COMP9417 – Machine Learning

Tutorial: Regression II

Wanqing Yang - z5325987

Question 1.

(a) Assume that $N \sim (\mu,\ \sigma^2)$ and $\theta = (\mu, \sigma)$, $X = (X_1, X_2 \dots X_n)$

Then we need to calculate the maximum value of θ

$\hat{\theta}_{MLE} = arg \max_{\theta \in \Theta} logL(\theta)$

Then $\hat{\theta}_{MLE} = arg \max_{\theta \in \Theta} logL(\theta) = arg \max_{\theta \in \Theta} log \prod_{\mathrm{i=1}}^{\mathrm{n}} P_\theta(X_i)$

Because $P_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right)$

Then $log \prod_{\mathrm{i=1}}^{\mathrm{n}} P_\theta(X_i) = \sum_{\mathrm{i=1}}^{\mathrm{n}} logP_\theta(X_i) = \sum_{\mathrm{i=1}}^{\mathrm{n}} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right)$

$$= \sum_{\mathrm{i=1}}^{\mathrm{n}} \log \frac{1}{\sqrt{2\pi\sigma^2}} - \left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right)$$

Then $\hat{\mu}_{MLE} = arg \max_{\mu} logL(\theta) = arg \max_{\mu} \sum_{i=0}^{\mathrm{n}} \left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right)$

$$= arg \min_{\mu} \sum_{i=0}^{\mathrm{n}} \left(\frac{(X_i-\mu)^2}{2\sigma^2}\right) = arg \min_{\mu} \sum_{i=0}^{\mathrm{n}} (X_i - \mu)^2$$

Let $\frac{\partial \sum_{i=0}^{\mathrm{n}} (X_i-\mu)^2}{\partial \mu} = 0$

Then $\hat{\mu}_{MLE} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i=0}}^{\mathrm{n}} X_i = \overline{\mathrm{x}}$

$$\widehat{\sigma^2}_{MLE} = arg \max_{\sigma} logL(\theta) = arg \max_{\sigma} \sum_{\mathrm{i=0}}^{\mathrm{n}} \left(-log\sigma - \frac{1}{2\sigma^2}(X_i, \mu)^2\right)$$

Let $\frac{\partial \sum_{\mathrm{i=0}}^{\mathrm{n}} \left(-log\sigma - \frac{1}{2\sigma^2}(X_i,\mu)^2\right)}{\partial \sigma} = 0$ Then $\sum_{\mathrm{i=0}}^{\mathrm{n}} \sigma^2 = \sum_{\mathrm{i=0}}^{\mathrm{n}} (X_i - \mu)^2$

Then $\widehat{\sigma^2}_{MLE} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i=0}}^{\mathrm{n}} (X_i - \hat{\mu}_{MLE})^2$

Therefore, $\hat{\mu}_{MLE} = \sum_{\mathrm{i=0}}^{\mathrm{n}} X_i$ …… ① $\widehat{\sigma^2}_{MLE} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i=0}}^{\mathrm{n}} (X_i - \hat{\mu}_{MLE})^2$ …… ②

Because neither equation ① and ② contains σ^2

Therefore, $X_1, X_2, \dots X_n \sim N(\mu, 1)$ gives exactly the same answer as solving the original problem.

(b) Because $P(X = k) = P^k(1-p)^{1-k}$ $k = 0,1$ $p \in [0,1]$

Then $P_\theta(X) = p^X(1-p)^{1-\mathrm{X}}$ $X = 0,1$ $p \in [0,1]$

$$\hat{\mathrm{p}}_{MLE} = arg\max_{\mathrm{p}} logL(\mathrm{p}) = arg\max_{\mathrm{p}} log\prod_{\mathrm{i}=0}^{\mathrm{n}} p_i^X(1-p)^{1-X_i}$$

$$= arg\max_{\mathrm{p}} \sum_{\mathrm{i}=0}^{\mathrm{n}} log[p_i^X(1-p)^{1-X_i}] = arg\max_{\mathrm{p}} \sum_{\mathrm{i}=0}^{\mathrm{n}} log[X_i logp + (1-X_i)log(1-p)]$$

Let $\frac{\partial L(p)}{\partial \mathrm{p}} = \sum_{\mathrm{i}=0}^{\mathrm{n}}\left[\frac{X_i}{\mathrm{p}} + \frac{1-X_i}{\text{p-1}}\right] = \sum_{\mathrm{i}=0}^{\mathrm{n}}\frac{p-X_i}{p(p-1)} = 0$

Then $\hat{\mathrm{p}}_{MLE} = \sum_{\mathrm{i}=0}^{\mathrm{n}} X_i$

Question 3.

(a) Because $y|x \sim N(x^T\beta, \sigma^2)$

$$L(\beta) = \prod_{\mathrm{i}=0}^{\mathrm{n}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i - x^T\beta)^2}{2\sigma^2}\right)$$

Then $logL(\beta) = \sum_{i=0}^{\mathrm{n}} \log\frac{1}{\sqrt{2\pi\sigma^2}} - \frac{\left(X_i - x^T\beta\right)^2}{2\sigma^2}$

Then $\hat{\beta}_{MLE} = arg\max_{\beta} logL(\beta) = arg\max_{\beta} \sum_{\mathrm{i}=0}^{n}\left(-\frac{\left(X_i - x^T\beta\right)^2}{2\sigma^2}\right)$

$= arg\min_{\beta} \sum_{i=0}^{\mathrm{n}}\left(\frac{\left(X_i - x^T\beta\right)^2}{2\sigma^2}\right) = arg\min_{\beta} \sum_{\mathrm{i}=0}^{\mathrm{n}}(X_i - x^T\beta)^2$

Let $\frac{\partial \sum_{\mathrm{i}=0}^{n}\left(X_i - x^T\beta\right)^2}{\partial\beta} = 0$

Then $\hat{\beta}_{MLE} = -2n\sum_{\mathrm{i}=0}^{\mathrm{n}} X_i\, x$