تمرین پنجم

۴۰۰۱۰۴۹۶۴ — آیلین بسته

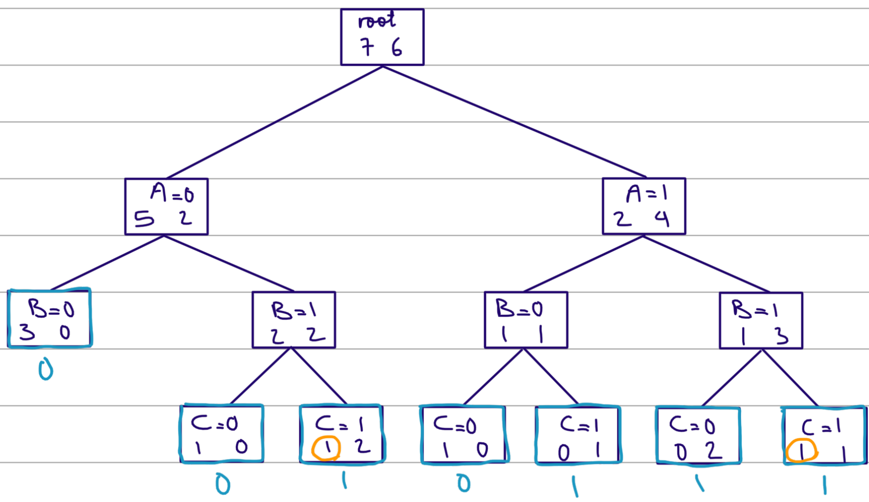
**۱) الف)** ۱) افزایش تعداد داده‌ها ۲) افزایش پیچیدگی مدل ۳) افزایش ویژگی‌ها

**ب)** وجود همبستگی بین داده‌ها ممکن است موجب افزایش پیچیدگی مدل شود که این موجب افزایش بایاس و واریانس می‌شود. با حذف این همبستگی‌ها مدل را به مدل ساده‌تری تبدیل می‌کنیم و بایاس کاهش می‌یابد (ممکن است واریانس افزایش بیابد ولی خیلی زیاد نیست) و در نهایت مدل ما بهتر می‌شود.

**پ)** گزاره اول زیرا با افزایش داده‌ها می‌توانیم اثر داده‌های نویزدار را بیشتر از بین ببریم و به مدل واقعی نزدیکتر شویم. گزاره دوم غلط است زیرا پیچیدگی کمتر از مقدار مورد نیاز سبب افزایش خطا می‌شود و پیچیدگی زیاد هم می‌توان با افزایش تعداد دیتا جبران کرد.

| C | B | A | X |
|---|---|---|---|
| ۰ | ۰ | ۰ | ۰ |
| ۱ | ۰ | ۰ | ۰ |
| ۱ | ۰ | ۰ | ۰ |
| ۰ | ۱ | ۰ | ۰ |
| ۱ | ۱ | ۰ | ۰ |
| ۱ | ۱ | ۰ | ۱ |
| ۱ | ۱ | ۰ | ۱ |
| ۰ | ۰ | ۱ | ۰ |
| ۱ | ۰ | ۱ | ۱ |
| ۰ | ۱ | ۱ | ۱ |
| ۰ | ۱ | ۱ | ۱ |
| ۱ | ۱ | ۱ | ۰ |
| ۱ | ۱ | ۱ | ۱ |

**۲) الف)**




درصد خطای مدل ← $\frac{2}{13} \times 100 = 15.4\%$

ب) اگر داده ها به صورت کاملا یکنواخت پخش شده باشند و نتوان تفکیکی بین آنها قائل شد (اعضای هر کلاس باهم برابر باشند) در این صورت همه آنها به یک کلاس نسبت می شوند و درصد خطا برابر می شود با: $\frac{K-1}{K} \times 100$

۳) الف)

$$p(y=i \mid X; W) = \frac{e^{w_i^T X}}{\sum_{j=1}^{K} e^{w_K^T x}} \qquad X = (1, x_1, \dots, x_m) \qquad W = \begin{bmatrix} w_{10} & w_{12} & \dots & w_{1m} \\ \vdots & & & \vdots \\ w_{K0} & w_{K2} & \dots & w_{Km} \end{bmatrix}$$

ب) در این حالت پارامتر W باید تخمین زده شود.

پ)

$$\ell(w_1, \dots, w_K) = \prod_{i=1}^{n} \prod_{K=1}^{K} \left( \frac{e^{w_K^T X_i}}{\sum_{h=1}^{K} e^{w_h^T X_i}} \right)^{\mathbb{1}(i,K)} \qquad \mathbb{1}(i,K) = y_i == K$$

$$L(w_1, \dots, w_K) = \sum_{i=1}^{n} \sum_{K=1}^{K} \mathbb{1}(i,K) \log\left( \frac{e^{w_K^T X_i}}{\sum_{h=1}^{K} e^{w_h^T X_i}} \right)$$

