمینان δ=0.05 داشته باشیم.

محاسبه تعداد نمونه ها، با استفاده از فرمول:

۱. محاسبه $\left(\frac{2}{0.1}\right)^2$: ⟵ $\left(\frac{2}{0.1}\right)^2 = (20)2 = 400$

۲. محاسبه $\mathrm{Ln}\left(\frac{4}{0.05}\right)$: ⟵ $\frac{4}{0.05} = 80$ $\quad \ln(80) \sim 4.382$

۳. ترکیب نتیجه ها: ⟵ $N_H(0.1\,,\,0.05) \leq 400\ .\ 4.382 \sim 1752.8$

بنابراین، برای دستیابی به دقت ϵ=0.1 با احتمال خطا δ=0.05، به حداقل ۱۷۵۳ نمونه نیاز داریم.

این مثال نشان می‌دهد که وقتی ظرفیت VC یک مدل کم و ناپایدار باشد (مانند ظرفیت d=2 در مثال ما)، به تعداد قابل توجهی از نمونه‌ها نیاز است تا بتوانیم به دقت مورد نظر برسیم.

۲-۳-۳-۶ همگرایی: برای بررسی رفتار تابع رشد و احتمال خطا با افزایش تعداد داده‌های آموزشی استفاده می‌شود. همچنین همگرایی در زمینه نظریه یادگیری ماشین به مفهومی اشاره دارد که در آن یک الگوریتم یادگیری به تدریج به یک مدل یا تابع بهینه نزدیک می‌شود، به خصوص با اضافه شدن نمونه‌های بیشتر.

بیان قضیه:

اگر H یک خانواده توابع با ظرفیت VC برابر با d باشد و با n نمونه تصادفی از توزیع D آموزش ببینید، آنگاه با احتمالی حداقل ۱ $\delta-$:

$$|R(h) - R'(h)| \leq \in$$

که در آن : خطای حقیقی تابع R(h)، خطای تجربی R'(h)، حداکثر خطای قابل قبول ∈، احتمال وقوع خطا δ.

مراحل اثبات:

تعریف خطاهای تجربی و حقیقی ، خطاهای تجربی R'(h) به صورت زیر تعریف میشود:

$$R'(h) = ½ \sum_{i=1}^{n} | (h\ (xi) \neq y\ i)$$

خطاهای حقیقی R (h) به صورت زیر تعریف میشود:

$$R\ (h)\ E\ (x,y) \sim D\ [\ (h\ (x) \neq y)]$$

استفاده از نامساوی VC :با استفاده از نامساوی VC و فرمول محدودیت‌ها، می‌توان نشان داد که برای هر $\epsilon>0$ و $\delta>0$، تعداد نمونه‌های لازم به‌گونه‌ای انتخاب شده‌اند که تفاوت بین خطای تجربی و خطای حقیقی محدود باشد.

استفاده از اصل شمول و استثنا: برای تحلیل نحوه همگرایی، از اصل شمول و استثنا استفاده می‌شود تا تعداد حالاتی را که برای n نمونه می‌توان مشاهده کرد، ارزیابی کنیم. این می‌تواند تضمین کند که با افزایش n، خطای تجربی به‌تدریج به خطای واقعی نزدیک می‌شود

مثال: فرض کنید ما یک خانواده توابع خطی با ظرفیت VC برابر با $d=1$ داریم. به‌عبارت دیگر، ما می‌خواهیم یک خط راست بسازیم که مجموعه‌ای از نقاط را طبقه‌بندی کند.

شرایط اولیه: فرض کنید ۱۰ نمونه تصادفی از داده‌ها داریم. ما می‌خواهیم بدانیم چه تعداد نمونه (از خطا) برای رسیدن به دقت ۰/۰۵ با احتمال ۰/۰۱ نیاز داریم.

محاسبات: از نامساوی VC استفاده می‌کنیم:

$$N_H\ (\in ، \delta) \leq \frac{1}{\epsilon^2}(d \ln\left(\frac{2n}{d}\right) + \ln\left(\frac{4}{\delta}\right))$$

۱. محاسبه ϵ^2: $\epsilon^2 = (۰.۰۵)^2 = ۰.۰۰۲۵$

۲. محاسبه $\ln\left(\frac{2n}{d}\right)$: با توجه به اینکه n=10 و d=1 $\ln\left(\frac{2*10}{1}\right) = \ln(20) \sim 2.996$

۳. محاسبه $\ln\left(\frac{4}{\delta}\right)$ $\ln\left(\frac{4}{0.01}\right) = \ln(400) \sim 5.991$

۴. $\ln\left(\frac{4}{\delta}\right)$ جایگذاری در فرمول $N_H\ (0.05\ ,\ 0.01) \leq \frac{1}{0.0025}\ (1\ .\ \ 2.996 + 5.991)$

بدین معنا $N_H\,(0.05\ ,\ 0.01) \leq \frac{1}{0.0025}\ .\ 8.987 = 3594.8$

در نتیجه ، برای رسیدن به دقت ۰/۰۵ با احتمال انحراف ۰/۰۱ به حداقل ۳۵۹۵ نمونه نیاز دارد.

این اثبات به ما نشان می‌دهد که با افزایش نمونه‌ها و استفاده از ظرفیت VC، خطای تجربی به خطای واقعی نزدیک‌تر می‌شود. این همگرایی نشان‌دهنده این است که اگر داده‌های کافی وجود داشته باشد، مدل‌های یادگیری قادر به تعمیم و پیش‌بینی درست خواهند بود.

۳-۳-۶ تابع‌های تولیدکننده:

در برخی موارد، از تابع‌های تولیدکننده برای حل مسائل شمارشی استفاده می‌شود.توابع تولیدکننده (Generating functions) ابزار قدرتمندی در ریاضیات ترکیبی و احتمال هستند که به ما امکان می‌دهند دنباله‌های عددی را به صورت توابع تحلیلی نمایش دهیم. این نمایش، بسیاری از عملیات روی دنباله‌ها را به عملیات ساده‌تری روی توابع تبدیل می‌کند و اغلب در اثبات هویت‌ها و نامساوی‌ها بسیار مفید است.

۱-۳-۳-۶ مراحل استفاده از توابع تولیدکننده در اثبات نامساوی VC:

مدل‌سازی مشکلات: برای تحلیل یک مساله، برای مثال، تعداد مختلفی از نواها یا دنباله‌ها، ابتدا باید تابع تولید کننده مناسب را برای مساله تعریف کنیم. برای مثال، می‌توانیم در نظر بگیریم که a_n نشان دهنده تعداد روش‌های تقسیم یک دنباله به nدسته است.

تحلیل توابع تولیدکننده: با کار کردن روی تابع تولیدکننده، می‌توانیم خواص مختلف دنباله‌ها را به دست آوریم. مثلا، با مشتق گیری از تابع تولیدکننده و یا دیگر تکنیک‌ها می‌توانیم ویژگی‌های مرتبط با ترکیب‌پذیری را کشف کنیم.

استفاده از خواص ترکیبات: تابع تولیدکننده به ما اجازه می‌دهد که از خواص ترکیباتی برای استنتاج مقدار های خاص استفاده کنیم. برای نمونه، برای دلایل احتمال، می‌توان از توزیع باینومیال یا دوجمله‌ای استفاده کرد.

برقراری نامساوی VC: در نهایت، با به‌دست آوردن نتایج و اطلاعات کافی از توابع تولیدکننده (مانند محاسبه توزیع احتمال)، ما می‌توانیم به مناسبت یک نامساوی VCدست یابیم. به عنوان مثال، ممکن است این نامساوی به صورت زیر باشد:

که در آن Cیک ثابت است و d تعداد نقاط است.

$$\text{VC Dimension} \leq C\ .\ d$$

۱-۱-۳-۳-۶ مثال: فرض کنید ما یک دنباله X داریم که شامل ۴ نقطه$\{x1,x2,x3,x4\}$ است، و می‌خواهیم نشان دهیم که VC Dimension یک تابع طبقه‌بندی خاص از خانواده H برابر با ۳ است.

۲-۱-۳-۳-۶محاسبه تعداد زیرمجموعه‌ها: برای تعداد نقاط $n=4$ و با توجه به فرمول نامساوی Sauer، تعداد زیرمجموعه‌های قابل ایجاد به صورت زیر محاسبه می‌شود:

$$|H(X)| \leq \sum_{i=0}^{d} \binom{n}{I}$$

برای :d=3

$$|H(X)| \leq \binom{4}{0} + \binom{4}{1} + \binom{4}{2} + \binom{4}{3}$$

حساب کنیم:

$$\binom{4}{0} = 1$$

$$\binom{4}{1} = 4$$

$$\binom{4}{2} = 6$$

$$\binom{4}{3} = 4$$

حالا مجموع این مقادیر: $1+4+6+4=15$

پس ما میتوانیم بگوییم: $|H(X)| \leq 15$

۳-۱-۳-۳-۶استفاده از تابع تولیدکننده: حال می‌خواهیم از تابع تولیدکننده برای محاسبه تعداد توابع طبقه‌بندی استفاده کنیم. تابع تولیدکننده برای دنباله زیرمجموعه‌ها به صورت زیر است

$$G(x)=(1+x)^n$$

برای :n=4

$$G(x)=(1+x)^4 = 1 + 4x + 6x^2 + 4x^2 + 4x^3 + x4$$

در این تابع: ضریب x^3 نماینگر زیر مجموعه خالی است، ضریب x^1 نماین گر ۴ نقطه هم ردیف (یک نقطه جدا) است، ضریب x^2 نماین گر حالات ۲ نقطه ای از ۴ نقطه است، ضریب x^3 نماین گر حالات ۳ نقطه ای از ۴ نقطه است، ضریب x^4 نماین گر تمتم نقاط (همه نقطه ها) است.

۴-۳-۳-۶نتیجه‌گیری:

با این محاسبات، ما نشان دادیم که تعداد زیرمجموعه‌های قابل ایجاد از ۴ نقطه عددی (۴ نقطه در فضای ۲ بعدی) به کمک نامساوی VCو توابع تولیدکننده می‌تواند تعداد توابع طبقه‌بندی را در VC Dimensionبررسی کند. در این حالت، با توجه به VC Dimensionبرابر با ۳، مشخص می‌شود که تعداد توابع طبقه‌بندی قابل ایجاد با ۴ نقطه حداکثر ۱۵ است.

۵-۳-۳-۶ تفسیر:

این تحلیل نشان می‌دهد که علی‌رغم داشتن ۴ نقطه، نمی‌توان تمامی حالت‌های ممکن را در یک معماری ساده‌تر از ۳ متفاوت نشاندار (تشخیص) کرد. این نوع تحلیل به صورت خاص در یادگیری ماشین و مدل‌سازی‌های آماری به کار می‌رود و نشان‌دهنده قدرت توابع تولیدکننده در تحلیل و اثبات این نوع نامساوی‌ها است.

۴-۳-۶ جبر خطی:

۱-۴-۳-۶ فضاهای برداری: برای نمایش داده‌ها و مدل‌ها به صورت بردار استفاده می‌شود. فضای برداری مجموعه‌ای از نقاط (یا بردارها) است که تحت عملیات جمع و ضرب عددی بسته است. در یادگیری ماشین، ما می‌توانیم از فضای برداری برای توصیف مجموعه‌ای از داده‌ها و تصمیم‌گیری‌ها استفاده کنیم.

ویژگی‌ها (Features): در هر داده، ویژگی‌ها به عنوان ابعاد فضای برداری محسوب می‌شوند.

برچسب‌ها (Labels): خروجی که می‌خواهیم پیش‌بینی کنیم.

به طور کلی، ابعاد VCیک مجموعه از داده‌ها U، بیشترین اندازه‌ی مجموعه‌ای از نقاط $X \subseteq U$ را که می‌توان با یک مدل خاص (مانند یک مدل خطی) به طور کامل تفکیک کرد، به صورت زیر تعریف می‌شود

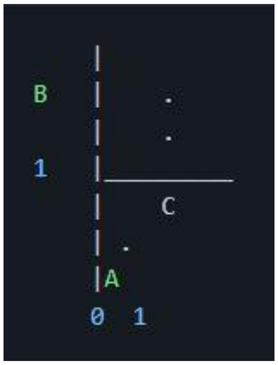
$$\text{VC Dimension} = \max\{m:\text{every labeling } X \text{ can be realized by the model}\}$$

مثال عددی، فرض کنیم، سه نقطه در فضای دو بعدی داریم:

A(0,0)

B(1,1)

C(1,0)



بررسی VC Dimension، محاسبه برچسب‌ها: فرض کنید می‌خواهیم دو برچسب +۱ و −۱ به این نقاط اختصاص دهیم. باید همه ترکیبات ممکن برچسب را بررسی کنیم.

