Deepak Meweda | 20CS91P02

1.1

Compute the MLE for the parameter ν of the distribution without using functional invariance property

Consider $\nu = g(\theta) = (1-\theta)$

Given $P(x|\nu) = \theta^x (1-\theta)^{1-x}$ for $x \in \{0,1\}$

$\Rightarrow$ Since $\theta = 1-\nu$ given so

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

$$= (1-\nu)^x (\nu)^{1-x}$$

Now, $$L_N(\nu) = \prod_{i=1}^{n} P(x_i|\nu)$$

$$= \prod_{i=1}^{n} (1-\nu)^{x_i} \nu^{1-x_i}$$

taking log of likelihood $L_N(\nu)$

$$\log L_N(\nu) = \sum_i x_i \log(1-\nu) + (n - \sum x_i) \log \nu$$

To maximise, we set partial diff w.r.t. ν to 0

Hence $$\frac{\partial}{\partial \nu} \mathcal{L}(\nu) \Rightarrow \frac{-\sum_i x_i}{1-\nu} + \frac{n - \sum x_i}{\nu} = 0$$

$$\Rightarrow \frac{-\nu \sum x_i + (1-\nu)(n - \sum_i x_i)}{(1-\nu)\nu} = 0$$

$$\Rightarrow -\nu \sum_i x_i + n - n\nu - \sum_i x_i + \nu \sum x_i = 0$$

$$\Rightarrow n\nu = n - \sum x_i$$

$$\Rightarrow \boxed{\nu = 1 - \frac{\sum x_i}{n}}$$

maximum likelihood estimation.

1.2

Given $p(x|\theta) = \frac{1}{2} e^{-|x-\theta|}$

$$L(\theta) = \prod_{i=1}^{2m} \left(\frac{1}{2} e^{-|x_i-\theta|}\right)$$

log likelihood

$$l(\theta) = \sum_{i=1}^{2m} \left[\log\frac{1}{2} + \log e^{-|x_i-\theta|}\right]$$

$$= \sum_{i=1}^{2n} \left[-\log 2 - |x_i-\theta|\right]$$

$$= -2n\log 2 - \sum_{i=1}^{2n} |x_i-\theta|\} \longrightarrow \text{(1)}$$

as its a mod function, it's not differentiable

So here to find Maximum in mod function two case arise, for $i = 1$ to P

Case (i) when $P = 2m+1$, $\hat{\theta}_{MLE} = X_{(m+1)}$ which is unique MLE

Case 2 (ii) when $P = 2m$

$$\hat{\theta}_{MLE} \in [X_m, X_{m+1}]$$

Hence $\hat{\theta}_{MLE} = \text{Median}(x_i)$

Here in our problem (1)

It follows case (ii) as $2m$ is even

So here $\hat{\theta}_{MLE} = \text{median of}(x_i)$

$$\hat{\theta}_{MLE} = \frac{X_n + X_{n+1}}{2}$$

2.1

$$P(y_i|x_i,\theta,d_1,d_2) = z(d_1,d_2)\,\frac{e^{d_1(y_i-\theta^T x_i)}}{(d_1 e^{2(y_i-\theta^T x_i)}+d_2)^{\frac{d_1+d_2}{2}}}$$

$$L_i(\theta) = P(y_i|x_i,\theta,d_1,d_2)$$

taking log of likelihood

$$l_i(\theta) = \log P(y_i|x_i,\theta,d_1,d_2)$$

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

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

to find $\frac{d l_i(\theta)}{d\theta}$, differentiating $l_i(\theta)$ w.r.t. θ

$$\frac{d l_i(\theta)}{d(\theta)} = 0 + d_1(0-x_i) - \left(\frac{d_1+d_2}{2}\right)\left(\frac{d_1 e^{2(y_i-\theta^T x_i)}\times 2\times -x_i}{d_1 e^{2(y_i-\theta^T x_i)}+d_2}\right)$$

