

کسری صید ۹۹۳۹۲۳۰۳۰
تربیات - آریزیا هاسین سیدی

(الف)

$$P(w|D) = \frac{P(w) P(D|w)}{P(D)} \rightarrow \text{وابسته به } w \text{ نیست}$$

$$\text{loss} = -\ln(P(w|D)) = -\ln(P(w)) = -\ln(P(D|w))$$

$$\ln(P(t_1, t_2, t_3, \dots, t_n | w)) = \ln(\prod P(t_i | w))$$

$$\sum \ln(P(t_i | w)) = \sum t_i \ln(\sigma(w^T x + b)) + (1 - t_i) \ln(1 - \sigma(w^T x + b))$$

$$\rightarrow \text{loss} = \frac{1}{N} (w^T (\alpha^{-1} I)^{-1} w) + \ln(\sqrt{2\pi\epsilon}) - \ln(\sqrt{N\epsilon}) + (1 - t_i) \ln(1 - \sigma(w^T x + b)) - \sum (\ln(t_i (\sigma(w^T x + b))))$$

$$\rightarrow \frac{\partial \text{loss}}{\partial b} = -\sum (t_i (1 - \sigma(w^T x + b)) - (1 - t_i) \sigma(w^T x + b))$$

$$\frac{\partial \text{loss}}{\partial w} = -\sum (t_i (1 - \sigma(w^T x + b)) + \frac{1}{N} (\alpha^{-1} I)^{-1} w - (1 - t_i) \sigma(w^T x + b)) x_i$$

$$P(w|D) = \frac{P(w)P(D|w)}{P(D)}$$

(۴)

$$\arg \max P(w|D) = \arg \max \log P(w|D) = \arg \max (\log P(w) + \log P(D|w))$$

$$\log P(D|w) = \log \prod_{i=1}^{|D|} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y_i - w^T x_i)^2}{2\sigma^2}\right) = -\frac{1}{2} |D| \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{|D|} (y_i - w^T x_i)^2$$

$$\log P(w|D) = -\frac{1}{2} |D| \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{|D|} (y_i - w^T x_i)^2 - \frac{1}{2\sigma^2} \sum_{j=1}^{|w|} w_j^2 - \frac{1}{2} |w| \log(2\pi\sigma^2)$$

$$\frac{\partial \log P(w|D)}{\partial w} = -\frac{1}{\sigma^2} \sum_{j=1}^{|w|} w_j - \frac{1}{\sigma^2} \sum_{i=1}^{|D|} (y_i - w^T x_i) x_i = 0 \rightarrow$$

$$\rightarrow w_i = \left(\sum_{i=1}^n y_i x_i \right) \left(\sum_{i=1}^n x_i^T x_i \right)^{-1}$$

$$y_i = a_1 \log x_1 + a_2 \log e + \epsilon_i$$

(۵) افق درست

می‌توان با رگرسیون خطی، پارامترهای مدل را به دست آورد

ب) غلط به دلیل اینکه مدل به صورت $y_i = a_1 x_1 + a_2 + \epsilon_i$ است و این مدل یک مدل غیر خطی است
می‌توان از رگرسیون خطی برای به دست آوردن پارامترهای این مدل استفاده کرد و مدل ساده‌تری شود که مدل خطی
تبدیل شود

(۶)

در مدل B چون یک چند صلبی درجه ۵ است وقت آن از مدل A که چند صلبی درجه ۲ است بیشتر است ولی احتمال overfit در این مدل بیشتر است. برای انتخاب مدل بهتر ابتدا عدد مدل را با داده های Train آموزش می دهیم تا W ها به دست آیند. حال بعد از آموزش مدل ها را با داده های Test آزمایش می کنیم و معیار دقت accuracy-score را برای انتخاب می آوریم. هر چقدر که برای accuracy-score بالاتری باشد را برای مدل بهتر برای داده های آموزشی می گیریم. برای انتخاب بهتر نیز می توانیم معیارهای f_1 -score، recall-score و precision-score را محاسبه کنیم و از روی این نتایج تصمیم بگیریم و مدل بهتر را انتخاب کنیم. در اینجا چون تعداد نمونه ها کم است احتمالاً مدل A مدل بهتری برای این مسئله باشد.