

محمد رضا احمدي تشنيزي

۹۸۱۷۰۶۴۶

تمرین ۵

سوال ۱:

(الف)

1. افزایش مجموعه داده‌ها: افزایش حجم و تنوع مجموعه داده‌های آموزشی می‌تواند به کاهش اریبی کمک کند. با جمع‌آوری و اضافه کردن نمونه‌های جدید به مجموعه داده‌ها، مدل بیشتر با موارد مختلف و تنوع داده‌ها آشنا می‌شود و می‌تواند بهتر درکی از الگوها داشته باشد.
2. تنظیم پارامترها: با بهینه‌سازی پارامترهای مدل مانند نرخ یادگیری، تعداد لایه‌ها و نوروها، نوع تابع فعال‌سازی و... می‌توانیم بهبود قابل توجهی در کاهش اریبی داشته باشیم.
3. استفاده از تکنیک‌های رگولاریزاسیون: تکنیک‌هایی مانند Dropout و L1/L2 regularization با کاهش انعطاف‌پذیری مدل و جلوگیری از بیش‌برازش، می‌توانند به کاهش اریبی کمک کنند.
4. استفاده از تکنیک‌های ترکیبی: استفاده از روش‌های ترکیبی مانند انتقال یادگیری یا آموزش مدل‌های موازی (ensemble learning)، می‌تواند بهبود قابل توجهی در کاهش اریبی داشته باشد.
5. بررسی معماری مدل: ارزیابی معماری مدل و تغییر آن به نحوی که بهترین عملکرد را در کاهش اریبی ارائه کند. این شامل تغییر تعداد لایه‌ها، اضافه کردن لایه‌ها و یا تغییر اندازه نوروها می‌شود.
6. تجزیه مقادیر خروجی: بررسی دقیق خروجی‌های مدل و شناسایی منابع اریبی می‌تواند راهکارهایی را برای کاهش آنها فراهم کند.

(ب)

اگر ویژگی‌های همبسته را حذف کنیم:

بایاس ممکن است بیشتر شود؛ زیرا مدل از همبستگی بین ویژگی‌ها استفاده کرده بود و حذف آنها می‌تواند منجر به از دست دادن اطلاعات مهم و کاهش دقت تقریب مدل شود.

واریانس ممکن است کاهش یابد؛ زیرا با حذف ویژگی‌های همبسته، مدل دچار بیش‌برازش کمتری شود و پراکندگی خروجی‌ها کاهش پیدا کند.

با این حال، این تغییرات نهایتاً بسته به ویژگی‌های مدل، حجم داده‌ها و ساختار مسئله‌ی مورد بررسی متفاوت خواهد بود.

این گزاره نادرست است. بایاس به خطایی اشاره دارد که ناشی از ساختار مدل و تخمین‌هایی است که بر روی داده‌های آموزشی صورت گرفته است. افزایش تعداد داده‌های آموزشی نمی‌تواند به طور مستقیم بایاس را کاهش دهد. در واقع، افزایش تعداد داده‌های آموزشی می‌تواند بهبود دقت و قابلیت تعمیم مدل را افزایش دهد، اما بایاس نیز می‌تواند همچنان وجود داشته باشد.

این گزاره نیز نادرست است. افزایش پیچیدگی مدل در رگرسیون ممکن است باعث بشود مدل بیشتر به داده‌های آموزشی *fit* شود و خطای آموزش کاهش یابد. اما این ممکن است باعث بشود مدل بیش از حد پیچیده شود و نمونه‌های تست را به خوبی تعمیم ندهد، که باعث افزایش خطای تست می‌شود. در نتیجه، این گزاره نمی‌تواند به طور عمومی برقرار باشد و بستگی به مسئله و داده‌های مورد استفاده دارد.

(سوال ۲)

(الف)

در آموزش داده‌های درخت تصمیم، در مواردی که ورودی‌ها  $A$  و  $B$  و  $C$  یکسان هستند اما متغیر  $X$  برای آنها متفاوت است، تعدادی دسته‌بندی اشتباه اتفاق افتاده. به طور خاص در این بخش، در ۲ مورد از ۱۳ مورد دسته‌بندی، خطای *misclassification* رخ داده. در نتیجه، خطای تخمینی حدوداً ۱۵.۴٪ است.

