Homework 3

Nguyễn Hoàng Long - 11202352 - DSEB 62

5/10/2022

1. Biến đổi lại linear regression trên lớp ra latex, từ $t = y(x, w) + noise => w = (X^TX)^{-1}X^Tt$

Suppose that the observations are drawn independently from a Gaussian distribution

$$t = y(x, w) + \mathcal{N}(0, \sigma^2) = N(y(x, w), \sigma^2)$$

With $\beta = \frac{1}{\sigma^2}$

$$p(t|x, w, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|y(x_n, \beta^{-1})$$

It is convenient to maximize the logarithm of the likelihood function:

$$\begin{aligned} logp(t|x, w, \beta) &= \sum_{n=1}^{N} log(\mathcal{N}(t_n|y(x_n, w), \beta^{-1})) \\ &= \sum_{n=1}^{N} log(\frac{1}{\sqrt{2\pi\beta^{-1}}} e^{\frac{-(t_n - y(x_n, w))^2\beta}{2}}) \\ &= \sum_{n=1}^{N} (\frac{-1}{2} log(2\pi\beta^{-1}) - (t_n - y(x_n, w))^2 - \frac{\beta}{2}) \\ &= -\sum_{n=1}^{N} (t_n - y(x_n - w))^2 \end{aligned}$$

Minimize: $\sum_{i=1}^{N} \left(t_n - y\left(x_n - w\right)\right)^2$

Set $L = \frac{1}{N} \sum_{i=1} N(t_n - y(x_n, w))$

$$x = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ ... & ... \\ 1 & x_n \end{bmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, t = \begin{bmatrix} t_1 \\ t_2 \\ ... \\ t_n \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix} = \begin{bmatrix} w_1x_1 + w_0 \\ w_1x_2 + w_0 \\ ... \\ w_1x_n + w_0 \end{bmatrix} = Xw$$

$$t - y = \begin{bmatrix} t_1 - y_1 \\ t_2 - y_2 \\ ... \\ t_n - y_n \end{bmatrix}$$

$$\Longrightarrow \|t - y\|_2^2 = (t_1 - y_1)^2 + (t_2 - y_2)^2 + ... + (t_n - y_n)^2 = \sum_{i=1}^{n} (t_i - y_i)^2 = L$$

$$\begin{aligned} &\Longrightarrow L = \|t - y\|_2^2 = \|t - Xw\|_2^2 \\ &\Longrightarrow \frac{\partial L}{\partial w} = 2X^T(t - Xw) = 0 \\ &\Longrightarrow X^T t = X^T X w \\ &\Longrightarrow w = \left(X^T X\right)^{-1} X^T t \end{aligned}$$

4. Chứng minh $X^T X$ invertible khi X full rank.

If X is full rank, X is linear independent.

$$\begin{aligned} &\Rightarrow \vec{v}^T X^T X \vec{v} = \vec{v}^T \vec{0} \\ &\Rightarrow (X\vec{v})^T X \vec{v} = 0 \\ &\Rightarrow (X\vec{v}) \cdot (X\vec{v}) = 0 \\ &\Rightarrow X\vec{v} = \vec{0} \end{aligned}$$

We have: if $\vec{v} \in N\left(X^T X\right) \Rightarrow \vec{v} \in N(X)$

$\Rightarrow \vec{v}$can only be $\vec{0} \Rightarrow N\left(X^T X\right) = N(X) = \{\vec{0}\}$

$\Rightarrow X^T X$ is linearly independent; and $X^T X$ is a square matrix $\Rightarrow X^T X$ is invertible