

https://colab.research.google.com/drive/1lhdmgHzd-5pp288biEgq9o79g0kt_EHo?usp=sharing

1. در روش حداقل مربعات خطا (LSE)، معیار خطا به صورت مجموع مربعات اختلاف بین مقادیر واقعی و پیش‌بینی شده تعریف می‌شود چون این معیار باعث می‌شود که خطاهای بزرگ‌تر تأثیر بیشتری روی تابع هدف داشته باشند، که تحلیل را نسبت به انحرافات بزرگ حساس‌تر می‌کند و علاوه بر آن تابع هدف (مجموع مربعات خطا) یک تابع پیوسته و مشتق‌پذیر است که می‌توان با روش‌های تحلیلی یا عددی، به راحتی بهینه‌سازی شود.
2. اگر داده‌ها دارای نویز غیر گوسی یا دارای داده‌های پرت باشند، استفاده از LSE ممکن است منجر به نتایج نادرست شود، زیرا LSE نسبت به خطاهای بزرگ حساس است. برای رفع این مشکل می‌توان به جای LSE از روش‌های مقاوم (Robust Estimation) مثل روش حداقل قدرمطلق خطاها (Least Absolute Deviations) یا استفاده از تکنیک‌هایی مثل حذف داده‌های پرت (Outlier Removal) استفاده کرد.
3. اگر توزیع نویز داده‌ها نرمال نباشد، روش LSE برآوردکننده‌های بایاس‌دار و ناکارا خواهد داشت. زیرا LSE در اصل برای حالت نویز گوسی بدون بایاس و واریانس ثابت طراحی شده است. در غیر این صورت، تخمین پارامترها دیگر ویژگی‌های بهینه‌ای مانند نااریب بودن یا حداقل واریانس را نخواهد داشت.
4. در شرایطی که داده‌ها دارای نویز غیر نرمال باشند یا داده‌ها دارای تعداد زیادی نقاط پرت باشند، استفاده از LSE مناسب نیست. زیرا چند داده‌ی پرت می‌توانند تابع هزینه را به شدت تحت تأثیر قرار دهند و منجر به تخمین‌های اشتباه شوند.
5. برای اثبات اینکه روش LSE بدون بایاس نیست باید فرض کنیم که مدل واقعی به صورت خطی باشد ولی نویز دارای امیدریاضی صفر نباشد یا واریانس یکنواخت نداشته باشد. در این شرایط می‌توان نشان داد که امیدریاضی تخمین پارامترهای به دست آمده از LSE برابر با پارامتر واقعی مدل نخواهد بود، پس بایاس وجود خواهد داشت.
6. روش MLE به جای کمینه کردن مربع خطا، احتمال مشاهده داده‌ها را بیشینه می‌کند. در این روش تابع درستنمایی (Likelihood Function) تعریف می‌شود و تخمین پارامترها به گونه‌ای انجام می‌شود که احتمال مشاهده داده‌های موجود بیشترین مقدار را داشته باشد.
7. در حالتی که داده‌ها دارای نویز گوسی باشند، روش MLE دقیقاً تخمین‌های مشابه LSE را تولید می‌کند. برای سایر توزیع‌ها، MLE همچنان قابل استفاده است زیرا این روش بر اساس بیشینه کردن احتمال داده‌ها

نسبت به پارامترهای مدل عمل می‌کند و فرمی خاص برای توزیع داده‌ها در نظر می‌گیرد (مثلاً توزیع نمایی، پواسون و...).

8. وقتی داده‌ها دارای توزیع غیرگوسی باشند (مثلاً توزیع پواسون یا نمایی)، روش MLE با استفاده از تابع درستنمایی متناظر با آن توزیع ساخته می‌شود. در نتیجه تخمین‌ها به جای اتکا بر فرض خطی بودن یا گوسی بودن نویز، به صورت سازگار با توزیع داده‌ها انجام می‌شود و عملکرد بهتری خواهد داشت.
9. در حضور داده‌های کم یا نمونه‌های کوچک، روش MLE ممکن است برآوردهای ناپایداری ارائه دهد، چون تابع درستنمایی ممکن است چندین ماکزیمم محلی داشته باشد یا به داده‌های محدود بسیار حساس باشد. برای رفع این مشکل می‌توان از تکنیک‌هایی مثل استفاده از تخمین‌های بیزی (Bayesian Estimation)، منظم‌سازی (Regularization)، یا روش‌های شبه‌بیزی بهره گرفت تا اطلاعات پیشین به تخمین اضافه شود و پایداری تخمین افزایش یابد.

