Yash Butala
17CS30038

Q1.1) $P(x|\theta) = \theta^x (1-\theta)^{1-x}$

$v = g(\theta) = 1-\theta$

$\rightarrow P(x|v) = \theta^x (1-\theta)^{1-x}$

$P(x|v) = (1-v)^x (v)^{1-x}$

$f(x_1, x_2, \dots x_n | v) = \prod_i (1-v)^{x_i} (v)^{1-x_i}$

$= (1-v)^{\Sigma x_i} (v)^{\Sigma 1-x_i}$

$\therefore Log(L(v)) = \Sigma x_i \log(1-v) + (n - \Sigma x_i) \log v$

$\frac{\partial Log(L(v))}{\partial v} = 0$

$\Rightarrow \frac{-\Sigma x_i}{(1-v)} + \frac{n - \Sigma x_i}{v} = 0$

$\Rightarrow (n - \Sigma x_i)(1-v) = (v)(\Sigma x_i)$

$\Rightarrow n - \Sigma x_i + \cancel{v \Sigma x_i} - nv = \cancel{v \Sigma x_i}$

$\Rightarrow v = \frac{n - \Sigma x_i}{n}$

$= 1 - \frac{\Sigma x_i}{n}$

$\therefore v_{MLE} = 1 - \frac{\Sigma x_i}{n}$

Also, v_{MLE} using functional invariance property $= 1 - \hat{\theta} = 1 - \frac{\Sigma x_i}{n}$.

Hence we find v_{MLE} and verify the property.

Yash Butala
17CS30038

Q1.2) $P(x|\theta) = \frac{1}{2} e^{-|x-\theta|}$

$f(x_1, x_2, \dots x_{2n}|\theta) = \prod f(x_i|\theta)$ $\quad$ ($\because$ all x_i are iid)

$= \left(\frac{1}{2}\right)^{2n} e^{-\sum_i |x_i - \theta|}$

$\therefore \log(L(\theta)) = 2n\log(1/2) - \sum_{i=1}^{2n} |x_i - \theta|$

$\frac{\partial \log(L)}{\partial \theta} = 0$

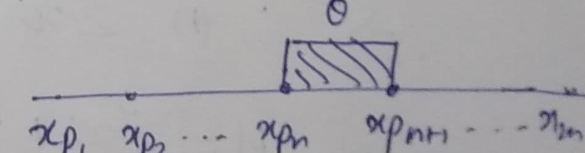
$\Rightarrow \frac{\partial}{\partial \theta}\left(2n(\log(1/2)) - \sum_{i=1}^{2n} |x_i - \theta|\right) = 0$

$\Rightarrow \sum_{i=1}^{2n} \mathrm{sgn}(x_i - \theta) = 0$ $\quad$ using $\frac{\partial}{\partial z}|z| = \mathrm{sgn}(z)$

$\therefore$ (no. of $x_i > \theta$) = (no. of $x_i < \theta$)

Let $p_1, p_2, \dots p_{2n}$ be permutation such that $\{x_{p_i}\}$ is sorted in non decreasing order.

$\therefore \theta_{MLE} \in [x_{p_n}, x_{p_{n+1}}]$



Yash Butala
17CS30038

Q2.1) $P(y_i | x_i, d_1, d_2, \theta) = \dfrac{z(d_1, d_2)\, e^{d_1 (y_i - \theta^T x_i)}}{\left(d_1 e^{2(y_i - \theta^T x_i)} + d_2\right)^{(d_1+d_2)/2}}$

$$L_i(\theta) = \log(z(d_1, d_2)) + d_1 (y_i - \theta^T x_i) - \left(\frac{d_1+d_2}{2}\right) \log\left(d_1 e^{2(y_i - \theta^T x_i)} + d_2\right)$$

$$\frac{\partial L_i}{\partial \theta} = -d_1 x_i - \left(\frac{d_1+d_2}{2}\right)\left(\frac{1}{d_1 e^{2(y_i - \theta^T x_i)} + d_2}\right)\left(d_1 e^{2(y_i - \theta^T x_i)}\right)(-2x_i)$$