شمارش ترکیبات: برای ۳ نقطه، تعداد ترکیبات ممکن برابر با $2^3=8$ است:

(+1,+1,+1)

(+1,+1,-1)

(+1,-1,+1)

(+1,-1,-1)

(-1,+1,+1)

(-1,+1,-1)

(-1,-1,+1)

(-1,-1,-1)

تفکیک خط: حالا باید بررسی کنیم که آیا می‌توان نقاط را به گونه‌ای برچسب‌گذاری کرد که برای همه ۸ ترکیب، تفکیک با یک خط ممکن باشد.

برای برچسب‌گذاری (+۱,−۱,−۱): را به عنوان +۱ و B وC را به عنوان −۱ در نظر می‌گیریم. می‌توان با کشیدن یک خط افقی در بین A وC این برچسب را تفکیک کرد. برای دیگر ترکیبات نیز می‌توان به همین صورت بررسی کرد و می‌بینیم که با دو نقطه A وB می‌توان به ترکیبات مختلف برچسب داد و تمام ویژگی‌ها را تفکیک کرد.

A(0,0):+1

B(1,1):-1

C(1,0):-1

نتیجه‌گیری: با بررسی ترکیب‌های مختلف برچسب‌ها و قابلیت تفکیک آن‌ها، می‌توانیم بگوییم که برای این سه نقطه، ابعاد VC برابر با ۳ است.

- VC Dimension = 3

نتیجه: VC Dimension نشان‌دهنده این است که یک مدل با چه تعداد نقاط می‌تواند به طور کامل تفکیک شود. در این مثال، VC Dimension آن مجموعه برابر با ۳ است زیرا می‌توانیم همه ترکیب‌های احتمالی برچسب‌ها را با استفاده از مدل‌های پیشنهادی (مانند خط) تفکیک کنیم.

۲-۴-۳-۶تبدیلات خطی: برای انجام عملیات‌های مختلف روی داده‌ها استفاده می‌شود.تبدیلات خطی عملگرهایی هستند که تابعی از یک فضای برداری را به فضای برداری دیگری می‌نگرند و دارای دو خاصیت زیر هستند:

جمع خطی:برای هر دو بردار u وv $\quad T(u+v) = T(u) + T(v)$

ضرب با عدد حقیقی:برای هر عدد حقیقی *c* و هر بردار u $\quad T(c.u) = c.T(u)$

به عنوان مثال، فرض کنید یک تبدیلات خطی به فرم: $T(x) = Ax$ که در آن A یک ماتریس $m×n$ است.

VC Dimension: یک مجموعه از نقاط S به صورت زیر تعریف می‌شود:

$$\text{VC Dimension}(S) = \max\{d:\text{every labeling } S \text{ can be realized}\}$$

۱-۲-۴-۳-۶مثال عددی:

یک مثال عددی را بررسی می کنیم که نشان می‌دهد چگونه می‌توان با استفاده از تبدیلات خطی VC Dimension کار کرد.

داده ها: فرض کنیم سه نقطه در فضای دو بعدی داریم:

A(0,0)

B(1,1)

C(1,0)

بررسی زیر مجموعه‌ها: برای هر یک از این نقاط می‌توان ۲ وضعیت ممکن (برچسب +۱ یا −۱) در نظر گرفت. بنابراین برای ۳ نقطه، تعداد ترکیب‌های ممکن برچسب‌ها برابر است با: $2^3=8$. برچسب های ممکن عبارتند از:

(+1,+1,+1)

(+1,+1,-1)

(+1,-1,+1)

(+1,-1,-1)

(-1,+1,+1)

(-1,+1,-1)

(-1,-1,+1)

(-1,-1,-1)

تحلیل با استفاده از تبدیلات خطی: حال بیایید یک ماتریس تبدیل خطی Aتصور کنیم:

$$A\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$$

که تغییر خاصی به وجود نمی‌آورد، و نقاط را به صورت زیر نمایش می‌دهد:

A(0,0)=(0,0)

B(1,1)=(1,1)

C(1,0)=(1,0)

حال فرض کنید که یک تبدیل جدید به شکل زیر داریم:

$$A‘=\begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}$$

وقتی این تبدیل بر نقاط اعمال می‌شود، نقاط به صورت زیر در می‌آید:

$$A'=A'\begin{bmatrix}0\\0\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$$

$$B'=A'\begin{bmatrix}1\\1\end{bmatrix}=\begin{bmatrix}2\\1\end{bmatrix}$$

$$C'=A'\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}1\\0\end{bmatrix}$$

بررسی تفکیک‌پذیری: برای این که قادر به برچسب‌گذاری این نقاط باشیم، می‌توانیم بررسی کنیم که آیا می‌توان این نقاط را تحت یک خط تقسیم کرد: اگر بخواهیم A'را برچسب +۱ انتخاب کنیم و بقیه را به −۱ برسانیم، خط تفکیک می‌تواند به شکل یک خط افقی در نظر گرفته شود، که توانسته ترکیب‌های برچسب‌ها را به صورت قابل تفکیک درآورد.

نتیجه‌گیری: با توجه به این مثال، ما می‌بینیم که با اعمال یک تبدیل خطی روی نقاط A،B ، و C، می‌توانیم همه ترکیب‌های برچسب‌ها را به طور مستقل برچسب‌گذاری کنیم. این نشان می‌دهد که VC Dimension برای این مجموعه نقاط برابر با ۳ است، زیرا توانسته‌ایم همه ترکیب‌ها را با استفاده از این نقاط تفکیک کنیم.

$$\text{VC Dimension} = 3$$

در نهایت: تبدیلات خطی به ما این امکان را می‌دهند که نقاط را به گونه‌ای جابجا و پردازش کنیم که قابلیت یادگیری و تفکیک مدل‌ها افزایش یابد. هرچند تعداد نقاط محدود است، اما سیستمی که فضای برداری را تست می‌کند، می‌تواند تا حد زیادی در فهم ظرفیت یادگیری ما اثرگذار باشد.

۵-۳-۶سایر تکنیک‌ها:

۱-۵-۳-۶ نظریه گراف: در برخی موارد، از نظریه گراف برای مدل‌سازی روابط بین داده‌ها استفاده می‌شود. نظریه گراف به مطالعه ساختارهای ریاضی به نام گراف می‌پردازد که شامل مجموعه‌ای از رئوس (نقاط) و یال‌ها (ارتباطات بین نقاط) است. به عبارت دیگر، یک گراف می‌تواند به صورت $G=(V,E)$ تعریف شود:

V :مجموعه رئوس، E :مجموعه یال‌ها که رابطه بین رئوس را نشان می‌دهد.

۱-۵-۳-۶ بررسی گراف‌ها به منظور اثبات VC Dimension :

به منظور بررسی ابعاد VCبا استفاده از نظریه گراف، می‌توانیم از گراف‌ها برای نمایش داده‌ها و روابط بین آن‌ها استفاده کنیم. هر گراف می‌تواند نشان‌دهنده‌ی نقاط و برچسب‌ها باشد و یک زیرگراف خاص می‌تواند برچسب‌ها را تفکیک کند.

۲-۵-۳-۶ ارتباط گراف VC Dimension

برای نشان دادن ارتباط بین گراف‌ها و ابعاد VC، می‌توانیم از گراف‌های دو حالته استفاده کنیم. در یک گراف دو حالته، رئوس به دو دسته تقسیم می‌شوند و یال‌ها تنها بین دسته‌های مختلف وجود دارند.

اگر یک مجموعه نقطه‌ای داشته باشیم که هر ترکیب محتمل برچسب‌ها قابل تفکیک باشد، می‌توانیم آن را به کمک یک گراف توصیف کنیم.

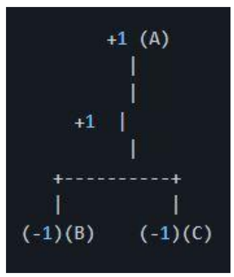
مثال، داده‌ها: فرض کنید سه نقطه داریم که می‌خواهیم آن‌ها را در یک گراف نمایش دهیم:

A (برچسب + 1)

B (برچسب - 1)

C (برچسب - 1)

این نقاط را میتوان به صورت زیر در یک گراف نمایش داد:



بررسی با استفاده از گراف: بر اساس این گراف، می‌توانیم نقاط را با استفاده از یال‌ها تفکیک کنیم. حال اگر بخواهیم برچسب‌ها را به شکل زیر در نظر بگیریم:

Aبا برچسب +۱

Bو Cبا برچسب −۱

با توجه به گراف می‌توانیم در نظر بگیریم که پروژه بر روی یک خط عوض می‌شود.

فرمول اعداد گراف: فرض کنید که می‌خواهیم ظرفیت یادگیری مدل را بر اساس تعداد رئوس nدر گراف بیان کنیم. برای نشان دادن این موضوع، تعداد ترکیب‌های برچسب ممکن برای nراس به صورت زیر نوشته می‌شود:

$$n^{۲}$$

که نشان‌دهنده‌ی تمامی ترکیب‌های ممکن برچسب‌ها برای nنقطه از داده‌ها است.

اثبات VC Dimension: اگر بتوانیم برای یک مجموعه از nنقطه، همه ترکیب‌های برچسب‌ها را تفکیک کنیم و یک گراف دو حالته بسازیم، بنابراین VC Dimensio آن مجموعه به صورت زیر خواهد بود:

VC Dimensio = n

نتیجه‌گیری: با استفاده از نظریه گراف و نمایش داده‌ها در قالب گراف، ما می‌توانیم به راحتی ارتباط بین نقاط و برچسب‌ها را تجزیه و تحلیل کنیم. این ارتباط نشان می‌دهد که چگونه می‌توانیم VC Dimensio را با استفاده از گراف‌ها توصیف کنیم و تعداد ترکیب‌های برچسب ممکن را محاسبه کنیم.

$$\text{VC Dimensio} = 3$$

در نهایت : گراف‌ها ابزارهای قدرتمندی هستند که به ما اجازه می‌دهند تا ساختار و روابط بین نقاط در مجموعه‌های داده را به روشی بصری و کمی بررسی کنیم. اگر نیاز به اطلاعات بیشتری دارید یا سوال‌های خاصی دارید، خوشحال می‌شوم کمک کنم!

۷- محاسبه vc dimension برای مدل های خطی در فضاهای مختلف:

محاسبه VC dimension برای مدل‌های خطی (مثل مدل‌های رگرسیون خطی یا دسته‌بندی خطی) در فضاهای مختلف (تعداد بعدها) می‌تواند بسته به بعد فضا متفاوت باشد. در اینجا به تفکیک VC dimension برای مدل‌های خطی در فضاهای مختلف پرداخته می‌شود:

۱-۷ فضای دو بعدی (R^2):

در فضای دو بعدی، ما می‌توانیم یک مدل خطی (خط) را تجسم کنیم.

نقاط: حداکثر سه نقطه می‌توانند به‌طور کامل توسط یک خط تفکیک شوند. سه نقطه ممکن است به‌گونه‌ای چیده شده باشند که خطی بتواند آن‌ها را به طور کامل تفکیک کند.

تحلیل:با یک نقطه، می‌توان آن را جدا کرد. با دو نقطه، می‌توان خطی گذاشت که آن‌ها را جدا کند.

با سه نقطه، هر آرایش عمومی از سه نقطه وجود دارد که می‌توانید با یک خط آن‌ها را جدا کنید.

با چهار نقطه، ممکن است نقاط به گونه‌ای بچینید که دیگر نتوانید خطی برای تفکیک کردن آن‌ها بگذارید (برای مثال، اگر چهار نقطه به شکل مستطیلی بچینید).

نتیجه: VC dimension مدل‌های خطی در R^2 برابر ۳ است.

۱-۱-۷ مثال عددی:

نقاط: بیایید فرض کنیم که دو نقطه داریم: $x_1=1$ و $x_2=2$

تحلیل:

۱ **نقطه**: x_1

۲ **نقطه**: x_1,x_2. می‌توانیم با یک خط (گذر از هر کدام) آن‌ها را تفکیک کنیم.

۳ **نقطه**: به‌عنوان مثال $x_1=1$, $x_2=2$, $x_3=3$, ما می‌توانیم برچسب‌ها را به شکل زیر تفکیک کنیم:

برچسب‌ها: (۰, ۱, ۱) → با خطی که از x_2 می‌گذرد.

```lua
|   1     2     3
|    o-----o-----o
|   0    1    0
|    -----------
```

نتیجه: VC dimension در فضای R^1 برابر ۲ است، زیرا با ۲ نقطه همیشه می‌توان آن‌ها را جدا کرد، اما با ۳ نقطه این تفکیک کار دشواری می‌شود.

۷-۲ فضای دو بعدی (R^2):

در فضای دو بعدی، ما می‌توانیم یک مدل خطی (خط) را تجسم کنیم.

نقاط: حداکثر سه نقطه می‌توانند به‌طور کامل توسط یک خط تفکیک شوند. سه نقطه ممکن است به‌گونه‌ای چیده شده باشند که خطی بتواند آن‌ها را به طور کامل تفکیک کند.