(ب)

بیشترین خطا در دسته‌بندی رخ می‌دهد زمانی که با ویژگی‌های یکسان، کمترین اطمینان را در انتخاب کلاس داشته باشیم. این کمترین اطمینان در حالتی رخ می‌دهد که تعداد داده‌های هیچ کلاسی در داده‌هایی که ویژگی‌های یکسانی دارند، بیشتر از تعداد داده‌های کلاس‌های دیگر نباشد یا به طور دقیق تعداد داده‌ها در هر کلاس یکسان باشد. در این حالت، به ناچار یکی از کلاس‌ها را به عنوان خروجی درخت تصمیم انتخاب می‌کنیم، اما در  $k-1/k$  موارد، خطای دسته‌بندی رخ می‌دهد.

سؤال ۳: الف) 
$$h(w) = \begin{bmatrix} p(y=1|x,w) \\ \vdots \\ p(y=k|x,w) \end{bmatrix} = \frac{1}{\sum_{i=1}^K e^{w_i^T x}} \begin{bmatrix} e^{w_1^T x} \\ \vdots \\ e^{w_K^T x} \end{bmatrix}$$

اگر برای هر کدام از کلاس احتمال تعیین کنیم پس دسته ای با بیشترین احتمال را انتخاب کنیم در حالت ۲ متغیر که جمع ۱ است داریم:

$$p(y=k | x=x) = \frac{e^{w_k^T x}}{1 + \sum_{i=1}^{K-1} e^{w_i^T x}} \quad k \neq K$$

$$\parallel \quad = \quad \left( \frac{1}{\parallel} \right) \quad k = K$$

ب) پارامتری که باید تعیین زد  $w$  است پس نیاز داریم پارامترهای  $w_1, \dots, w_K$  را تعیین کنیم این کار را می توان با روش هایی مثل گرادیان نزولی انجام داد.

$$\mathcal{L}(w_1, \dots, w_{K-1}) = \sum_{i=1}^n \ln p(y=y_i | x=x_i) =$$

$$= \sum w_{y_i}^T x_i - \sum \ln \left( 1 + \sum_{i=1}^{K-1} e^{w_{y_i}^T x_i} \right)$$

$$\frac{\partial \mathcal{L}}{\partial w_k} = \sum x_i - \frac{x_i e^{w_{y_i}^T x_i}}{1 + \sum_{i=1}^{K-1} e^{w_{y_i}^T x_i}} = \sum x_i (1 - p_k)$$

$$\frac{\partial f}{\partial w_k} = \frac{\partial l}{\partial w_k} + \frac{\partial \left( \frac{\lambda}{r} \sum \|w_i\|_r^r \right)}{\partial w_k} \quad (c)$$

$$= \sum n_i (1 - p_k) - \lambda \|w_k\|_r$$

$$MSE = (y - w_j q_j)^T (y - w_j q_j) \quad \text{نقال ٤ الف}$$

مشتق نسبت به  $w_j$

$$\Rightarrow -q_j^T (y - w_j q_j) + (y - w_j q_j)^T (-q_j) =$$

$$= -2 q_j^T y + 2 w_j^T q_j^T q_j \rightarrow w_j = \frac{q_j^T y}{q_j^T q_j}$$

$$MSE = (y - Xw)^T (y - Xw) \xrightarrow{\text{مشتق نسبت به } w} \quad (b)$$

$$\Rightarrow -2 X^T y + 2 X^T X w + w^T X^T X = -2 X^T y + 2 X^T X w \Rightarrow$$

$$w = (X^T X)^{-1} X^T y$$

ستون های متغیر  $X$   $\Rightarrow w_j = \frac{q_j^T y}{X_j^T X_j}$

$$MSF = \sum (y_i - w_j x_{ji} - w_0)^2 \xrightarrow{w_j \text{ سست } = 0} \quad (1)$$

$$\Rightarrow \frac{\partial}{\partial w_j} \left( \sum (y_i - w_j x_{ji} - w_0)^2 \right) (-x_{ji}) = 2 \left( -\sum y_i x_{ji} + w_j \sum x_{ji}^2 + w_0 \sum x_{ji} \right)$$

از طرفی  $\frac{\partial}{\partial w_0} = 0$   $\Rightarrow \sum (y_i - w_j x_{ji} - w_0) = 0$

$$\Rightarrow w_0 = \bar{y} - w_j \bar{x}_j = E(y) - w_j E(x_j)$$

$$\Rightarrow w_j = \frac{\sum y_i x_{ji} - \frac{\sum y_i \sum x_{ji}}{n}}{\sum x_{ji}^2 - \frac{(\sum x_{ji})^2}{n}} = \frac{\text{Cov}(x, y)}{\text{Var}(x)}$$