STAT409 Homework4
2021320322 윤민서

1.

(a)

Likelihood $L(\beta)=\prod_{i=1}^{n}\{p_\beta(x_i)\}^{y_i}\cdot\{1-p_\beta(x_i)\}^{1-y_i}$

$$\begin{aligned}
l(\beta) \triangleq \log L(\beta) &= \log \prod_{i=1}^{n}\{p_\beta(x_i)\}^{y_i}\cdot\{1-p_\beta(x_i)\}^{1-y_i} \\
&= \sum_{i=1}^{n} \log \{p_\beta(x_i)\}^{y_i}\cdot\{1-p_\beta(x_i)\}^{1-y_i} \\
&= \sum_{i=1}^{n}\left[y_i \log p_\beta(x_i) + (1-y_i)\log\{1-p_\beta(x_i)\}\right] \\
&= \sum_{i=1}^{n}\left\{y_i \log \frac{\exp(\beta^T x_i)}{1+\exp(\beta^T x_i)} + (1-y_i)\log\frac{1}{1+\exp(\beta^T x_i)}\right\} \\
&= \sum_{i=1}^{n}\left[y_i(\beta^T x_i) - y_i \log\{1+\exp(\beta^T x_i)\} - (1-y_i)\log\{1+\exp(\beta^T x_i)\}\right] \\
&= \sum_{i=1}^{n}\left[y_i(\beta^T x_i) - \log\{1+\exp(\beta^T x_i)\}\right]
\end{aligned}$$

(b)

$l_i(\beta) = y_i \log p_\beta(x_i) + (1-y_i)\log\{1-p_\beta(x_i)\}$ 라고 하면

$$\frac{\partial l_i(\beta)}{\partial \beta} = \frac{\partial(\beta^T x_i)}{\partial\beta}\frac{\partial p_\beta(x_i)}{\partial(\beta^T x_i)}\frac{\partial l_i(\beta)}{\partial p_\beta(x_i)} \text{ by chain rule}$$

$$\frac{\partial l_i(\beta)}{\partial p_\beta(x_i)} = \frac{y_i}{p_\beta(x_i)} - \frac{1-y_i}{1-p_\beta(x_i)} = \frac{y_i - p_\beta(x_i)}{p_\beta(x_i)\{1-p_\beta(x_i)\}}$$

$$\begin{aligned}
\frac{\partial p_\beta(x_i)}{\partial(\beta^T x_i)} &= \frac{\partial}{\partial(\beta^T x_i)}\frac{\exp(\beta^T x_i)}{1+\exp(\beta^T x_i)} = \frac{\exp(\beta^T x_i)\{1+\exp(\beta^T x_i)\} - \{\exp(\beta^T x_i)\}^2}{\{1+\exp(\beta^T x_i)\}^2} \\
&= \frac{\exp(\beta^T x_i)}{\{1+\exp(\beta^T x_i)\}^2} = \frac{\exp(\beta^T x_i)}{1+\exp(\beta^T x_i)}\frac{1}{1+\exp(\beta^T x_i)} \\
&= p_\beta(x_i)\{1-p_\beta(x_i)\} \cdots ⓐ
\end{aligned}$$

$$\frac{\partial(\beta^T x_i)}{\partial\beta} = \frac{\partial}{\partial\beta}(\beta_1 x_{i1} + \cdots + \beta_p x_{ip}) = \begin{bmatrix} \frac{\partial}{\partial\beta_1}(\beta_1 x_{i1}+\cdots+\beta_p x_{ip}) \\ \vdots \\ \frac{\partial}{\partial\beta_p}(\beta_1 x_{i1}+\cdots+\beta_p x_{ip})\end{bmatrix} = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip}\end{bmatrix} = x_i$$

$$\Rightarrow \frac{\partial l_i(\beta)}{\partial\beta} = x_i(y_i - p_\beta(x_i))$$