تحلیل:

با یک نقطه، می‌توان آن را جدا کرد. با دو نقطه، می‌توان خطی گذاشت که آن‌ها را جدا کند. با سه نقطه، هر آرایش عمومی از سه نقطه وجود دارد که می‌توانید با یک خط آن‌ها را جدا کنید. با چهار نقطه، ممکن است نقاط به گونه‌ای بچینید که دیگر نتوانید خطی برای تفکیک کردن آن‌ها بگذارید (برای مثال، اگر چهار نقطه به شکل مستطیلی بچینید).

نتیجه: VC dimension مدل‌های خطی در R^2 برابر ۳ است.

۱-۲-۷ مثال عددی:

نقاط: $P_1(1,1), P_2(2,2), P_3(3,1), P_4(2,0)$

تحلیل:

۱ **نقطه**: می‌توان به راحتی جدا شد.

۲ **نقطه**: می‌توان با یک خط آن‌ها را جدا کرد.

۳ **نقطه**: به‌طور کلی می‌توان تفکیک کرد، مانند:

(۱, ۱) و (۳, ۱) توسط خط عمودی یا با خطی عبور از (۲٬۲) تا (۲٬۰) انجام می‌پذیرد.

۴ **نقطه**: در صورتی که نقاط مانند شکل زیر باشند، می‌توان آن‌ها را هم جدا کرد:

نتیجه: VC dimension در فضای R_2 برابر ۳ است. نقاط می‌توانند تا ۳ تا با یک خط تفکیک شوند، اما با ارزیابی ۴ نقطه، این امکان وجود ندارد.

۷-۳ مدل‌های خطی در R^3 (فضای سه بعدی):

اینجا ما می‌توانیم نقاط را با یک صفحه تفکیک کنیم.**نقاط**:

برای ۱ **نقطه**: تفکیک‌پذیر.

برای ۲ **نقطه**: تفکیک‌پذیر.

برای ۳ **نقطه**: تفکیک‌پذیر.

برای ۴ **نقطه**: نقاط می‌توانند در یک صفحه تفکیک شوند.

برای ۵ **نقطه**: اگر پنج نقطه به گونه‌ای چیده شوند که هیچ چهار تا از آن‌ها در یک صفحه قرار نگیرند، نمی‌توان با یک صفحه آن‌ها را جدا کرد.

نتیجه: VC dimension در فضای R^3 برابر ۴ است.

۱-۳-۷ مثال عددی:

نقاط: $P_1(1,1,1), P_2(2,2,2), P_3(3,3,3), P_4(2,0,2)$

تحلیل:

۱ **نقطه**: جدا جدا.

۲ **نقطه**: با خط یا صفحه مشخص.

۳ **نقطه**: به‌طور کلی می‌توان آن‌ها را با یک صفحه جدا کرد.

۴ **نقطه**: در صورتی که نقاط مانند شکل زیر باشند، می‌توان آن‌ها را هم جدا کرد

```scss
      P3(3,3,3)
         *
        /|
       / |
      /  |
     *---*
    P1   P4 (2,0,2)
```

نتیجه: VC dimension در فضای R^3 برابر ۴ است. به‌طوری که حتی ۴ نقطه نیز می‌توانند با استفاده از یک صفحه جدا شوند، ولی با ۵ نقطه، این تفکیک ممکن است از بین برود.

۴-۷ فضای -*d* بعدی (R^d)

در فضای -*d* بعدی، مدل‌های خطی می‌توانند به‌صورت صفحات $(d-1)$ بعدی تجسم شوند.

نقاط:

با k نقطه که آرایش عمومی دارند، شما می‌توانید با $k-1$ درجه آزادی کار کنید.

برای d نقطه، می‌توانید سطحی از $d-1$ بعدی بسازید و آن‌ها را در d بعد جدا کنید.

با $d+1$ نقطه، اگر به‌طور غیرخطی یا متراکم چیده شوند، تفکیک‌پذیری آن‌ها ممکن است به دلیل قابلیت شکل‌دهی سطح‌ها از بین برود.

نتیجه: VC dimension مدل‌های خطی در R^d برابر d است.

جمع‌بندی:

VC dimension برای مدل‌های خطی در R^d برابر با **d** است. برای مثال:

R^1: VC dimension = 2

R^2: VC dimension = 3

R^3: VC dimension = 4

R^d: VC dimension = d

۵-۷ اهمیت VC Dimension برای مدل‌های خطی:

دانستن VC Dimension برای مدل‌های خطی به ما کمک می‌کند تا:

پیچیدگی مدل را ارزیابی کنیم: هرچه VC Dimension بالاتر باشد، مدل پیچیده‌تر است و احتمال اورفیت بیشتر می‌شود.

تعداد نمونه‌های آموزشی مورد نیاز را تخمین بزنیم: با استفاده از نامساوی VC می‌توانیم تعداد نمونه‌های آموزشی مورد نیاز برای رسیدن به یک سطح مشخص از دقت را تخمین بزنیم.

مدل مناسب را انتخاب کنیم: برای یک مسئله خاص، می‌توانیم مدلی با VC Dimension مناسب را انتخاب کنیم تا از یک طرف پیچیدگی مدل زیاد نباشد و از طرف دیگر دقت مدل نیز کاهش نیابد.

در نتیجه: محاسبه VC Dimension برای مدل‌های خطی نسبتا ساده است و به بعد فضای داده بستگی دارد. این مفهوم به ما کمک می‌کند تا درک بهتری از پیچیدگی مدل‌های خطی داشته باشیم و مدل‌های مناسب‌تری را برای مسائل مختلف انتخاب کنیم.

نکته: برای مدل‌های پیچیده‌تر مانند شبکه‌های عصبی، محاسبه VC Dimension بسیار دشوارتر است و روش‌های تقریبی برای تخمین آن استفاده می‌شود.

۸- محاسبه VC Dimension برای مدل های چند جمله ای در فضاهای مختلف:

در بخش‌های قبلی به بررسی VC Dimension برای مدل‌های خطی پرداختیم. در این بخش، به سراغ مدل‌های پیچیده‌تری می‌رویم که توانایی مدل‌سازی روابط غیرخطی را دارند:

۸-۱ مدل‌های چندجمله‌ای:

مدل‌های چندجمله‌ای به ما این امکان را می‌دهند تا با استفاده از ترکیبات مختلف توان‌های ویژگی‌ها، روابط پیچیده‌تری را بین ورودی و خروجی مدل کنیم.

۸-۲ VC Dimension در مدل‌های چندجمله‌ای:

VC Dimension یک مدل چندجمله‌ای به درجه d در فضای d بعدی، به طور کلی بسیار بزرگتر از VC Dimension یک مدل خطی است. این بدان معنی است که مدل‌های چندجمله‌ای پیچیده‌تر هستند و توانایی بیشتری برای یادگیری الگوهای پیچیده در داده‌ها دارند. اما چرا VC Dimension مدل‌های چندجمله‌ای بسیار بزرگتر است؟

انعطاف‌پذیری بیشتر: مدل‌های چندجمله‌ای با درجه بالاتر، انعطاف‌پذیری بیشتری برای تطبیق با داده‌های پیچیده دارند.

تعداد پارامترهای بیشتر: هرچه درجه چندجمله‌ای بالاتر باشد، تعداد پارامترهای مدل نیز بیشتر می‌شود که این امر به نوبه خود باعث افزایش پیچیدگی مدل می‌شود

۸-۳ محاسبه دقیق VC Dimension برای مدل‌های چندجمله‌ای:

محاسبه دقیق VC Dimension برای مدل‌های چندجمله‌ای به طور کلی بسیار پیچیده است و فرمول بسته‌ای برای آن وجود ندارد. با این حال، می‌توانیم برخی از موارد خاص را بررسی کنیم و مرزهای بالایی و پایینی برای VC Dimension ارائه دهیم.

مدل‌های چندجمله‌ای می توانند به صورت زیر تعریف شوند:

$$F(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

که در آن n درجه چند جمله ای و a_iضرایب آن هستند.

۸-۳-۱ فضای یک بعدی (R^1):

توضیحات: مدل‌های چند جمله‌ای در این فضا خطوط، منحنی‌های درجه یک، یا منحنی‌های درجه بالاتر را شامل می‌شوند.

۸-۳-۱-۱ مثال عددی:

نقاط: فرض کنید نقاط زیر را داریم:

$P_1(1), P_2(2), P_3(3), P_4(4)$با برچسب‌های (1,0,1,0) (به عنوان مثال).

تحلیل:

۱ نقطه: واضح است که می‌توان هر نقطه را جدا کرد.

۲ نقطه: به‌راحتی می‌توان آن‌ها را جدا کرد. با یک خط یا منحنی از درجه ۱.

۳ نقطه: یک منحنی از درجه ۲ می‌تواند این نقاط را به‌دست آورد.

۴ نقطه: حدوداً می‌توان با یک چند جمله‌ای از درجه ۳ آن‌ها را جدا کرد. برای مثال، نقاط در یک شکل مثلثی و با برچسب‌های خاص تفکیک‌شوند.

```scss
y
|        *
|   (0)  |   (1)
|        |
|     (2)|    *
| _____|_____|______ x
|        |
|   (3)  |       *
|        |
```

برچسب ها : ۱ (پایین)، 0 (بالا)

نتیجه: VC dimensionبرای مدل‌های چند جمله‌ای در R^1 برابر ۳ است. به این معنی که با ۳ نقطه می‌توان تمام حالات را تجزیه و تحلیل کرد، اما برای ۴ نقطه این کار ممکن نیست.

۸-۳-۲ فضای دو بعدی (R^2):

توضیحات:در این فضا می‌توانیم منحنی‌ها را در دو بعد خواهید داشت، که می‌تواند دایره، بیضی یا سایر منحنی‌ها باشد.

۸-۳-۲-۱ مثال عددی:

نقاط: فرض کنیم نقاط زیر را داریم:

$$P_1(1,1),P_2(2,3),P_3(3,1),P_4(4,4)$$

تحلیل:

۱ **نقطه:** می‌توان به راحتی آن را جدا کرد.

۲ **نقطه:** مثل قبل، می‌توان با یک خط آن‌ها را جدا کرد.

۳ **نقطه:** اینجا می‌توان با یک منحنی از درجه ۲ (مثل دایره) جدا کرد.

۴ **نقطه:** نقاط نمی‌توانند همیشه جدا شوند در صورتی که نقاط به شکل مثلثی یا یک دایره بچینند.

```scss
        (2,3)
          *          (4,4)
         /
        /
    *  /
    (1,1)--------(3,1)
```

نتیجه: VC dimension در فضای R^2 برای چند جمله‌ای های درجهٔ ۲ برابر ۳ است، زیرا می‌توان ۳ نقطه را به راحتی با یک منحنی تنش فراگیر جدا کرد.

۸-۳-۳ فضای d-بعدی (R^d):

توضیحات: برای مدل‌های چند جمله‌ای در فضای d-بعدی، هر اکسترم حداقل به اندازه n با جابجایی‌هایی که می‌توانند با هر درجه n تفکیک شوند، فعالیت می‌کند.

```
        (3,3,3)
          *
         /|
        / |
       /  |
(1,1,1)*---* (4,4,4)
       |   /
       |  /
(2,2,2)* /
          (4,3,2)
```

در اینجا یک صفحه میتواند نقاط P_1 , P_2 , P_3 , P_4 را تفکیک کرد.

۸-۳-۴ فضای d-بعدی (R^d):

توضیحات: برای مدل‌های چند جمله‌ای در فضای d-بعدی، هر اکسترم حداقل به اندازه n با جابجایی‌هایی که می‌توانند با هر درجه n تفکیک شوند، فعالیت می‌کند.

۸-۴ مرزهای بالایی و پایینی برای VC Dimension

مرز بالایی: VC Dimension یک مدل چندجمله‌ای درجه d در فضای d بعدی، از مرتبه $O(d^d)$ است. این مرز بسیار بزرگ است و نشان می‌دهد که مدل‌های چندجمله‌ای می‌توانند بسیار پیچیده باشند.

مرز پایینی: یافتن یک مرز پایینی دقیق برای VC Dimension مدل‌های چندجمله‌ای بسیار دشوار است.

۸-۵ چالش‌ها و راهکارها

اورفتینگ: به دلیل پیچیدگی بالای مدل‌های چندجمله‌ای، خطر اورفتینگ بسیار زیاد است.

تنظیم پارامترها: انتخاب درجه مناسب برای چندجمله‌ای بسیار مهم است. درجه خیلی کم ممکن است منجر به مدل‌های ساده‌ای شود که نتوانند الگوهای پیچیده را مدل کنند و درجه خیلی زیاد ممکن است منجر به اورفتینگ شود.

روش‌های کاهش پیچیدگی: برای کاهش پیچیدگی مدل‌های چندجمله‌ای، می‌توان از روش‌هایی مانند (Regularization) استفاده کرد.