(using rules of derivatives for vectors)

$$= \frac{-d_1 d_2 x_i - \cancel{d_1^2 e^{2(y_i - \theta^T x_i)} x_i} + \cancel{d_1^2 x_i e^{2(y_i - \theta^T x_i)}} + d_1 d_2 x_i e^{2(y_i - \theta^T x_i)}}{d_1 e^{2(y_i - \theta^T x_i)} + d_2}$$

$$= \frac{(d_1 d_2 x_i)\left[e^{2(y_i - \theta^T x_i)} - 1\right]}{d_1 e^{2(y_i - \theta^T x_i)} + d_2}$$

Yash Butala
17CS30038

Q3.1) $P(y_i = 1 \mid \theta, x_i) = ?$

$\theta \sim N(0, \sigma_\theta)$

$\epsilon \sim \text{Logistic}(0, \sigma_\epsilon)$

$y = 1[X\theta + \epsilon \geq 0]$ (boolean vector)

$$\therefore P(y_i = 1 \mid \theta, x_i) = P(\theta^T x_i + \varepsilon \geq 0)$$

$$= 1 - P(\cancel{\theta^T x_i}\ \varepsilon_i < -\theta^T x_i) \quad \text{—} \ (1)$$

As ε follows Logistic distribution, cumulative distribution function is logistic function

$$F_{\text{logistic}} = \frac{1}{1 + e^{-x/\sigma_\epsilon}} \quad \text{—} \ (2)$$

From (1), (2)

$$P(y_i = 1 \mid \theta, x_i) = 1 - F_{\text{logistic}}(-\theta^T x_i, \sigma_\epsilon)$$

$$= \left(1 - \frac{1}{1 + e^{\theta^T x_i / \sigma_\epsilon}}\right)$$

$$= \left(\frac{e^{\theta^T x_i / \sigma_\epsilon}}{1 + e^{\theta^T x_i / \sigma_\epsilon}}\right)$$

$$= \left(\frac{1}{1 + e^{-\theta^T x_i / \sigma_\epsilon}}\right)$$

$$\therefore P(y_i = 1 \mid \theta, x_i) = \text{Logistic}\left(\frac{\theta^T x_i}{\sigma_\epsilon}\right)$$

Yash Butala
17CS30038

Q 3.2) $P(y_i|\theta, x_i) = ?$

Given : $y_i = 1[\theta^T x_i + \varepsilon_i \geq 0]$

$\rightarrow P(y_i|x_i,\theta) = P(y_i=1|x_i,\theta)^{y_i} + P(y_i=0|x_i,\theta)^{1-y_i}$ — ①

(As y is boolean function)

$P(y_i=1|x_i,\theta) = \text{logistic}(\theta^T x_i)$

$P(y_i=0|x_i,\theta) = 1 - \text{logistic}(\theta^T x_i)$

② (using result in 3.1)

From ① and ②,

$\rightarrow \therefore P(y_i|x_i,\theta) = [\text{logistic}(\theta^T x_i)]^{y_i} [1-\text{logistic}(\theta^T x_i)]^{1-y_i}$

Hence proved.

Q 3.3) $\log P(y_i|x_i,\theta)$

$= \log\left(\text{logistic}(\theta^T x_i)^{y_i}(1-\text{logistic}(\theta^T x_i))^{1-y_i}\right)$ ...proved in 3.2

$= \log\left[\left(\frac{1}{1+e^{-\theta^T x_i}}\right)^{y_i}\left(1-\frac{1}{1+e^{-\theta^T x_i}}\right)^{1-y_i}\right]$

$= \log\left[\left(\frac{e^{\theta^T x_i}}{1+e^{\theta^T x_i}}\right)^{y_i}\left(\frac{e^{-\theta^T x_i}}{1+e^{-\theta^T x_i}}\right)^{1-y_i}\right]$

$= \log\left[\left(\frac{e^{\theta^T x_i}}{1+e^{\theta^T x_i}}\right)^{y_i}\left(\frac{1}{1+e^{\theta^T x_i}}\right)^{1-y_i}\right]$

$$= \log \left[\frac{\left(e^{\theta^T x_i}\right)^{y_i}}{1 + e^{\theta^T x_i}} \right]$$

$$= y_i \theta^T x_i - \log\left(1 + e^{\theta^T x_i}\right)$$

Hence proved.