$$\frac{\partial l(\beta)}{\partial\beta} = \sum_{i=1}^{n} x_i(y_i - p_\beta(x_i)) = X^T(y - p(\beta))$$

$$\frac{\partial^2 l_i(\beta)}{\partial\beta\partial\beta^T} = \frac{\partial}{\partial\beta^T}\{x_i(y_i - p_\beta(x_i))\} = -x_i\frac{\partial(p_\beta(x_i))}{\partial\beta^T}$$

$$= -x_i\frac{\partial(\beta^T x_i)}{\partial\beta^T}\frac{\partial p_\beta(x_i)}{\partial(\beta^T x_i)} \text{ by chain rule}$$

$$\frac{\partial(\beta^T x_i)}{\partial\beta^T} = x_i^T \Rightarrow \frac{\partial^2 l_i(\beta)}{\partial\beta\partial\beta^T} = -x_i x_i^T p_\beta(x_i)\{1-p_\beta(x_i)\} \text{ by (a)}$$

$$\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta^T} = -X^TWX$$

$$\left(\because X^TWX = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix}\begin{bmatrix} p_\beta(x_1)\{1-p_\beta(x_1)\} & & 0 \\ & \ddots & \\ 0 & & p_\beta(x_n)\{1-p_\beta(x_n)\} \end{bmatrix}\begin{bmatrix} - x_1^T - \\ \vdots \\ - x_n^T - \end{bmatrix}\right)$$

2.

(a)

primal function을 Lp라고 하면

$$\frac{\partial}{\partial\beta}L_p = \beta - \sum_{i=1}^n \alpha_i y_i x_i \overset{set}{=} 0 \Rightarrow \beta = \sum_{i=1}^n \alpha_i y_i x_i \quad \left(\because \frac{\partial(\beta^T x_i)}{\partial\beta} = x_i\right)$$

$$\frac{\partial}{\partial\beta_0}L_p = -\sum_{i=1}^n \alpha_i y_i \overset{set}{=} 0 \Rightarrow \sum_{i=1}^n \alpha_i y_i = 0$$

$$\frac{\partial}{\partial\xi_i}L_p = C - \alpha_i - \gamma_i \overset{set}{=} 0 \Rightarrow \alpha_i = C - \gamma_i$$

(b)

(a)의 결과를 Lp에 대입하면 dual function $g(\alpha,\gamma)$는

$$g(\alpha,\gamma) = \frac{1}{2}\sum_{i=1}^n(\alpha_i y_i x_i^T)\sum_{j=1}^n(\alpha_j y_j x_j) + \underbrace{C\sum_{i=1}^n \xi_i}_{①} + \sum_{i=1}^n \alpha_i - \beta_0 \cancel{\sum_{i=1}^n \alpha_i y_i} - \sum_{i=1}^n \alpha_i y_i\Big(\sum_{j=1}^n \alpha_j y_j \underbrace{x_j^T}\Big)x_i - \underbrace{\sum_{i=1}^n(\alpha_i+\gamma_i)\xi_i}_{②}$$

$$= \sum_{i=1}^n \alpha_i - \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \alpha_i\alpha_j y_i y_j x_i^T x_j \Rightarrow \text{objective function}$$

$x_j^T x_i = x_i^T x_j$

$① + ② = 0$

(a)에서 $\sum_{i=1}^n \alpha_i y_i = 0$

KKT의 dual feasible 조건에 의하여

$\alpha_i \ge 0,\ \gamma_i = C - \alpha_i \ge 0$을 연립하면

$0 \le \alpha_i \le C$

따라서 dual problem:

$$\underset{\alpha_1,\cdots,\alpha_n}{\text{maximize}} \ \sum_{i=1}^n \alpha_i - \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \alpha_i\alpha_j y_i y_j x_i^T x_j$$

$$\text{subject to } 0 \le \alpha_i \le C,\ i=1,\cdots,n$$

$$\sum_{i=1}^n \alpha_i y_i = 0$$