۸-۶ نتیجه‌گیری

مدل‌های چندجمله‌ای به دلیل انعطاف‌پذیری بالایی که دارند، می‌توانند روابط پیچیده‌ای را در داده‌ها مدل کنند. با این حال، پیچیدگی بالای این مدل‌ها می‌تواند منجر به اورفتینگ شود. بنابراین، انتخاب درجه مناسب برای چندجمله‌ای و استفاده از روش‌های کاهش پیچیدگی بسیار مهم است.

۹- محاسبه vc dimension برای مدل های شبکه های عصبی در فضاهای مختلف:

VC Dimension در مدل‌های شبکه‌های عصبی به مراتب پیچیده‌تر از مدل‌های خطی یا چندجمله‌ای است. این به دلیل پیچیدگی ساختاری شبکه‌های عصبی، وجود لایه‌های پنهان و تنوع توابع فعال‌سازی است.

۹-۱ مفهوم VC Dimension در شبکه‌های عصبی:

VC dimension یک مقدار است که نشان می‌دهد یک مدل یادگیری چقدر می‌تواند داده‌ها را سازگار کرده و نقاط داده مختلف را تفکیک کند. برای شبکه‌های عصبی، پیش‌بینی VC dimension معمولاً به تعداد لایه‌ها، تعداد نورون‌ها در هر لایه، و معماری شبکه بستگی دارد.

۹-۲ شبکه‌های عصبی و VC Dimension:

۹-۲-۱ محاسبه VC Dimension:

VC dimension برای یک شبکه عصبی به‌طور تقریبی از فرمول زیر به‌دست می‌آید

$$\text{VC dimension} \sim O\left(n \cdot \ln(m)\right)$$

که در آن : تعداد ورودی ها (تعداد ویژگی ها)، m تعداد نورون ها در لایه های پنهان .

این به این معنی است که با افزایش تعداد نورون‌ها و ویژگی‌ها، ظرفیت شبکه ما برای تفکیک و یادگیری افزایش می‌یابد.

۲-۲-۹ مثال عددی:

فرضیات: فرض کنید یک شبکه عصبی ساده، با:

تعداد ورودی‌ها: $n=2$ (دو ورودی)

تعداد نورون‌های لایه پنهان: $m=3$ (سه نورون در لایه پنهان)

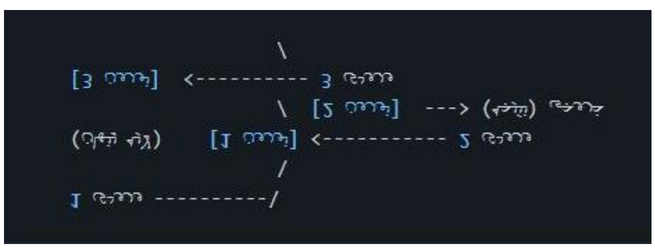
تعداد لایه‌ها: یک لایه پنهان و یک لایه خروجی

تحلیل VC Dimension: با توجه به فرمول above، ما داریم:

$$\text{VC Dimension} \sim O\,(2\ .\ 1.585) \approx 3.17$$

۳-۲-۹ نمودار:

در اینجا، ما می‌توانیم یک تصویر ساده از یک شبکه عصبی با ویژگی‌های ورودی و لایه مخفی و خروجی را رسم کنیم:



در این مثال، سه نورون در لایه پنهان می‌توانند ورودی‌هایی را دریافت کرده و به خروجی منتقل کنند. این توانایی آنها برای تفکیک داده‌ها و نحوه یادگیری کارکرد مدل را نشان می‌دهد.

۴-۲-۹ تحلیل ظرفیت یادگیری:

در حقیقت، ظرفیت VC dimension می‌تواند با تغییر در تعداد نورون‌ها و تعداد لایه‌ها به‌طور معنایی تغییر کند. برای مثال:

با اضافه کردن لایه‌های دیگر به شبکه (دو یا سه لایه پنهان)، VC dimension نیز به‌طور قابل توجهی افزایش می‌یابد.

برای یک شبکه عمیق و پیچیده، می‌توان VC dimension را حتی بیشتر افزایش داد.

باید توجه داشت که این محاسبات تقریباً تخمینی و مبتنی بر نظریه هستند و ممکن است در عمل به‌دلیل معماری‌های مختلف، نتایج متفاوتی حاصل شود.

۹-۳ چرا محاسبه VC Dimension برای شبکه‌های عصبی دشوار است؟

پیچیدگی ساختاری: شبکه‌های عصبی دارای ساختاری سلسله مراتبی هستند و هر لایه می‌تواند ویژگی‌های پیچیده‌تری را از داده‌ها استخراج کند.

توابع فعال‌سازی: انواع مختلف توابع فعال‌سازی (مانند ReLU، Sigmoid ، Tanh) باعث ایجاد غیرخطیگی در شبکه شده و محاسبه VC Dimension را دشوارتر می‌کند.

تعداد پارامترها: با افزایش تعداد لایه‌ها و نورون‌ها، تعداد پارامترهای شبکه نیز افزایش می‌یابد که این امر باعث افزایش پیچیدگی مدل می‌شود.

۹-۳-۱ مرزهای بالایی و پایینی:

با وجود این پیچیدگی‌ها، برخی مرزهای بالایی و پایینی برای VC Dimension شبکه‌های عصبی ارائه شده است. اما این مرزها اغلب بسیار محافظه‌کارانه هستند و به واقعیت نزدیک نیستند.

۹-۳-۲ روش‌های تقریبی:

به دلیل دشواری محاسبه دقیق VC Dimension برای شبکه‌های عصبی، از روش‌های تقریبی و تجربی استفاده می‌شود. برخی از این روش‌ها عبارتند از:

استفاده از مرزهای بالایی ساده شده: می‌توان از مرزهای بالایی ساده شده‌ای برای تخمین VC Dimension استفاده کرد. اما این مرزها ممکن است بسیار بزرگتر از VC Dimension واقعی باشند.

شبیه‌سازی: می‌توان با استفاده از شبیه‌سازی، تعداد نقاطی را که یک شبکه عصبی می‌تواند به طور کامل جدا کند، تخمین زد.

استفاده از معیارهای پیچیدگی دیگر: می‌توان از معیارهای پیچیدگی دیگری مانند تعداد پارامترها، تعداد لایه‌ها و تعداد نورون‌ها برای ارزیابی پیچیدگی شبکه استفاده کرد.

۹-۳-۳ مثال:

در نظر بگیرید یک شبکه عصبی ساده با یک لایه پنهان و یک تابع فعال‌سازی ReLU داریم. این شبکه می‌تواند توابع بسیار پیچیده‌ای را مدل کند. برای مثال، می‌تواند مرزهای تصمیم‌گیری بسیار پیچیده‌ای را در فضای ورودی ایجاد کند.

محاسبه دقیق VC Dimension این شبکه بسیار دشوار است. اما می‌توانیم بگوییم که VC Dimension این شبکه بسیار بزرگتر از VC Dimension یک مدل خطی با همان تعداد پارامتر است.

۹-۴ اهمیت VC Dimension در شبکه‌های عصبی:

انتخاب معماری شبکه: با دانستن اینکه VC Dimension بالا به معنای پیچیدگی بیشتر و احتمال اورفتینگ بالاتر است، می‌توانیم معماری شبکه را به گونه‌ای طراحی کنیم که پیچیدگی آن متناسب با حجم داده‌ها و مسئله مورد نظر باشد.

تنظیم پارامترهای شبکه: می‌توان از VC Dimension برای تنظیم پارامترهایی مانند تعداد لایه‌ها، تعداد نورون‌ها و نرخ یادگیری استفاده کرد.

مقایسه مدل‌ها: می‌توان از VC Dimension برای مقایسه پیچیدگی مدل‌های مختلف و انتخاب بهترین مدل استفاده کرد.

۵-۹ جمع‌بندی:

نسبت به مدل‌های خطی: شبکه‌های عصبی معمولاً دارای VC dimension بالاتری هستند که می‌تواند توانایی آن‌ها در یادگیری و تفکیک داده‌ها را به طرز چشمگیری افزایش دهد.

اجزای اصلی: تعداد ورودی‌ها، تعداد نورون‌ها در لایه پنهان و تعداد لایه‌ها، همگی بر VC dimension تأثیر می‌گذارند.

محاسبه VC Dimension برای شبکه‌های عصبی یک چالش بزرگ است. با این حال، درک این مفهوم به ما کمک می‌کند تا پیچیدگی شبکه‌های عصبی را بهتر درک کنیم و مدل‌های بهتری را طراحی کنیم.

۱۰-۹ مثال بررسی چالش های محاسبه برای مدل شبکه عصبی و معرفی مفاهیم موثر (Effective VC Dimension) :

مثال: برای بررسی چالش‌های محاسبه VC Dimension در مدل‌های پیچیده‌تر مانند شبکه‌های عصبی، یک مثال عمیق و عددی را بررسی می‌کنیم. در این مثال، ما یک **شبکه عصبی عمیق** با چندین لایه و خصوصیات متنوع را بررسی خواهیم کرد که چالش‌های محاسبه VC Dimension را به وضوح نشان می‌دهد.

مثال:شبکه عصبی عمیق با ۳ لایه پنهان.

مشخصات شبکه:

لایه ورودی: تعداد نرون ها : ۳

ورودی ها: (X1,X2,X3)

لایه پنهان اول: تعداد نرون ها :۵

لایه پنهان دوم: تعداد نرون ها :۴

لایه پنهان سوم: تعداد نرون ها :۳

لایه خروجی: تعداد نرون ها :۱

خروجی باینری: (۰ یا ۱)

بین لایه پنهان دوم و لایه پنهان سوم

تعداد نرون های لایه پنهان دوم : ۴

تعداد نرون های لایه پنهان سوم :۳

تعداد وزن ها: ۱۲=۳*۴

بین لایه پنهان سوم و لایه خروجی

تعداد نرون های لایه پنهان سوم:۳

تعداد نرون های لایه خروجی:۱

تعداد وزن ها: ۳=۱*۳

جمع کل وزن ها

۵۰=۱۵+۲۰+۱۲+۳=جمع وزن ها

محاسبات وزن ها

۱.تعداد وزن ها:

وزن ها بین لایه ها به شرح زیر محاسبه می شوند

بین لایه ورودی و لایه پنهان اول:

تعداد نرون های لایه ورودی:۳

تعداد نرون های لایه پنهان اول:۵

۱۵=۳*۵ تعداد وزن ها

بین لایه پنهان اول و لایه پنهان دوم

تعداد نرون های لایه پنهان اول:۵

تعداد نرون های لایه پنهان دوم:۴

۲۰=۵*۴ : تعداد وزن ها

پارامترهای شبکه ۲:

نرون ها (Neurons)

ورودی ها: ۳ تا نرون ورودی

لایه پنهان اول:۵ نرون

لایه پنهان دوم:۴ نرون

لایه پنهان سوم:۳ نرون

لایه خروجی:۱نرون

وزن ها (Weights):

وزن که در فرآیند یادگیری مدل تنظیم میشوند ۵۰

بایاس ها (Biases):

فرض میکنیم که به هر نرون یک بایاس اضافی وجود دارد

لایه پنهان اول:۵ بایاس

لایه پنهان دوم:۴ بایاس

لایه پنهان سوم:۳ بایاس

لایه خروجی: ۱ بایاس

۱۳=۵+۴+۳+۱ : جمع بایاس ها

۶۳=(بایاس ها) ۱۳+(وزن ها) ۵۰ : در نتیجه ، جمع کل پارمترها

۹-۱۱ محاسبه VC Dimension:

تخمینی داریم که به شکل زیر است VC Dimension در محاسبه:

$$\text{VC Dimension} \approx O(N \,.\, \text{Log}(n)) + d$$

که در آن :تعداد نرون = N، تعداد وزن ها = d، تعداد نرون ها : (N)

N = 3 (ورودی) + 5 (پنهان اول) +4 (پنهان دوم) +3(پنهان سوم) + 1 (خروجی) = 16

تعداد وزن ها (d): d=50

۹-۱۲ تخمین VC Dimension:

$$\text{VC Dimension} \approx O(16 \,.\, \text{Log}(16)) + 50$$

محاسبه Log(16)

$$\text{VC Dimension} \approx O(16 \,.\, 4\,) + 50 = O(64) + 50$$

بنابراین

$$\text{VC Dimension} \approx O(16 \,.\, 4\,) + 50 = O(64) + 50$$

این مثال نشان می‌دهد که محاسبه VC Dimensionدر شبکه‌های عصبی عمیق به دلیل تعدد پارامترها و ظرفیت یادگیری بالا بسیار پیچیده است. لازم است از تخمین‌های تجربی و معیارهای مختلف برای ارزیابی مدل استفاده شود، چرا که VC Dimensionبه تنهایی نمی‌تواند تصویر دقیقی از کیفیت و کارایی مدل‌های عصبی عمیق ارائه دهد.

