

تعمیم، منظم‌سازی، دادوستد بایاس-واریانس

درس: مقدمه‌ای بر داده‌کاوی

استاد: دکتر مائده السادات طاهائی

گردآورندگان: کاظم فرقانی، علیرضا کفاشها

مسائل تعمیم، منظم‌سازی و دادوستد بایاس-واریانس

پرسش ۱

(الف) بالا بودن واریانس در مدل چه معنایی دارد؟ یک روش ممکن برای کاهش واریانس در مدل خود بیان کنید.

(ب) چگونه رگولاریزیشن L_2 در های classifier خطی بر روی تعادل بایاس-واریانس تأثیر می‌گذارد؟

پاسخ ۱

(الف) این بدان معناست که مدل به داده‌های آموزشی بیش از حد تطبیق یافته است (overfitting) و قابلیت تعمیم ندارد. استفاده از تکنیک‌های منظم‌سازی مانند dropout، ساختن مدل با پارامترهای قابل یادگیری کمتر و غیره به حل این مسئله می‌تواند کمک کند.

(ب) رگرسیون L_2 با اعمال جریمه‌ای به مقادیر بزرگ وزن‌ها، پیچیدگی مدل را کنترل می‌کند. این امر از بیش‌برازش (overfitting) جلوگیری کرده و واریانس مدل را کاهش می‌دهد بدون اینکه بایاس به‌طور قابل‌توجهی افزایش یابد. در نتیجه، Generalization مدل بهبود می‌یابد.

پرسش ۲

شما در حال طراحی یک مدل برای یک تسک طبقه‌بندی (classification) هستید. در ابتدا مدل خود را بر روی ۲۰ نمونه آموزش می‌دهید و مشاهده می‌کنید که با وجود همگرا شدن آموزش، خطای آموزش بر روی این نمونه‌ها زیاد است. بنابراین تصمیم می‌گیرید که مدل خود را این بار بر روی ۱۰,۰۰۰ نمونه آموزش دهید. آیا روش شما برای حل این مشکل صحیح است؟ اگر بله، محتمل‌ترین نتایج مدل خود را در این حالت توضیح دهید. اگر خیر، راه‌حل مناسب برای رفع این مشکل را بیان کنید.

پاسخ ۲

مدل از مشکل بایاس رنج می‌برد.

افزایش مقدار داده‌ها واریانس را کاهش می‌دهد و محتمل نیست که مشکل را حل کند.

یک رویکرد بهتر کاهش بایاس مدل با احتمالاً افزایش پارامترهای قابل یادگیری بیشتر خواهد بود. ممکن است که فرایند آموزش به یک نقطه بهینه محلی همگرا شود. آموزش طولانی‌تر، استفاده از یک بهینه‌ساز بهتر و یا شروع مجدد از یک مقداردهی اولیه برای وزن‌ها همچنین می‌تواند کارساز باشد.

پرسش ۳

فرض کنید می‌خواهیم سن افراد را بر اساس اسکن مغز آن‌ها با استفاده از Regression تخمین بزنیم. تعداد افرادی که اسکن مغز آن‌ها در دسترس است، ۱۰ نفر است و برای هر فرد، یک بردار ویژگی ۲۰,۰۰۰ تایی بعد از پردازش اسکن مغز در دسترس است.

در این حالت، استفاده از روش L_2 -Regularization یا روش L_1 -Regularization را ترجیح می‌دهیم؟

پاسخ ۳

تعداد داده‌ها بسیار کم و تعداد ویژگی‌ها بسیار زیاد است. لذا، برای جلوگیری از پیچیده شدن مدل، مایلیم تعداد زیادی از ویژگی‌ها را حذف کنیم.

روش L_1 -Regularization باعث صفر شدن تعداد بیشتری از ویژگی‌ها نسبت به روش L_2 -Regularization می‌شود. بنابراین، انتخاب L_1 -Regularization مناسب‌تر است.

پرسش ۴

فرض کنید قیمت یک رمزارز از تابع زیر تبعیت می‌کند: $y = (1 + \sin(\frac{\pi x}{7})) + \epsilon$ که در آن x شماره‌ی روز از ماه است. سه مدل زیر را در نظر بگیرید:

$$\hat{y}_1 = \theta_1 x + \theta_0$$

$$\hat{y}_3 = \theta_3 x^3 + \theta_2 x^2 + \theta_1 x + \theta_0$$

$$\hat{y}_9 = \theta_9 x^9 + \dots + \theta_2 x^2 + \theta_1 x + \theta_0$$

کم یا زیاد بودن مقدار Bias و Variance هر یک از این سه مدل را در حالت‌های زیر با دلیل مشخص کنید. نیازی به محاسبات نیست.

- وقتی ۱۰۰ داده از قیمت این رمزارز در طول یک ماه داریم.

- وقتی ۵ داده از قیمت این رمزارز در طول یک ماه داریم.

پاسخ ۴

- حالت اول (۱۰۰ داده):

مدل \hat{y}_1 بایاس بالا و واریانس پایینی دارد؛ زیرا مدل بسیار ساده است و نمی‌تواند به خوبی الگوی نوسانی داده‌ها را یاد بگیرد.

مدل \hat{y}_3 تعادلی بهتر بین بایاس و واریانس دارد. با توجه به پیچیدگی متوسط آن، می‌تواند الگو را با دقت بیشتری مدل کند بدون آنکه واریانس زیادی ایجاد شود.

مدل \hat{y}_9 واریانس بالایی دارد؛ زیرا بسیار پیچیده است و احتمال دارد بر داده‌های آموزشی بیش‌برازش کند (overfitting)، هرچند بایاس آن کم است.

- حالت دوم (۵ داده):

مدل \hat{y}_1 همچنان بایاس بالایی دارد، اما به دلیل کم بودن داده‌ها، واریانس پایین آن می‌تواند مزیت باشد.

مدل \hat{y}_3 در این حالت واریانس بیشتری پیدا می‌کند و احتمالاً داده‌های محدود برای آموزش آن کافی نخواهد بود، اما همچنان می‌تواند عملکردی متوسط ارائه دهد.

مدل \hat{y}_9 به دلیل پیچیدگی بالا و کم بودن داده‌ها، شدیداً به داده‌های آموزشی حساس خواهد بود و واریانس بسیار بالایی خواهد داشت که عملکرد مدل را تضعیف می‌کند.

منابع و مراجع

- درس یادگیری ماشین، دانشکده مهندسی کامپیوتر دانشگاه صنعتی شریف، دکتر علی شریفی زارچی