$\because$ since z is constant

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

$$\boxed{\frac{d l_i(\theta)}{d\theta} = \left(-d_1 + (d_1+d_2) d_1 \frac{e^{2(y_i-\theta^T x_i)}}{d_1 e^{2(y_i-\theta^T x_i)}+d_2}\right) x_i}$$

3.1 To prove that

$$P(y_i | \theta_i, x_i) = \text{logistic}\left(\frac{\theta^T x_i}{\sigma_\epsilon}\right)$$

Now $\epsilon_i \sim \text{logistic}(0, \sigma_\epsilon)$

We know that CDF is the probability that the variable takes a value less than or equal to x

Thus $\text{CDF}(\theta^T x_i) = 1 - \text{logistic}\left(\frac{\theta^T x_i}{\sigma_\epsilon}\right)$ — (1)

Now $P(y=1 | \theta, x_i) = P(\theta^T x_i + \epsilon > 0)$

$$= P(\epsilon > -\theta^T x_i)$$

$$= P(\epsilon \leq \theta^T x_i)$$

$$= 1 - \text{CDF}(\theta^T x_i)$$

$$= 1 - \left[1 - \text{logistic}\left(\frac{\theta^T x_i}{\sigma_\epsilon}\right)\right] \text{ from (1)}$$

$$\Rightarrow \boxed{P(y_i = 1 | \theta, x_i) = \text{logistic}\left(\frac{\theta^T x_i}{\sigma_\epsilon}\right)}$$

(3.2) Using logistic Reg. classifier we have

$$\text{logistic}(z) = \sigma(z) = \frac{1}{1+e^{z}}$$

Also $y = \sigma(h_\theta(x)) = \sigma(\theta^T x)$

from 3.1

$$P(y_i = 1 \mid \theta; x_i) = \sigma(\theta^T x_i) \quad \text{— (1)}$$

$$P(y_i = 0 \mid \theta; x_i) = 1 - \sigma(\theta^T x_i) \quad \text{— (2)}$$

Thus we know y is binary given to us

$$P(y_i \mid x_i, \theta) = \sigma(\theta^T x_i)^{y_i} (1 - \sigma(\theta^T x_i))^{1-y_i}$$

Since $\sigma_t = 1$

$$\sigma(\theta^T x_i) = \frac{1}{1-e^{\theta^T x_i}}$$

or $\sigma(\theta^T x_i) = \text{logistic}(\theta^T x_i)$

hence

$$\boxed{P(y_i \mid x_i; \theta) = \text{logistic}(\theta^T x_i)^{y_i} (1 - \text{logistic}(\theta^T x_i))^{1-y_i}}$$

— (3.2)

(3.3) from eqn (3.2)

taking log of both sides

$$\log P(y_i|\theta, x_i) = y_i[\log 1 - \log(1+e^{-\theta^T x_i})] + (1-y_i)[\log e^{-\theta^T x_i} - \log(1+e^{-\theta^T x_i})]$$

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

$$\log P(y_i|\theta, x_i) = \boxed{y_i \theta^T x_i - \log(1+e^{\theta^T x_i})}$$

Proved

log

3.4 Taking summation of both side for all i in result of eqn from question 3.3

$$\mathcal{L}_{MLE}(\theta) = \log(y \mid X, \theta) = \sum_i y_i \theta^T x_i - \sum_i \log 1 + e^{\theta^T x}$$

$$\Rightarrow \boxed{\mathcal{L}_{MLE}(\theta) = y^T X \theta - 1_{n \times 1} \cdot \log(1_{n \times 1} + e^{X\theta})}$$

(3.5)

Partial derivative w.r.t θ on both RHS and LHS for 3.4 we get

$$\frac{\partial \mathcal{L}_{MLE}(\theta)}{\partial \theta} = X^T y - \frac{1}{1+e^{X\theta}}$$

$$\boxed{\frac{\partial \mathcal{L}_{MLE}(\theta)}{\partial \theta} = X^T\left(y - \text{logistic}(X\theta)\right)}$$

hence proved