در مثال ارائه‌شده، VC Dimension به عنوان یک معیار برای اندازه‌گیری ظرفیت یادگیری مدل، به ما نشان می‌دهد که یک مدل چقدر می‌تواند بر روی داده‌های مختلف عمومی‌سازی کند. در واقع، VC Dimensionمعیاری برای تعیین قدرت تعمیم‌پذیری یک مدل یادگیری ماشین است.

۹-۱۲-۱ VC Dimensionمؤثر در این مثال:

برای شبکه عصبی عمیق که توضیح داده شد، با توجه به پارامترهای زیر:

(تعداد کل نرون ها در لایه های مختلف) (n) تعداد نرون ها:۱۶

(وزن های متصل به نرون و بایاس ها) (d) تعداد وزن ها : ۵۰

۹-۱۳تخمین VC Dimension:

محاسبات برای تخمین VC Dimensionبه شکل زیر است:

$$\text{VC Dimension} \approx O\ (N \,.\log(N)) + d$$

محاسبه $O\ (N\,.\log(N))$

محاسبه $\log(16))$ $\quad \log(۱۶) = ۴$

محاسبه $0\ (16\ .\ 4)$ $\quad 0\ (16\ .\ 4) = O(64)$

<u>افزودن تعداد وزن ها</u>: بنابر این VC Demensionتخمین کلی به شکل مقابل است

$$\text{VC Dimension} \approx O\ (64) + 50$$

نتیجه:نتیجه نهایی این تخمین به ما می‌گوید که VC Dimensionمؤثر برای این شبکه شاید چیزی در حدود:

VC Dimension≈ ۱۱۴

این به این معنی است که مدل می‌تواند تقریباً ۱۱۴ نمونه مختلف از داده‌ها را بدون خطا تقسیم‌کند. به عبارتی دیگر، اگر VC Dimensionمدل بیشتر از تعداد داده‌های آموزشی باشد، احتمال بالایی وجود دارد که مدل بر روی آن‌ها بیش‌برازش شده و نتواند به درستی داده‌های جدیدی که تاکنون ندیده را پیش‌بینی کند.

۹-۱۴ اهمیت VC Dimensionمؤثر:

تحلیل قابلیت تعمیم: VC Dimensionبه ما کمک می‌کند تا درک کنیم که آیا مدل توانایی یادگیری الگوهای پیچیده و تعمیم به داده‌های جدید را دارد یا خیر.

اجتناب از بیش‌برازش: وقتی VC Dimensionبالا می‌رود، خطر بیش‌برازش نیز افزایش می‌یابد. بنابراین، لازم است هنگامی که VC Dimension از تعداد نمونه‌های آموزشی بیشتر شود، مدلی با ظرفیت کمتر اتخاذ کنیم.

طراحی بهینه مدل: با استفاده از اطلاعات VC Dimension ، طراحان مدل می‌توانند تصمیم بگیرند که آیا مدل طراحی‌شده برای مسئله خاصی مناسب است یا خیر و بهینه‌سازی‌های لازم را انجام دهند.

۱۰- چالش های کلی VC Dimension:

۱۰-۱ محاسبات پیچیده:

محاسبه برای مدل‌های پیچیده: تعیین VC Dimensionبرای مدل‌های پیچیده مانند شبکه‌های عصبی عمیق دشوار است و ممکن است نیاز به روش‌های تحلیلی و شبیه‌سازی پیچیده داشته باشد.

مسئله NP-hard: برای بسیاری از کلاس‌های توابع، محاسبه دقیق VC Dimensionمی‌تواند به مشکلات NP-hardتبدیل شود.

۱۰-۲ عدم ارتباط مستقیم با تعمیم‌پذیری:

تبعیض بین اندازه‌گیری و عملکرد واقعی: VC Dimensionممکن است نشان‌دهنده قابلیت تعمیم یک مدل نباشد. یک مدل با VC Dimensionبالا می‌تواند به راحتی بر روی داده‌های آموزشی بیش‌برازش شود و نتواند به درستی بر روی داده‌های جدید عمل کند.

یادگیری غیرخطی: در مورد مدل‌های غیرخطی، افزایش VC Dimensionمی‌تواند همزمان با افت کیفیت تعمیم همراه باشد.

۱۰-۳ عدم دقت در ارزیابی کیفیت مدل:

تأثیر پارامترهای مختلف: VC Dimensionبه صورت یک دستی آن را به پارامترهای مدل نسبت می‌دهد، اما تأثیر تکنیک‌های یادگیری مانند regularizationیا dropoutمعمولاً قابل محاسبه در این زمینه نیست.

کلاسی خاص: VC Dimensionممکن است برای برخی از کلاس‌های داده یا ویژگی‌ها به خوبی عمل نکند و تحلیل‌های کم‌وزن کند.

۱۰-۴ تفاوت در وضعیت‌های مختلف داده:

داده‌های توزیع‌شده نامتعادل: در حالات داده‌های نامتعادل، VC Dimensionنمی‌تواند تضمین کند که مدل به طور مؤثر تعمیم می‌یابد.

تطبیق مدل با تغییرات داده: VC Dimensionثابت نیست و می‌تواند برای داده‌های جدید یا حوادث ناگهانی تغییر کند.

۱۰-۵ تخمین نادرست:

تفاوت بین VC Dimension وEffective VC Dimension: VC Dimension نمی‌تواند به تنهایی از ظرفیت واقعی مدل در یادگیری و تعمیم اطلاعات دقیق بدهد و برای نتایج دقیق‌تر به استفاده از Effective VC Dimensionنیاز داریم.

عدم انطباق با مدل‌های واقعی: ممکن است VC Dimensionبا برخی از مدل‌های یادگیری به صورت غیرقابل پیش‌بینی عمل کند و تحلیل‌های انجام‌شده دائما نیاز به ارزیابی مجدد دارند.

۱۰-۶ پیچیدگی در پیاده‌سازی:

پشتیبانی از مدل‌های چندمنظوره: در مدل‌های پیچیده‌ای که شامل چندین وظیفه یادگیری multi-task (learning) هستند، VC Dimensionبه تنهایی نمی‌تواند قابلیتی برای تحلیل آنها فراهم کند.

همپوشانی و وابستگی ویژگی‌ها: در مواقعی که ویژگی‌ها با یکدیگر همپوشانی دارند، محاسبه VC Dimensionمی‌تواند به صورت غیرقابل پیش‌بینی باشد.

۱۱- کاربرد های کلی VC Dimension

۱۱-۱ تحلیل قدرت تعمیم:

VC Dimensionبه ما کمک می‌کند تا بفهمیم یک مدل چقدر می‌تواند به طور مؤثر بر روی داده‌های جدید عمل کند. هرچه VC Dimensionیک مدل بیشتر باشد، به طور نظری ظرفیت یادگیری آن بیشتر است، و بنابراین می‌تواند الگوهای پیچیده‌تری را شناسایی کند.

مثال: فرض کنید ما دو مدل داریم: یک مدل خطی و یک مدل غیرخطی با پیکسل‌های اضافی. VC Dimensionمدل غیرخطی بیشتر است، بنابراین این مدل احتمال بیشتری برای یادگیری ویژگی‌های پیچیده در داده‌های آموزشی دارد. اگر این مدل به درستی روی داده‌های جدید تعمیم نیابد، ممکن است به بیش‌برازش مبتلا شود.

۱۱-۲ اندازه‌گیری ظرفیت مدل:

VC Dimensionنمایانگر تعداد بالاترین نقاطی است که یک مدل می‌تواند به طور کامل تفکیک کند. این معیار به ما کمک می‌کند تا بفهمیم یک مدل چقدر داده‌های مختلف را می‌تواند پردازش کند.

مثال: فرض کنید که شما در حال کار بر روی یک الگوریتم طبقه‌بندی تصویر برای شناسایی سگ‌ها و گربه‌ها هستید. اگر دقیقاً دو ویژگی (مثلاً رنگ و اندازه) برای شناسایی داشته باشید، VC Dimension می‌تواند نشان دهد که آیا این ویژگی‌ها کافی هستند یا نیاز به ویژگی‌های بیشتری است. اگر VC Dimension بیش از دو باشد، این نشان می‌دهد که شما نیاز به ویژگی‌های بیشتری دارید تا به طور کامل دو دسته را تفکیک کنید.

۱۱-۳ راهنمایی در انتخاب مدل:

با مقایسه VC Dimensionمدل‌های مختلف، می‌توانیم در انتخاب مدل مناسب برای یک مسئله خاص بهتر تصمیم‌گیری کنیم. مدل‌هایی با VC Dimensionمناسب برای یک نوع مسأله ممکن است عملکرد بهتری داشته باشند.

مثال: شما دو مدل داریم: یک درخت تصمیم و یک شبکه عصبی. اگر VC Dimensionدرخت تصمیم به طور معقول در مقایسه با شبکه عصبی باشد، ممکن است درخت تصمیم برای داده‌های ساده‌تر بهتر عمل کند و در نتیجه، انتخاب بهتری برای آن نوع داده‌ها باشد.

۱۱-۴ تحلیل پیچیدگی محاسباتی:

VC Dimensionمی‌تواند نمایانگر پیچیدگی محاسباتی یک مدل باشد. الگوریتم‌هایی با VC Dimension بالا معمولاً به زمان بیشتری برای یادگیری و پیش‌بینی نیاز دارند.

مثال: در هنگام مدل‌سازی برای یک مشکل خاص، اگر یک الگوریتم خاص VC Dimensionبالایی داشته باشد، ممکن است نیاز به داده‌های بیشتری برای جلوگیری از بیش‌برازش و تأمین زمان محاسباتی برای یادگیری و پیش‌بینی دقیق داشته باشد.

۱۱-۵ توسعه الگوریتم‌های جدید:

در طراحی و توسعه الگوریتم‌های جدید، دانش از VC Dimensionمی‌تواند به ما کمک کند تا از ایجاد الگوریتم‌هایی که ظرفیت یادگیری آن‌ها بیش از حد بالا است (و منجر به بیش‌برازش می‌شود) جلوگیری کنیم.

مثال: اگر محققان در حال طراحی یک الگوریتم جدید برای دسته‌بندی متون هستند، می‌توانند از VC Dimension برای ارزیابی نوع ویژگی‌هایی که باید استفاده کنند، بهره‌برداری کنند. اگر VC Dimension الگوریتم خیلی بالاست، می‌توانند ویژگی‌های اضافی را حذف کنند تا خطر بیش‌برازش را کاهش دهند.

۱۱-۶ مطالعه نظریه یادگیری:

VC Dimensionبه عنوان یک معیار برای اندازه‌گیری اثرات طبقه‌بندی و یادگیری از داده‌ها مورد استفاده قرار می‌گیرد. این معیار به ما کمک می‌کند تا درک بهتری از اصول بنیادی یادگیری ماشین پیدا کنیم.

مثال: یک مطالعه نظری ممکن است بررسی کند که چگونه VC Dimensionبر روی قابلیت تعمیم یک مدل خاص تأثیر می‌گذارد و چه نوع مدل‌هایی با VC Dimension مختلف می‌توانند در شرایط خاص به خوبی عمل کنند.

۱۱-۷ ارزیابی مدل‌های یادگیری عمیق:

در یادگیری عمیق، VC Dimensionمی‌تواند به عنوان یک شاخص اصلی برای سنجش ظرفیت و قدرت یادگیری شبکه‌های عصبی استفاده شود.

مثال: یک کارشناس یادگیری عمیق می‌تواند مدل‌های عمیق مختلف را بر اساس VC Dimensionآنها ارزیابی کند تا بفهمد کدام مدل توانایی بیشتری در یادگیری خاصیت‌های پیچیده داده‌ها دارد.

۱۱-۸ تحلیل منبع معتبر (Confidence Bounds):

در برخی از کاربردها، مثل تشخیص ناهنجاری (anomaly detection)یا بخش‌بندی داده‌های پزشکی، VC Dimension می‌تواند به تعیین محدوده‌های اعتماد برای پیش‌بینی‌های یک مدل کمک کند.

مثال: اگر در حال کار بر روی یک مدل برای شناسایی بیماری‌ها هستید، می‌توانید با استفاده از VC Dimension و ترکیب آن با دیگر معیارها، اطمینان بیشتری از قابلیت‌های پیش‌بینی مدل خود به دست آورید.

۱۲- استفاده از VC Dimensionدر انتخاب مدل های یادگیری ماشین مقایسه بین مدل های مختلف رگرسیون

استفاده از VC Dimensionبرای انتخاب مدل‌های یادگیری ماشین، به ویژه در مقایسه بین مدل‌های مختلف رگرسیون، می‌تواند به ما کمک کند تا بهترین مدل را برای داده‌ها و مسئله‌ای که با آن مواجه هستیم انتخاب کنیم.