3.4) $L_{MLE}(\theta) = \sum_i \log P(y_i \mid \theta, x_i)$

$$= \sum_{i=1}^{n} \left[y_i \theta^T x_i - \log\left(1 + e^{\theta^T x_i}\right) \right]$$

$$= \sum_{i=1}^{n} y_i \left(\theta_1 x_{i1} + \theta_2 x_{i2} + \dots + \theta_d x_{id}\right) - \sum_{i=1}^{n} \log\left(1 + e^{\theta^T x_i}\right)$$

$$= \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} \begin{bmatrix} \theta_1 x_{11} + \theta_2 x_{12} + \dots + \theta_d x_{1d} \\ \vdots \\ \theta_1 x_{n1} + \theta_2 x_{n2} + \dots + \theta_d x_{nd} \end{bmatrix} - 1_{n\times 1}^T \begin{bmatrix} \log\left(1 + e^{\theta^T x_1}\right) \\ \log\left(1 + e^{\theta^T x_2}\right) \\ \vdots \\ \log\left(1 + e^{\theta^T x_n}\right) \end{bmatrix}$$

$$= y^T X \theta - 1_{n\times 1}^T \log \begin{bmatrix} 1 + e^{\theta^T x_1} \\ 1 + e^{\theta^T x_2} \\ \vdots \\ 1 + e^{\theta^T x_n} \end{bmatrix}$$

$$= y^T X \theta - 1_{n\times 1}^T \log \left(1_{n\times 1} + \begin{bmatrix} e^{x_1^T \theta} \\ e^{x_2^T \theta} \\ \vdots \\ e^{x_n^T \theta} \end{bmatrix} \right)$$

Yash Butala
17CS30038

$$= y^T X\theta - 1^T_{n\times1} \log(1_{n\times1} + e^{X\theta})$$

Hence proved.

3.5) $$\frac{\partial L_{MLE}(\theta)}{\partial\theta} = \frac{\partial}{\partial\theta}\left(y^T X\theta - 1^T_{n\times1}\log(1_{n\times1} + e^{X\theta})\right) \rightarrow \text{(using 3.4)}$$

$$= \frac{\partial}{\partial\theta}\left(\sum_{i=1}^{n}\left(y_i\theta^T x_i - \log(1+e^{\theta^T x_i})\right)\right) \longrightarrow \text{(from 3.3)}$$

$$= \sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta}\left(\underbrace{y_i}_{1\times1}\underbrace{\theta^T}_{d\times1}\underbrace{x_i}_{d\times1} - \log(\underbrace{1+e^{\theta^T x_i}}_{1\times1})\right)\right)$$

$$= \sum_{i=1}^{n}\left(y_i x_i - \frac{e^{\theta^T x_i} x_i}{1+e^{\theta^T x_i}}\right) \longrightarrow \text{using } \frac{\partial a^T b}{\partial a} = b$$

$$= \sum_{i=1}^{n} x_i\left(y_i - \frac{e^{\theta^T x_i}}{1+e^{\theta^T x_i}}\right)$$

$$= \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & & \vdots \\ x_{1d} & x_{2d} & \cdots & x_{nd} \end{bmatrix}_{d\times n} \left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} e^{\theta^T x_1}/(1+e^{\theta^T x_1}) \\ \vdots \\ e^{\theta^T x_n}/(1+e^{\theta^T x_n}) \end{bmatrix} \right)$$

$$= X^T\left(y - \begin{bmatrix} \text{logistic}(\theta^T x_1) \\ \vdots \\ \text{logistic}(\theta^T x_n) \end{bmatrix}_{n\times1}\right)$$

$$= X^T(y - \text{logistic}(X\theta))$$

Hence proved.