ت)

$$\nabla_{w_j} L(w_1, \dots, w_K) = \sum_{i=1}^{n} \nabla_{w_j} \sum_{K=1}^{K} \mathbb{1}(i,K) \log\left( \frac{e^{w_K^T X_i}}{\sum_{h=1}^{K} e^{w_h^T X_i}} \right)$$

$$= \sum_{i=1}^{n} \mathbb{1}(i,j) \frac{X_i \sum_{h=1}^{K} e^{w_h^T X_i} - X_i e^{w_j^T X_i}}{\sum_{h=1}^{K} e^{w_h^T X_i}} + (1 - \mathbb{1}(i,j)) \frac{-X_i e^{w_i^T X_i}}{\sum_{h=1}^{K} e^{w_h^T X_i}}$$

$$= \sum_{i=1}^{n} X_i \left( \mathbb{1}(i,j) - \frac{e^{w_i^T X_i}}{\sum_{h=1}^{K} e^{w_h^T X_i}} \right)$$

ث)

$$f(w_1, \dots, w_{k-1}) = L(w_1, \dots, w_{k-1}) - \frac{\lambda}{2} \Sigma_{j=1}^{K-1} ||w_j||_2^2$$

$$\nabla_{w_K} \mathbb{1}(w_1, \dots, w_{K-1}) = \nabla_{w_K} L(w_1, \dots, w_{K-1}) - \frac{\lambda}{2} \sum_{j=1}^{K-1} \nabla_{w_K} ||w_j||_2^2$$

$$\frac{\lambda}{2} \sum_{j=1}^{K-1} \nabla_{w_K} ||w_j||_2^2 = \frac{\lambda}{2} \times 2 \times w_K = \lambda w_K$$

$$\nabla_{w_K} \mathbb{1}(w_1, \dots, w_{K-1}) = \sum_{i=1}^{n} X_i \left( \mathbb{1}(i,j) - \frac{e^{w_i^T X_i}}{\sum_{h=1}^{K} e^{w_h^T X_i}} \right) + \lambda w_K$$

ستون ۱ به آن اضافه شده است

۴) الف)

$$E(w) = \frac{1}{2}\|Xw - y\|_2^2 = \frac{1}{2}(Xw-y)^T(Xw-y)$$

$$\nabla E(w) = 0 \longrightarrow \nabla E(w) = \frac{1}{2} \times 2\, X(Xw-y) = X^TXw - X^Ty = 0$$

$$w = (X^TX)^{-1}X^Ty$$

اگر فقط بر روی یک ویژگی مدل را آموزش دهیم $X$ ما فقط یک ستون دارد ($x_j$) و معادله بالا به صورت زیر در می‌آید:

$$w_j = (x_j^T x_j)^{-1} x_j^T y = \frac{x_j^T y}{x_j^T x_j}$$

ب) $x_i$ و $x_j$ باهم متعامدند:

$$x_i^T x_j = 0$$

$$X^TX = \begin{bmatrix} x_1^Tx_1 & 0 & 0 & \cdots & 0 \\ 0 & x_2^Tx_2 & 0 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & x_m^Tx_m \end{bmatrix} \qquad (X^TX)^{-1} = \begin{bmatrix} \frac{1}{x_1^Tx_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{x_2^Tx_2} & 0 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{x_m^Tx_m} \end{bmatrix} \qquad X^Ty = \begin{bmatrix} x_1^Ty \\ \vdots \\ x_m^Ty \end{bmatrix}$$

$$w_j = \frac{cov[x_j, y]}{var[x_j]}$$

$$w_. = E[y] - w_j E[x_j]$$

$$w_j = \left[(X^TX)^{-1}X^Ty\right]_j = \frac{1}{x_j^Tx_j} \times x_j^Ty \longrightarrow$$

همانند بالا به دست آمد.

پ)

$$X = [1 \;\; x_1] \qquad X^TX = \begin{bmatrix} N & \sum x_1^i \\ \sum_i x_1^i & \sum_i (x_1^i)^2 \end{bmatrix} \qquad X^Ty = \begin{bmatrix} \sum_i y_i \\ \sum x_1^i y_i \end{bmatrix}$$

$$w = \frac{1}{\underbrace{N\sum_i (x_1^i)^2 - (\sum_i x_1^i)^2}_{N^2 var(x_1)}} \begin{bmatrix} \sum (x_1^i)^2 & -\sum x_1^i \\ -\sum x_1^i & N \end{bmatrix} \begin{bmatrix} \sum y_i \\ \sum x_1^i y_i \end{bmatrix}$$

$$-\frac{1}{N}\left(\sum x_1^i\right)^2 \sum y_i + \frac{1}{N}\left(\sum x_1^i\right)^2 \sum y_i$$

$$= \frac{1}{N^2 var(x_1)} \begin{bmatrix} \sum (x_1^i)^2 \sum y_i - \sum x_1^i \sum x_1^i y_i \\ \underbrace{N\sum x_1^i y_i - \sum x_1^i \sum y_i}_{N^2 cov(x_1, y)} \end{bmatrix} = \begin{bmatrix} E[y_i] - w_1 E[x_1] \\ \frac{cov(x_1, y)}{var(x_1)} \end{bmatrix} \begin{matrix} \rightarrow w_0 \\ \rightarrow w_1 \end{matrix}$$

خواسته سوال اثبات شد.