۱۲-۱ مفهوم VC Dimensionدر رگرسیون:

در زمینه رگرسیون، مدل‌ها می‌توانند به شکل‌های مختلفی تعریف شوند، از جمله:رگرسیون خطی، رگرسیون چندجمله‌ای، رگرسیون درخت تصمیم.

۱۲-۲ مراحل انتخاب مدل با استفاده از VC Dimension:

مسیر انتخاب مدل با توجه به VC Dimensionمی‌تواند به صورت زیر باشد:

شناسایی مدل‌های بالقوه: مدل‌های مختلف مورد نظر (به عنوان مثال، رگرسیون خطی و رگرسیون چندجمله‌ای) را شناسایی کنید.

محاسبه VC Dimension: VC Dimensionهر مدل را محاسبه یا تخمین بزنید. هرچه VC Dimension بالاتر باشد، ظرفیت مدل برای یادگیری الگوهای پیچیده بیشتر است.

مقایسه ظرفیت و خطر بیش‌برازش: ظرفیت یادگیری بیشتر می‌تواند به خودی خود نشانه‌ای از توانایی مدل باشد، اما همچنین خطر بیش‌برازش را نیز باید در نظر گرفت. مدل‌هایی با VC Dimension بالا ممکن است به بیش‌برازش نسبت به داده‌های آموزشی دچار شوند.

انتخاب مدل مناسب: بر اساس تحلیل VC Dimensionو نیازمندی‌های خاص داده (مانند پیچیدگی و ویژگی‌ها)، بهترین مدل را انتخاب کنید.

۱۲-۳ مثال:

فرض کنید که ما داده‌هایی داریم که ارتباط بین قد و وزن افراد را نشان می‌دهد و می‌خواهیم مدلی برای پیش‌بینی وزن بر اساس قد پیدا کنیم.

۱۲-۳-۱ مدل‌های مورد نظر:

رگرسیون خطی: مدل ساده‌ای که رابطه‌ی خطی بین کمیت‌ها را تصور می‌کند.

رگرسیون چندجمله‌ای: مدلی که به شما اجازه می‌دهد تا رابطه‌های غیرخطی را بین قد و وزن مدل کنید.

رگرسیون درخت تصمیم: مدلی که می‌تواند تصمیمات را بر اساس ویژگی‌های مختلف (مانند سن، جنسیت و غیره) بگیرد و یک ساختار درختی ایجاد کند.

۱۲-۳-۲ محاسبه VC Dimension:

رگرسیون خطی: VC Dimensionاین مدل برابر ۲ است (زیرا می‌توان تنها دو نقطه را به طور مستقل تفکیک کرد).

رگرسیون چندجمله‌ای: فرض کنید از یک چندجمله‌ای درجه ۳ استفاده می‌کنیم، VC Dimension این مدل معمولاً ۴ است.

رگرسیون درخت تصمیم: VC Dimensionیک درخت تصمیم بسته به عمق و پیچیدگی درخت می‌تواند متفاوت باشد، اما معمولاً این مقدار بسیار بالاتر است (در اینجا فرض کنیم VC Dimension برابر ۱۰ است).

۱۲-۳-۳ مقایسه و تحلیل:

رگرسیون خطی : (VC = 2) این مدل ممکن است نتواند پیچیدگی‌های رابطه بین قد و وزن را به درستی مدل کند و ممکن است تحت‌برازش باشد.

رگرسیون چندجمله‌ای :(VC = 4)این مدل می‌تواند به طور متوسط پیچیدگی داده‌ها را به خوبی برآورده کند و خطرات کمتری از بیش‌برازش داشته باشد.

رگرسیون درخت تصمیم :(VC = 10)اگرچه این مدل می‌تواند اطلاعات بیشتری را در خود ذخیره کند، اما به خطر بیش‌برازش بالاتری مبتلا خواهد شد، به ویژه اگر داده‌های کمی داشته باشیم.

۱۲-۳-۴ انتخاب نهایی:

با توجه به تحلیل VC Dimensionو درجه پیچیدگی داده‌ها، می‌توانیم انتخاب کنیم که آیا رگرسیون چندجمله‌ای (در اینجا انتخاب مناسب) خوب عمل می‌کند یا رگرسیون درخت تصمیم نیاز به تنظیم بیشتری دارد.

۱۳- تنظیم پارامتر های پیچیده در VC Dimension، انتخاب تعداد لایه ها و نورون ها در شبکه های عصبی:

تنظیم پارامترهای پیچیده در مدلسازی با استفاده از VC Dimensionمی‌تواند به بهبود عملکرد و تعمیم‌پذیری مدل‌های یادگیری ماشین، به ویژه در شبکه‌های عصبی کمک کند. در اینجا به توضیح جزئیات و نحوه تاثیر VC Dimensionبر انتخاب تعداد لایه‌ها و نورون‌ها در شبکه‌های عصبی می‌پردازیم.

۱۳-۱ مفهوم VC Dimensionدر شبکه‌های عصبی

VC Dimensionبه ما کمک می‌کند تا درک کنیم یک شبکه عصبی چه تعداد از الگوها را می‌تواند یاد بگیرد و به چه اندازه می‌تواند به داده‌های جدید تعمیم یابد. به طور خاص:

۱-۱-۱۳ تعداد نورون‌ها: هیچ محدودیت دقیقی بر روی VC Dimensionبر اساس تعداد نورون‌ها وجود ندارد، اما به طور کلی، افزایش تعداد نورون‌ها باعث افزایش VC Dimensionمی‌شود. (زیرا مدل می‌تواند پیچیدگی بیشتری را یاد بگیرد).

۲-۱-۱۳ تعداد لایه‌ها: افزودن لایه‌های متعدد به شبکه عصبی معمولاً VC Dimensionرا افزایش می‌دهد، چون هر لایه می‌تواند ویژگی‌های مختلف داده‌ها را استخراج کند و از آن‌ها برای ساخت تصمیمات پیچیده استفاده کند.

۲-۱۳ تأثیر پارامترها بر VC Dimension:

۱-۲-۱۳ تعداد نورون‌ها:

افزایش تعداد نورون‌ها در یک لایه خاص می‌تواند ظرفیت مدل را برای یادگیری الگوهای مختلف افزایش دهد. با این حال، تعداد زیاد نورون‌ها می‌تواند به **بیش‌برازش** منجر شود، جایی که مدل به خوبی بر روی داده‌های آموزشی عمل می‌کند اما عملکرد ضعیفی بر روی داده‌های جدید دارد.

۱-۱-۲-۱۳ مثال:

فرض کنید یک شبکه عصبی با یک لایه ورودی، یک لایه پنهان با ۵ نورون و یک لایه خروجی داریم. VC Dimensionاین مدل نسبتا پایین است و ممکن است تنها قادر به شناسایی الگوهای ساده باشد.

اگر این لایه پنهان به ۲۰ نورون افزایش یابد، VC Dimensionافزایش می‌یابد و مدل احتمالاً می‌تواند نظریه‌های پیچیده‌تری را یاد بگیرد.

۲-۲-۱۳ تعداد لایه‌ها

با افزایش تعداد لایه‌ها در شبکه neuronal، ظرفیت یادگیری مدل به طور قابل توجهی افزایش می‌یابد و VC Dimensionمعمولاً افزایش می‌یابد. با این حال، این افزایش نیز خطر بیش‌برازش را به دنبال خواهد داشت.

۱-۲-۲-۱۳مثال:

فرض کنید یک شبکه عصبی ساده با ۲ لایه پنهان تشکیل شده است. این شبکه می‌تواند برخی از ویژگی‌های پیچیده را یاد بگیرد، اما VC Dimensionمتوسطی دارد.

اگر تعداد لایه‌ها به ۵ افزایش یابد، این مدل می‌تواند الگوهای بسیار پیچیده‌تری را شناسایی کند و VC Dimensionبه شدت افزایش می‌یابد. اما به همین اندازه، ممکن است این شبکه به داده‌های آموزشی بیش‌برازش پیدا کند.

۳-۱۳ انتخاب پارامترها با توجه به VC Dimension:

برای انتخاب تعداد لایه‌ها و نورون‌ها، می‌توان مراحل زیر را دنبال کرد:

۱-۳-۱۳ آزمایش مدل‌های مختلف:

برای مقایسه مدلسازی، می‌توان مدل‌هایی با تعداد نورون‌ها و لایه‌های متفاوت را آزمایش کرد و مقدار VC Dimensionآن‌ها را محاسبه کرد.

۲-۳-۱۳ استفاده از اعتبارسنجی متقابل (Cross-Validation):

با استفاده از اعتبارسنجی متقابل، می‌توان کارایی هر یک از مدل‌ها را مورد بررسی قرار داد و مشخص کرد کدام مدل با تنظیمات مشخص شده از نظر VC Dimensionبه خوبی بر داده‌های جدید تعمیم می‌یابد.

۳-۳-۱۳ شناسایی تعادل:

هدف پیدا کردن تعادل بین VC Dimensionو عملکرد مدل است. یک مدل بسیار پیچیده‌تر می‌تواند نتایج خوبی بر روی داده‌های آموزشی نشان دهد، اما اگر قدرت تعمیم کافی نداشته باشد، بر روی داده‌های جدید ضعیف خواهد بود.

۴-۱۳مثال: شبکه عصبی برای پیش‌بینی قیمت خانه:

مشخصات پروژه: ما در حال توسعه یک مدل یادگیری ماشین برای پیش‌بینی قیمت خانه‌ها بر اساس ویژگی‌هایی مانند متراژ، تعداد اتاق‌ها، سن خانه و موقعیت قرارگیری هستیم.

آرایش شبکه عصبی: ما سه مدل مختلف از شبکه‌های عصبی را طراحی کردیم و مقادیر مربوط به هرکدام را بررسی می‌کنیم

ساده ترین مدل A

تعداد لایه ها: ۱ لایه پنهان

تعداد نورون ها : ۵ نورون

تابع فعال ساز : ReLU

VC Dimension: 10

متوسط ترین مدل B

تعداد لایه ها: 3 لایه پنهان

تعداد نورون ها : 10نورون

تابع فعال ساز : ReLU

VC Dimension: 40

پیچیده ترین مدل CA

تعداد لایه ها: ۵ لایه پنهان

تعداد نورون ها : ۱۵ نورون

تابع فعال ساز : ReLU

VC Dimension: 100

۱۳-۴-۱ توضیحات پارامترها:

مدل A: وبا تعداد نورون‌های کم و تنها یک لایه پنهان، ظرفیت یادگیری این مدل محدود است، به طوری که می‌تواند تنها الگوهای خطی و ساده را شناسایی کند.

VC Dimensionبرابر ۱۰ به این معناست که این مدل قادر به شناسایی ۱۰ نمونه مستقل و تفکیک‌پذیر است. بنابراین، احتمال دارد بر روی مجموعه‌داده‌های ساده و کم‌پیچیدگی به خوبی عمل کند، اما در صورت ورود به داده‌های پیچیده‌تر، ممکن است تحت‌برازش شود.

مدل B: اینجا با افزایش تعداد لایه‌ها (۳ لایه) و نورون‌ها (۱۰ نورون در هر لایه)، ظرفیت یادگیری مدل افزایش می‌یابد. این شبکه قادر به شناسایی الگوهای پیچیده‌تری است و می‌تواند ویژگی‌های غیرخطی بیشتری را مدل‌سازی کند. VC Dimensionبرابر ۴۰ نشان می‌دهد که این مدل ظرفیت بهتری برای یادگیری داده‌های پیچیده و تنوع بیشتری دارد.

مدل C: با افزایش تعداد لایه‌ها به ۵ و تعداد نورون‌ها به ۱۵ در هر لایه، پیچیدگی مدل به شدت افزایش می‌یابد. این مدل می‌تواند ویژگی‌های پیچیده را به خوبی یاد بگیرد و الگوهای غیرخطی زیادی را شناسایی کند. VC Dimensionبرابر ۱۰۰ نشان‌دهنده میزان بالای ظرفیت یادگیری است. با این حال، با این پیچیدگی، خطر بیشتری از بیش‌برازش نیز وجود دارد، به ویژه اگر داده‌ها محدود باشند.

۱۳-۴-۲ نتیجه‌گیری و ارزیابی:

مدل‌ها را به منظور پیش‌بینی قیمت خانه‌های مختلف آموزش می‌دهیم و با اعتبارسنجی متقابل -cross (validation) کارایی هر یک را بررسی می‌کنیم:

مدل A : احتمال زیادی وجود دارد که عملکرد خوبی بر روی داده‌های ساده داشته باشد، اما در داده‌های پیچیده‌تر با قابلیت تعمیم کم، دچار مشکل می‌شود.

مدل B: قادر است تعادل خوبی بین قدرت یادگیری و تعمیم برقرار کند و معمولاً بهترین انتخاب برای این نوع داده‌ها خواهد بود.

