Logistic Regression

Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

$$P(y=1|x)=\mu(x)=\frac{1}{1+e^{-\dot{\theta}^T x}}=\frac{e^{X\theta}}{1+e^{X\theta}}$$

Finding the Parameter, θ

$$X\theta=\log\left(\frac{P(Y|X)}{1-P(Y|X)}\right)$$

- **Maximum Likelihood Estimation (MLE) of $\boldsymbol{\theta}$**
  - Choose θ that maximizes the probability of observed data

$$\hat{\boldsymbol{\theta}}=\boldsymbol{argmax_{\theta}P(D|\theta)}$$

- **This is Maximum Conditional Likelihood Estimation (MCLE)**
- $\hat{\theta}=argmax_{\theta}P(D|\theta)=argmax_{\theta}\prod_{1\le i\le N}P(Y_i|X_i;\theta)$

$$=argmax_{\theta}log(\prod_{1\le i\le N}P(Y_i|X_i;\theta))=argmax_{\theta}\sum_{1\le i\le N}log(P(Y_i|X_i;\theta))$$

- $P(Y_i|X_i;\theta)=\mu(X_i)^{Y_i}(1-\mu(X_i))^{1-Y_i}$
- $log\left(P(Y_i|X_i;\theta)\right)=Y_i\log\left(\mu(X_i)\right)+(1-Y_i)\log\left(1-\mu(X_i)\right)$

$$=Y_i\left\{\log\left(\mu(X_i)\right)-\log\left(1-\mu(X_i)\right)\right\}+\log\left(1-\mu(X_i)\right)$$

$$=Y_i\log\left(\frac{\mu(X_i)}{1-\mu(X_i)}\right)+\log\left(1-\mu(X_i)\right)$$

$$=Y_iX_i\theta+\log\left(1-\mu(X_i)\right)=Y_iX_i\theta-\log\left(1+e^{X_i\theta}\right)$$

Finding the Parameter, $\boldsymbol{\theta}$, contd.

Linear Regression (Closed Form):

$$\hat{f} = X\theta \qquad \nabla_\theta(\theta^T X^T X\theta - 2\theta^T X^T Y) = 0$$
$$2X^T X\theta - 2X^T Y = 0$$
$$\theta = (X^T X)^{-1} X^T Y$$

- $\hat{\theta} = argmax_\theta \sum_{1\le i\le N} log(P(Y_i|X_i;\theta))$
- $= argmax_\theta \sum_{1\le i\le N}\{Y_i X_i\theta - \log(1 + e^{X_i\theta})\}$
- Partial derivative to find a certain element in θ
- $\frac{\partial}{\partial\theta_j}\{\sum_{1\le i\le N} Y_i X_i\theta - \log(1 + e^{X_i\theta})\}$

$$P(y = 1|x) = \frac{e^{X\theta}}{1 + e^{X\theta}}$$

$$= \left\{\sum_{1\le i\le N} Y_i X_{i,j}\right\} + \left\{\sum_{1\le i\le N} -\frac{1}{1 + e^{X_i\theta}} \times e^{X_i\theta} \times X_{i,j}\right\}$$

$$= \sum_{1\le i\le N} X_{i,j}\left(Y_i - \frac{e^{X_i\theta}}{1 + e^{X_i\theta}}\right) = \sum_{1\le i\le N} X_{i,j}(Y_i - P(Y_i = 1|X_i;\theta)) = 0$$

- There is no way to derive further
  - There is no closed form solution!
  - Open form solution → approximate!

Cannot be easily solved in the closed form because of the logistic function