-
1. مدل‌های ساده‌ای مانند رگرسیون خطی نسبت به نویز حساس‌ترند چون فرض می‌کنند که رابطه بین متغیرها خطی و بدون انحراف است. در حالی که مدل‌هایی مثل جنگل تصادفی (Random Forest) از طریق تجمیع تصمیم‌های چندگانه (Ensemble Learning) عمل می‌کنند و می‌توانند نویز را بهتر فیلتر کنند و بر روی الگوهای پایدار داده‌ها تمرکز کنند. به همین دلیل در داده‌های واقعی و نویزی، مدل‌هایی مانند جنگل تصادفی عملکرد بهتری نسبت به مدل‌های خطی دارند.
2. کمترین مقدار MSE به تنهایی نشان‌دهنده بهترین مدل برای پیش‌بینی نیست. ممکن است مدلی که کمترین MSE را دارد، دچار بیش‌برازش (Overfitting) شده باشد و فقط روی داده‌های آموزشی خوب عمل کند ولی در داده‌های جدید پیش‌بینی‌های ضعیفی داشته باشد. بنابراین، ارزیابی کیفیت مدل باید با معیارهایی مثل Generalization Error روی داده‌های تست یا اعتبارسنجی انجام شود، نه صرفاً مقدار MSE روی داده‌های آموزش.
3. مدل‌های پیچیده مثل رگرسیون چندجمله‌ای می‌توانند داده‌های آموزشی را با دقت بسیار بالا برازش کنند ولی دچار مشکل بیش‌برازش شوند. این مدل‌ها انعطاف‌پذیری بالایی دارند ولی به کوچک‌ترین تغییرات در داده حساس می‌شوند. مدل‌های ساده‌تر اغلب تعمیم‌پذیری (Generalization) بهتری دارند چون فرض‌های قوی‌تر و محدودتری روی شکل تابع دارند، بنابراین کمتر درگیر نویز و جزئیات غیرضروری داده می‌شوند.
4. در شرایطی که داده‌ها دارای ویژگی‌های همبسته (Collinear Features) باشند، استفاده از روش‌هایی مثل Ridge Regression و Lasso مناسب‌تر است چون این روش‌ها با افزودن جریمه (Regularization) به مدل، از بزرگ شدن بیش از حد ضرایب جلوگیری می‌کنند. Ridge باعث کوچک شدن ضرایب می‌شود و Lasso می‌تواند حتی برخی ضرایب را دقیقاً صفر کند و به انتخاب ویژگی‌ها کمک کند. در نتیجه مدل پایدارتر و قابل تعمیم‌تر خواهد بود.
5. برای تصمیم‌گیری درباره انتخاب بهترین مدل باید بسته به ویژگی‌های مسئله تحلیل دقیقی انجام شود. مثلاً اگر داده‌ها نویزی باشند مدل‌های مقاوم لازم است، اگر حجم داده‌ها زیاد باشد مدل‌های سریع‌تر ترجیح دارند، اگر ویژگی‌های زیادی وجود داشته باشد روش‌هایی که انتخاب ویژگی انجام می‌دهند بهترند. استفاده

- از تکنیک‌هایی مانند اعتبارسنجی متقاطع (Cross Validation)، بررسی منحنی‌های یادگیری (Learning Curves) و تحلیل بایاس-واریانس می‌تواند به انتخاب مدل مناسب کمک کند.
6. روش‌هایی مانند KNN و SVR در مواجهه با داده‌های بزرگ محاسباتی بسیار سنگین می‌شوند چون در KNN برای هر پیش‌بینی نیاز است کل داده‌ها جستجو شوند و در SVR مسئله بهینه‌سازی پیچیده‌ای حل می‌شود. برای بهینه‌سازی این روش‌ها می‌توان از تکنیک‌هایی مانند کاهش ابعاد (Dimensionality Reduction)، استفاده از داده‌های نمونه‌گیری شده (Sampling)، یا نسخه‌های سریع‌تر الگوریتم‌ها مثل Approximate KNN یا استفاده از هسته‌های تصادفی (Random Fourier Features) استفاده کرد.
7. زمانی که تعداد داده‌ها کم باشد یا عدم قطعیت بالایی در داده‌ها وجود داشته باشد، رگرسیون خطی بیزی مزیت دارد؛ چون می‌تواند عدم قطعیت در پیش‌بینی را به‌طور مستقیم مدل کند. این روش همیشه دقیق‌تر نیست؛ در بسیاری از مسائل با داده‌های زیاد، روش‌های کلاسیک ممکن است عملکرد مشابه یا حتی بهتری داشته باشند و محاسبات ساده‌تر باشند.
8. می‌توان از مدل‌های مقاوم (Robust Models) استفاده کرد که حساسیت کمتری نسبت به داده‌های پرت دارند. همچنین استفاده از روش‌های بیزی یا مدل‌های ensemble مثل Random Forest می‌تواند دقت را در این شرایط افزایش دهد.
9. خیر، همیشه مناسب نیستند. برای داده‌هایی که عدم تعادل دارند یا دارای نویز زیاد هستند، معیارهایی مثل MAE (میانگین قدر مطلق خطا) یا مدل‌های طبقه‌بندی با معیارهایی مثل AUC-ROC بهتر عمل می‌کنند.
10. رگرسیون خطی کلاسیک (OLS) حساسیت زیادی به داده‌های پرت دارد. راهکارها: استفاده از رگرسیون مقاوم (مثل RANSAC)، حذف داده‌های پرت یا مقیاس‌بندی (Scaling) داده‌ها و استفاده از روش‌های وزن‌دهی به نمونه‌ها.
11. افزایش تعداد مؤلفه‌ها باعث می‌شود اطلاعات بیشتری از داده‌ها حفظ شود، اما ممکن است مدل پیچیده‌تر و ریسک overfitting بیشتر شود. در مدل‌های رگرسیون، این ممکن است باعث کاهش خطای آموزش ولی افزایش خطای تست شود.
12. با استفاده از نمودار واریانس تجمعی (Cumulative Explained Variance Plot)؛ اگر ۱۰ مؤلفه بیش از ۹۰٪ واریانس داده‌ها را پوشش دهد، معمولاً کافی است.
13. نرمال‌سازی باعث می‌شود ویژگی‌ها در مقیاس مشابه قرار بگیرند و مدل سریع‌تر همگرا شود و نتایج بهتری ارائه کند. مخصوصاً در مدل‌هایی که به فاصله و اندازه داده حساس هستند (مثل رگرسیون Ridge یا Lasso، نرمال‌سازی بسیار مهم است).