مدل C: در شرایط مناسب، این مدل می‌تواند پیش‌بینی‌های دقیقی انجام دهد، اما دقت سنجی نشان می‌دهد ممکن است به بیش‌برازش دچار شود، به ویژه اگر داده‌های آموزشی کافی وجود نداشته باشد.

۱۳-۴-۳ پیشنهاد نهایی:

با توجه به این مثال عددی، انتخاب تعداد نورون‌ها و لایه‌ها بهینه‌ترین راه برای استنتاج مدل است. مدل B به دلیل VC Dimension معقول و ظرفیت یادگیری کافی، احتمالاً بهترین انتخاب برای پیش‌بینی قیمت خانه در این سناریو است. استفاده از اعتبارسنجی متقابل و ارزیابی بر روی داده‌های جدید به ما کمک می‌کند تا اطمینان حاصل کنیم که این انتخاب به طور صحیحی انجام شده است.

۱۴- مقایسه VC Dimension با معیار های عملکرد مانند خطای آموزش و خطای آزمون و معیار های پیچیدگی مانند پارامتر ها و درجه آزادی مدل:

ابعاد VC (VC Dimension) یک مفهوم اساسی در یادگیری ماشین است که به ما کمک می‌کند پیچیدگی یک مدل را اندازه‌گیری کنیم. اما ابعاد VC تنها معیار برای سنجش پیچیدگی یک مدل نیست.

پارامترها و درجه آزادی مدل نیز دو معیار مهم دیگر هستند که به طور گسترده‌ای استفاده می‌شوند.که می‌تواند توصیف دقیقی از ظرفیت یادگیری و قدرت تعمیم مدل‌های یادگیری ماشین به ما بدهد.

اما برای ارزیابی عملی عملکرد یک مدل، به معیارهای دیگری مانند خطای آموزش و خطای آزمون نیاز داریم. این دو معیار به صورت مستقیم به عملکرد مدل بر روی داده‌ها مرتبط هستند.و به درک عمیق‌تری از تعامل بین ظرفیت مدل و عملکرد واقعی آن بر روی داده‌های واقعی کمک کند.

به توضیح هر یک از این مفاهیم و ارتباطات آنها خواهیم پرداخت و با یک مثال عملی، این تحلیل را روشن‌تر خواهیم کرد.

۱۴-۱ VC Dimension:

VC Dimension (Vapnik-Chervonenkis Dimension) به قابلیت یک مدل یادگیری ماشین برای یادگیری الگوها اشاره دارد. VC Dimension نشان می‌دهد که یک مدل می‌تواند چندین نقطه را به طور مستقل شناسایی کند و در نتیجه ظرفیت تعمیم‌پذیری آن را مشخص می‌کند.

مزایا: می‌تواند به پیش‌بینی رفتار مدل برای داده‌های جدید و ارزیابی ظرفیت یادگیری آن کمک کند.

محدودیت‌ها: VC Dimension نمی‌تواند به تنهایی کارایی مدل را نشان دهد، بلکه باید به انتظارات آموزشی و آزمون‌ها پرداخت.

۱۴-۲ خطای آموزش و خطای آزمون:

خطای آموزش: مقدار خطا (معمولاً دقت) بر روی داده‌های آموزشی که مدل با آن‌ها آموزش دیده است.

خطای آزمون: مقدار خطا بر روی داده‌های جدید و ناشناخته (داده‌های آزمون) که مدل در آن‌ها آموزش نداشته است.

مزایا: این معیارها به ما کمک می‌کنند تا بفهمیم مدل چگونه روی داده‌های آموزشی و جدید عمل می‌کند.

محدودیت‌ها: ممکن است یک مدل برای داده‌های آموزشی بسیار خوب عمل کند (کم خطا) اما بر روی داده‌های آزمون ضعیف باشد، که به وضعیتی به نام **بیش‌برازش** اشاره دارد.

۱۴-۳ مقایسه و ارتباط بین VC Dimension و خطاها:

۱۴-۳-۱ تأثیر VC Dimension برخطای آموزش و آزمون:

یک VC Dimension بالاتر معمولاً نشان‌دهنده ظرفیت بالای یادگیری مدل است. این موضوع می‌تواند منجر به کاهش خطای آموزش (مدل به خوبی داده‌های آموزشی را یاد می‌گیرد) شود.

اما یک VC Dimension بالا همچنین می‌تواند به افزایش خطای آزمون منجر شود، زیرا مدل ممکن است الگوهایی را که فقط در داده‌های آموزشی وجود دارند، یاد بگیرد و به نتایج خوب در داده‌های جدید نرسد (بیش‌برازش).

۱۴-۳-۲ تأثیر خطای آموزش و آزمون بر :VC Dimension:

اگر خطای آموزش بسیار پایین باشد و خطای آزمون بالا باشد، این می‌تواند نشان‌دهنده یک VC Dimension بسیار بالا (بیش از حد) باشد که باعث یادگیری بیش از حد مدل شده است.

در مقابل، اگر خطای آموزش و آزمون نزدیک به هم باشند (و هر دو در سطح نسبتاً بالایی باشند)، احتمالاً VC Dimension به اندازه کافی تنظیم نشده و به ظرفیت یادگیری مشکل دارد.

۱۴-۳-۲-۱ مثال عددی:

فرض کنید چهار مدل خطی مختلف برای پیش‌بینی قیمت یک خانه بر اساس ویژگی‌هایی مانند متراژ، تعداد اتاق‌ها و سن خانه داریم

مدل ها و مشخصات آنها

رگرسیون خطی مدل ۱

VC Dimension:۲

خطای آموزش :۱۵٪

خطای آزمون : ۲۰٪

رگرسیون درجه دوم۲

VC Dimension:۳

خطای آموزش :۱۰٪

خطای آزمون : ۲۵٪

رگرسیون درجه سوم۳

VC Dimension:۴

خطای آموزش :۵٪

خطای آزمون : ۳۰٪

رگرسیون با چند جمله ای درجه چهارم۴

VC Dimension:۵

خطای آموزش :۲٪

خطای آزمون : ۵۰٪

تحلیل این مدل‌ها: **مدل ۱ (رگرسیون خطی)**: ظرفیت یادگیری پایینی دارد (VC Dimension=2) و با وجود اینکه در داده‌های آموزشی عملکرد خوبی دارد، اما خطای آزمون بالایی دارد. این می‌تواند ناشی از عدم توانایی مدل برای یادگیری الگوهای پیچیده‌تری در داده‌ها باشد.

مدل ۲ و ۳: این مدل‌ها با افزایش ظرفیت یادگیری خود (VC Dimensionبیشتر) شروع به کاهش خطای آموزش می‌کنند، اما خطای آزمون آن‌ها در حال افزایش است. این نشان‌دهنده این است که در حال نزدیک شدن به فاز بیش‌برازش هستند.

مدل ۴: دارای ظرفیت یادگیری بالایی است که آن را قادر به یادگیری جزئیات زیاد از داده‌های آموزشی می‌کند. گرچه خطای آموزش به شدت کاهش یافته و تقریباً به صفر رسیده، اما خطای آزمون به شدت افزایش یافته که نشان‌دهنده یادگیری "صدای" داده‌ها و عملاً به معنای بیش‌برازش است.

در نتیجه :**تطابق** VC Dimension**با خطای آموزش و آزمون**: با افزایش VC Dimension، خطای آموزش بهبود می‌یابد اما خطای آزمون ممکن است افزایش یابد که به وضعیت بیش‌برازش اشاره دارد.

مراقبت از پیچیدگی مدل: به کار بردن VC Dimensionبه ما کمک می‌کند تا فشار بیشتری را بر روی تعادل بین ظرفیت یادگیری و قابلیت تعمیم در مدل‌های یادگیری ماشین قرار دهیم.

۱۴-۴ تعداد پارامترها:

تعداد **پارامترها** اشاره به تعداد متغیرهای قابل یادگیری در مدل دارد. پارامترها شامل وزن‌ها و بایاس‌ها در یک مدل مانند شبکه‌های عصبی هستند.

مزایا: تعداد پارامترها معمولاً به ما نشان می‌دهد که یک مدل چقدر می‌تواند داده‌ها را به دقت یاد بگیرد. هرچه پارامترها بیشتر باشند، مدل می‌تواند اطلاعات بیشتری را در خود حفظ کند.

محدودیت‌ها: تنها به تعداد پارامترها نگاه کردن ممکن است گمراه‌کننده باشد. یک مدل با تعداد پارامتر بالا ممکن است دچار بیش‌برازش شود، به‌خصوص اگر داده‌های آموزشی کمی داشته باشیم.

۱۴-۵ درجه آزادی:

درجه آزادی (Degrees of Freedom) به تعداد پارامترهای مستقل در یک مدل اشاره دارد که می‌تواند تحت شرایط خاص تغییر کند. درجه آزادی معمولاً به تعداد پارامترهای آزاد مدل در زمانی که برخی محدودیت‌ها بر روی آن تحمیل نشده است، اشاره می‌کند.

مزایا: درجه آزادی به ما می‌گوید که چه مقدار از اطلاعات می‌تواند به کار گرفته شود برای تعیین پیش‌بینی‌ها؛ به این صورت که درجه آزادی بالا به معنای انعطاف‌پذیری بیشتر مدل است.

محدودیت‌ها: همانند تعداد پارامترها، یک مدل با درجه آزادی بالا ممکن است دچار بیش‌برازش شود که در آن مدل توانایی دقیق یادگیری از داده‌ها را دارد اما به خوبی بر روی داده‌های جدید تعمیم نمی‌کند.

۶-۱۴ مقایسه و ارتباط بین VC Dimension و تعداد پارامترها و درجه آزادی :

۱-۶-۱۴مثال عددی: مدل‌های رگرسیونی ساده

فرض کنید ما سه مدل رگرسیونی برای پیش‌بینی قیمت خانه داریم

رگرسیون خطی مدل A

تعداد پارامتر ها : ۲(ضرایب برای ویژگی های متراژ و سن ، و یک بایاس)

VC Dimension:۲

درجه آزادی :۲

رگرسیون درجه دوم مدل B

تعداد پارامتر ها :۳(ضرایب برای ویژگی های متراژ و سن ، و یک بایاس)

VC Dimension:۳

درجه آزادی :۳

رگرسیون چند جمله ای درجه سوم مدل C

تعداد پارامتر ها :۴(ضرایب برای ویژگی های متراژ و سن ، و یک بایاس)

VC Dimension:۴

درجه آزادی :۴

۲-۶-۱۴تحلیل و مقایسه:

مدل A(رگرسیون خطی): ساده‌ترین مدل است که فقط روابط خطی را شناسایی می‌کند و ظرفیت یادگیری کمی دارد. هم VC Dimension و هم تعداد پارامترها و درجه آزادی آن پایین است. این مدل ممکن است به خوبی بر روی داده‌های ساده عمل کند، اما در حالت‌های پیچیده در خطر تحت‌برازش است.

مدل B(رگرسیون درجه دوم): این مدل نسبت به مدل Aپیچیده‌تر است و می‌تواند الگوهای غیرخطی ساده‌تری را شناسایی کند. VC Dimensionو درجه آزادی این مدل افزایش یافته‌اند و بنابراین احتمال دارد که بهتر از مدل Aبر روی داده‌های پیچیده‌تر عمل کند.

مدل C(رگرسیون چندجمله‌ای درجه سوم): دارای بیشترین پارامترها، VC Dimensionو درجه آزادی است. این مدل می‌تواند الگوهای پیچیده‌تری را یاد بگیرد، اما همچنین در خطر بیشتری برای بیش‌برازش بر روی داده‌های آموزشی قرار دارد.

در نتیجه می‌توان گفت که: VC Dimensionبه ما می‌گوید که یک مدل چگونه می‌تواند به طور مستقل الگوهایی را یاد بگیرد و آیا قادر به تعمیم بر روی داده‌های جدید است یا خیر. تعداد پارامترها و درجه آزادی به ما نشان می‌دهند که یک مدل چقدر می‌تواند انعطاف‌پذیر باشد و چگونه می‌تواند منابع مختلف را برای یادگیری به کار گیرد. نهایتاً، برای انتخاب بهترین مدل، باید به ترکیبی از این معیارها توجه کنیم تا تضمین کنیم که مانع هر دو مشکل بیش‌برازش و تحت‌برازش شویم.

۱۵- VC Dimension در دنیای واقعی:

VC Dimensionدر بسیاری از زمینه‌های یادگیری ماشین و تحلیل داده‌ها کاربردهای عملی دارد. چندین کاربرد عملی VC Dimensionرا به همراه مثال‌های واقعی بررسی خواهیم کرد:

۱۵-۱ انتخاب و طراحی مدل‌ها:

کاربرد: VC Dimensionبه محققان و مهندسان کمک می‌کند تا مدل‌های مناسب‌تری برای داده‌های خاص انتخاب کنند.

