

①

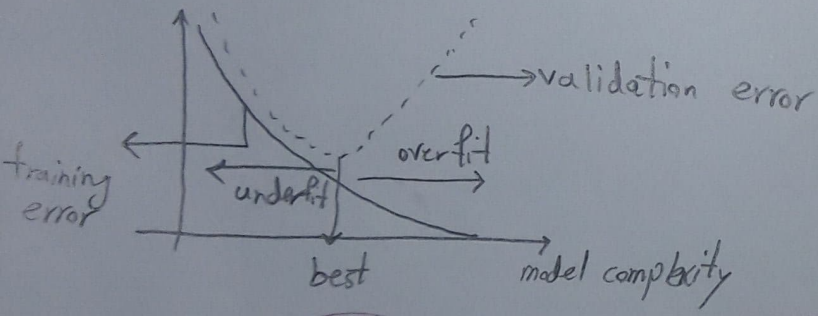
model selection / assessment خردمان چگونه از اطمینان پیداست فرایند خطی مقایسه می‌کنند؛ در model ما مدلی را از میان انتخاب می‌کنیم

model ML برای داده‌های در نظر می‌گیریم (بایستی به dataset و ...) model assessment ما مدل انتخاب شده را روی داده‌ها آزمون (validation) ارزیابی می‌کنیم برای اطمینان آن از روش‌های مدل‌های ما شدیم ما در این انتخاب استفاده می‌کنیم

②

generalization: نیکی که مدلی که انتخاب کرده ایم با داده‌های جدید (داده‌ی دیدنی) در وقت عمل می‌کند این داده‌ها باید دارای ویژگی‌هایی باشند از جمله آن این که ۱- نباید در آن train آورده شود ۲- باید با توزیع داده‌های train باشد

نیکی از روش‌های generalization استفاده از validation است. به طور کلی می‌دانیم اگر مدل را با داده‌های train آموزش دهیم و خطای آن را محاسبه کنیم (در آن تست می‌کنیم) خطای variance را کاهش می‌دهیم ولی خطای bias زیاد می‌شود (بسیار زیاد است) و این به دلیل این است که ما در train داده‌ها را دیده‌ایم و در validation تست می‌کنیم؛ پس مدل را با استفاده از داده‌های train آموزش می‌دهیم و بعد از آن با استفاده از داده‌های validation اعتبار نمی‌دهیم به این نتیجه می‌توانیم از overfit (از دست دادن بایاس) و underfit (از دست دادن واریانس) جلوگیری می‌کنیم



cross-validation استاندارد می‌شود. در این روش داده‌ها را به چندین بخش تقسیم کرده و در هر بار با یکی از این بخش‌ها به آموزش و اعتبار بخشی مدل می‌پردازیم

③

و باقیه مدل‌ها (test) تست می‌کنیم

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

mean squared error متریک: برای ارزیابی مدل‌ها استفاده می‌شود. در واقع تفاوت بین مقدار واقعی و مقدار پیش‌بینی شده را محاسبه می‌کنیم. MSE همیشه مثبت است. اول به فاصله‌هایی که فاصله‌ها دارد. (احتمالی است $p > 0$)

④

در این که تخمین برابری که قابلیت تولید تخمین دقیق‌تری دارد را حساب می‌کنند. هر چه بیشتر صفر شود این تخمین بهتر است.

① گاهی ممکن است تعداد ویژگی‌های داده کم‌بیشتر بسیار زیاد باشد؛ اما در حقیقت بتوان داده‌ها را برپایه با تعداد گهتی ویژگی (feature) تفحص کرد؛ یعنی از دلایلی می‌تواند این ویژگی‌ها را به feature هایش تبدیل کنیم استقلال ندارند (در طبیعت و حقیقت است). پس لازم است ویژگی‌ها را طیفی در هم، اگر از تعداد feature های موجود استفاده کنیم و حذف انجام ندهیم می‌تواند به کمک این ویژگی‌ها (train) وابسته به نام؛ overfit ایجاد کند. به طوری اهداف و برای feature عبارتند از: ① گاهی پیچیدگی زمانی و $O(n)$ و ② عدم وابستگی زیاد در حافظه ③ سببیت data و قابلیت در یافت به تعداد داده و فهم آن (insight) و ... visualization

ب) ضمیمه جواب بعضی است. اگر کلمات معانی باشند معیار فاعله معیار و مفعول نیست چون نقاط ابتدا و انتهای خط با وجود فاعله زیاد،
مخبر اول
باری عقول بیان باشند. هر چند اگر داده ها بصورت داده ای باشند نیز این معیار مناسب نیست و دلیل آن نیز مشابه آن چه ذکر شد.

انت.

خوبه: این روشی از الگوریتمهای density-based می باشد. که بر مبنای نزدیکی نقاط به هم و اندازه آنها در یک کلاستر (cluster) قرار می دهد و اگر نقطه (point) دورتر از این داده ها باشد بعنوان outlier در نظر گرفته می شود. مراحل الگوریتم ۱- انتخاب نقطه دلخواه ۲- تمام نقاط قابل قبول از نظر density را انتخاب می کنیم (با فرض ۴ minpts) ۳- اگر نقطه اولی (انتخابی) core باشد؛ مجموعه کلی cluster می دهد. ۴- اگر نقطه مدزی باشد؛ نقطه دلخواه دیگری را بیابیم. ۵- الگوریتم را ادامه می دهیم تا تمام نقاط پیری خوشه یا کلاسترهای جدا را پیدا کنیم. ۶- همچنین اگر یک نقطه از این مجموعه را پیدا کنیم و با هیچ داده ای تعلق cluster نداشته باشد (point) نوین / outlier است.

density reachability

density reachability

9 در رابطه

density reachability

density reachability

و نقطه‌ای در هر سالی ع (رایگویی) از آن مقدار کنید.

چند نکته: روش DBSCAN بر اساس $yes/no \in \text{boolean}$ و $value$ برای تفسیر تراکم نقاط به صورت مفهومی عمل می‌کند (مثلاً غنی/فقره) و در حالت کلی برای تفسیر تراکم عملیات pop برای تفسیر میزان تراکم

از $reachability$ استعاره می‌کنیم تا معنی حالت های میانی داشته توصیف کنیم و بجای yes/no ، مقادیر عددی داریم همینطور $distance$

در این روش برای توصیف رابطه بجای yes/no از ضرایب عددی استفاده می‌نور. علاوه بر این روش $OPTICS$ برخلاف $DBSCAN$

روشی گریزانه (greedy) است و به بیان از نظر حافظه زمان در حال پیدایش کوتاه تر است (روش دیگر الگوریتم رایج صورت

~~احساس کامل خورشید و تکامل زمین~~
بکار می‌رود - $random$ آغاز می‌کند $iterational$ /

اثر دیوولاما PBS(A) می توان - تقابل غیر متقابل کلاس های طیفال نزدیک بهم اشاره کرد. حال آن که در $opt(A)$ دیالیت

9

بصورت ظلی قریب می شود.

⑦ در PCA ما به دنبال محلی هستیم که بیشترین تغییرات داده در آن جهت

می یابند. (موفق اول برای) همچنین این الگوریتم را می توان تکمیل کرد تا داده های جهت محلی که بیشترین اطلاعات را

می دهد. (تفسیر داده) جهت محلی فواید دارد (موفق دوم). در LDA ما به دنبال محلی هستیم که با انتخاب آن و مقیاس

داده ها روی آن بیشترین تغییرات پذیری / حساسیت را بین کلاس ها صورت گیرد.

- برای وقت بیشتر نظر گذارن آورده می شود. به همراه توالات کد دار (3,4,5)