مثال: فرض کنید در حال توسعه مدلی برای شناسایی چهره‌های انسانی در تصاویر هستید. از مدل‌های مختلفی مانند رگرسیون لجستیک (VC Dimensionکم) یا شبکه‌های عصبی پیچیده (VC Dimensionبالا) استفاده می‌کنید. با سنجش VC Dimensionهر مدل، می‌توانید تصمیم بگیرید که آیا برای شناسایی الگوهای پیچیده نیاز دارید از یک مدل با ظرفیت بالاتر (مثل شبکه عصبی) استفاده کنید یا یک مدل ساده‌تر کافی است.

۱۵-۲ ارزیابی و مدیریت ریسک:

کاربرد: VC Dimensionمی‌تواند در تشخیص اینکه آیا مدل‌های مالی مناسب برای پیش‌بینی رفتار بازار هستند، مفید باشد.

مثال: شما به عنوان تحلیلگر داده در یک شرکت مالی، می‌خواهید روندهای آینده قیمت سهام را پیش‌بینی کنید. با استفاده از مدل‌هایی مانند ماشین‌های بردار پشتیبان(SVM) یا درخت‌های تصمیم، می‌توانید VC Dimensionآنها را محاسبه کنید. اگر VC Dimensionیک مدل خیلی بالا باشد، این مدل ممکن است به داده‌های آموزشی بیش از حد وابسته شده و در پیش‌بینی داده‌های جدید عملکرد ضعیفی داشته باشد (بیش‌برازش).

۱۵-۳ طبقه‌بندی ایمیل‌ها:

کاربرد: تشخیص اینکه یک ایمیل اسپم است یا خیر.

مثال: برای سیستم‌های فیلتر اسپم، می‌توانید چند مدل مختلف (مثل رگرسیون لجستیک، درخت تصمیم و شبکه عصبی) برای پیش‌بینی اینکه آیا یک ایمیل اسپم است یا نه استفاده کنید. با مقایسه VC Dimensionاین مدل‌ها، می‌توانید تصمیم بگیرید که کدام مدل بهتر عمل می‌کند. اگر ایمیل‌های جدید با مدل انتخابی خوب طبقه‌بندی نشوند، ممکن است نیاز به انتخاب یک مدل با VC Dimensionبالاتر داشته باشید.

۱۵-۴ یادگیری عمیق در بینایی ماشین:

کاربرد: طراحی شبکه‌های عصبی پیچیده.

مثال: در پروژه‌ای برای تشخیص اجسام در تصاویر، می‌توانید شبکه‌های عصبی عمیق طراحی کنید. با توجه به VC Dimensionاین شبکه‌ها، می‌توانید تعداد لایه‌ها و نورون‌ها را بهینه کنید تا مدلی بسازید که هم قدرت تعمیم خوبی داشته باشد و هم در خطر بیش‌برازش نباشد. بالابردن VC Dimensionمی‌تواند منجر به یادگیری جزئیات بیشتر از تصاویر شود، اما باید مراقب باشید که از حدود ظرفیت تعمیم مدل فراتر نروند.

۱۵-۵ پردازش زبان طبیعی

کاربرد: مدل‌سازی و تحلیل متن.

مثال: هنگام توسعه یک مدل برای تحلیل احساسات (مثل تشخیص مثبت یا منفی بودن یک نظر)، می‌توانید از مدل‌های مختلف یادگیری ماشین استفاده کنید. با محاسبه VC Dimension، می‌توانید ظرفیت هر مدل را بسنجید و متوجه شوید که آیا مدل انتخابی شما قادر به پردازش ویژگی‌های پیچیده زبان طبیعی است یا خیر. اگر VC Dimensionبسیار پایین باشد، مدل ممکن است نتواند تنوع و پیچیدگی احساسات را شناسایی کند.

۱۵-۶ بیوانفورماتیک

کاربرد: شناسایی الگوهای ژنی.

مثال: در شناسایی الگوهای ژنی برای تشخیص بیماری‌ها، می‌توانید از مدل‌های یادگیری ماشین برای تحلیل داده‌های بزرگ ژنتیکی استفاده کنید. با در نظر گرفتن VC Dimension، می‌توانید مدل‌هایی انتخاب کنید که از ویژگی‌های ژنتیکی به شیوه‌ای معنی‌دار استفاده کنند و در عین حال در خطر بیش‌برازش نباشند.

۱۶- پیش بینی روندهای آینده در تحقیقات مرتبط با VC Dimension:

مواردی که ذکر میکنیم شامل ترکیبی از تحقیقات فعلی و پیش‌بینی‌های آینده هستند.

۱۶-۱ موارد تحقیقی جاری:

ادغام با تکنیک‌های یادگیری عمیق: مطالعات زیادی در حال حاضر در این زمینه در حال انجام است. محققان در حال بررسی VC Dimensionدر شبکه‌های عصبی عمیق و تأثیر آن بر تعمیم‌پذیری مدل‌ها هستند.

بهبود تعمیم‌پذیری: پژوهش‌هایی در حال حاضر به بررسی روش‌های جلوگیری از بیش‌برازش و افزایش تعمیم‌پذیری با استفاده از VC Dimensionمی‌پردازند.

کاربردهای چندتخصصی: کاربرد VC Dimensionدر حوزه‌های مختلف، از جمله پزشکی و علوم اجتماعی، در حال حاضر مورد بررسی قرار دارد.

مدل‌سازی داده‌های غیرساختاری: تحقیقات فعلی به بررسی VC Dimensionدر یادگیری از داده‌های غیرساختار مانند متون و تصاویر می‌پردازند.

۲-۱۶ موارد پیش‌بینی‌شده برای آینده:

توسعه ابزارهای محاسباتی: این موضوع همچنان در حال شکل‌گیری است و احتمالاً در آینده نرم‌افزارهای بهتری برای محاسبه VC Dimensionایجاد خواهند شد.

تحقیق بر روی مدل‌سازی غیرخطی: این حوزه هنوز به طور کامل شکوفا نشده و انتظار می‌رود در آینده بیشتر مورد بررسی قرار گیرد.

تحلیل داده‌های کم‌مقدار: این موضوع نیز در حال حاضر در حال تحقیق است، اما بهبودهای بیشتری ممکن است در آینده روی دهد.

آموزش و کاربردهای عملی: پژوهش گران در حال حاضر، از اهمیت VC Dimensionدر دوره‌های آموزشی آگاه هستند، اما ممکن است در آینده توجه بیشتری به این موضوع شود.

۱۷- نتیجه گیری:

VC Dimensionمعیاری کلیدی در نظریه یادگیری ماشین است که حداکثر تعداد نقاطی را که یک مدل می‌تواند بدون خطا تفکیک کند، مشخص می‌کند. این مفهوم به ما کمک می‌کند تا قدرت تعمیم‌پذیری و ظرفیت یادگیری مدل‌ها را ارزیابی کنیم و تعادل مناسبی بین این دو ایجاد کنیم. استفاده از VC Dimensionدر انتخاب و طراحی مدل‌ها می‌تواند به شناسایی و جلوگیری از مشکلاتی مانند بیش‌برازش و تحت‌برازش کمک کند و به توسعه مدل‌هایی منجر شود که در حل مسائل واقعی عملکرد بهتری دارند.

به‌طور کلی، VC Dimensionابزاری حیاتی برای محققان و مهندسان یادگیری ماشین است و در کاربردهایی مانند پردازش زبان طبیعی، بینایی ماشین، و تحلیل داده‌های بزرگ به کار می‌رود. درک و استفاده صحیح از این مفهوم می‌تواند به بهینه‌سازی مدل‌ها و بهبود نتایج در پروژه‌های مختلف منجر شود.

۱۸- منابع:

[1]- Using book: Pattern Recognition and Machine Learning written by: Christopher M. Bishop.Book link :

//www.rasa-ai.com/wp-content/uploads/2022/02/Bishop-Pattern-Recognition-And-Machine-Learning-Springer-2006.pdf:

[2]- https://ieeexplore.ieee.org/document/6001314/

[3]- https://ieeexplore.ieee.org/document/9121280/

[4]- https://ieeexplore.ieee.org/document/6488859/

[5]- https://ieeexplore.ieee.org/document/6977327/

[6]- https://ieeexplore.ieee.org/document/6033352/

[7]- https://ieeexplore.ieee.org/document/7008521/

[8]- https://ieeexplore.ieee.org/document/933889/

[9]- https://ieeexplore.ieee.org/document/7354386/

[10]- https://ieeexplore.ieee.org/document/6790116/

[11]- https://ieeexplore.ieee.org/document/8710465/

[12]- https://www.sciencedirect.com/science/article/pii/S0012365X24003236

[13]- https://www.sciencedirect.com/science/article/pii/S0166218X24000830

[14]- https://www.sciencedirect.com/science/article/pii/S0195669824000131

[15]- https://www.sciencedirect.com/science/article/pii/S089360802300566X

[16]- https://www.sciencedirect.com/science/article/pii/S030505482200288X

[17]- https://www.sciencedirect.com/science/article/pii/S0885064X21000558

[18]- https://www.sciencedirect.com/science/article/pii/S0022169421011859

[19]- https://www.sciencedirect.com/science/article/pii/S0925231220301144

[20]- https://www.sciencedirect.com/science/article/pii/S002001901930002X

[21]- https://www.sciencedirect.com/science/article/pii/S0166218X19304482

[22]- https://www.sciencedirect.com/science/article/pii/S0925772118301949

[23]- https://www.sciencedirect.com/science/article/pii/S0020019014001252

[24]- https://www.sciencedirect.com/science/article/pii/S0166218X16301755

[25]- https://www.sciencedirect.com/science/article/pii/S0012365X15002174

[26]- https://www.sciencedirect.com/science/article/pii/S092523121500939X

[27]- https://www.sciencedirect.com/science/article/pii/S0925772113001053

[28]- https://www.sciencedirect.com/science/article/pii/S0097316512000672

[29]- https://www.sciencedirect.com/science/article/pii/S0024384109001740

[30]- https://www.sciencedirect.com/science/article/pii/S1359646213003047

[31]- https://www.sciencedirect.com/science/article/pii/S0888613X09000693

[32]- https://www.nature.com/articles/s43588-021-00084-1

[33]- https://www.sciencedirect.com/science/article/pii/S0893608024008530

[34]- https://synapse.koreamed.org/pdf/10.4258/hir.2016.22.4.351

[35]- https://link.springer.com/article/10.1007/s00158-020-02748-4

[36]- https://link.springer.com/book/10.1007/978-1-4471-4610-0

[37]- https://www.tandfonline.com/doi/abs/10.1080/0143116031000114851

1st International Conference on
Artificial Intelligence
in the Era of Digital Transformation
Event Place: Tbilisi,Georgia
www.Aicnf.ir
اولین کنفرانس بین المللی
هوش مصنوعی در عصر تحول دیجیتال | گرجستان
1st International Conference on Artificial Intelligence in the Era of Digital Transformation
PUBLISH IN JOURNALS — INTERNATIONAL CERTIFICATION

VC Dimension: A Comprehensive Review

Fatemeh Benahamzai
Department of Computer Engineering, Artificial Intelligence and Robotics Engineering, Islamic Azad University, Tehran, Iran

Since the evolution of humanity, mankind has used a variety of tools to perform various tasks in a simpler way. The creativity of the human brain has led to the invention of various machines. These machines have made human life easier and enabled people to fulfill various needs of life including travel, industries, and computing. And machine learning is one of them. According to Arthur Samuel, machine learning is defined as the field that gives computers the ability to learn without explicit programming and is divided into eight general learning methods: supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, multi-task learning, ensemble learning, neural network, and instance-based learning. There is no one type of algorithm that is best for solving a problem. The type of algorithm employed depends on the type of problem you want to solve, the number of variables, the type of model it is suitable for, etc.

The VC dimension is a fundamental concept in machine learning theory used to measure the complexity and generalization ability of learning models. Although it is not directly a machine learning algorithm or model, it is a very important tool for analyzing and understanding the performance of machine learning models.

This concept is directly related to the probability of model overfitting and helps us select models that not only learn well from training data, but also have the ability to accurately predict new data.

Therefore, we conducted our study in the field of VC Dimension and analyzed their theoretical and numerical examples. And we examined VC Dimension from various aspects, including its importance, challenges, and general applications. In addition, we analyzed and examined its relationship with Overfitting. And we also described the study of VC inequality and its calculations. Then we mentioned the use of VC Dimension in model selection. Finally, we evaluated VC Dimension with real-world examples. And we also observed that the science of machine learning is the answer to many of the challenges raised. In this regard, we received review articles and research articles from reputable scientific databases in the world and thoroughly reviewed and analyzed them. So that it can be used accurately and practically by researchers in this field.

Keywords: Machine learning models, learning theory, depth of generalization, generalizability, complexity analysis, pre-fitting, complex models, VC Dimension, learning rate.