Homework week 3

Nguyễn Minh Đức
Student code: 11204838

Octorber 2022

1 Question 1:

We have:

$$t = y(x, w) + \epsilon$$

Suppose that the observations are drawn independently from the Gaussian distribution. Then we wish to find

$$\begin{aligned} p(t_n) &\approx y(x_n, w) \\ p(t_n) &\approx \mathcal{N}(t_n|y(x_n, w), \epsilon^2) \end{aligned}$$

Let X = $\begin{bmatrix} X_1 \\ X_2 \\ ... \\ X_n \end{bmatrix}$; T = $\begin{bmatrix} t_1 \\ t_2 \\ ... \\ t_n \end{bmatrix}$; W = $\begin{bmatrix} w_0 \\ w_1 \\ ... \\ w_d \end{bmatrix}$; Y = $\begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix}$; $\beta = \frac{1}{\epsilon^2}$

We have:

$$p(T|X, W, \beta) = \prod_{i=1}^{N} \mathcal{N}(t_i|y(x_i, w); \beta^{-1})$$

and

$$\mathcal{N}(t_i|y(x_i, w), \beta^{-1}) = \frac{1}{\sqrt{2\pi\beta^{-1}}} e^{-(t_i - y(x_i,w))^2 \frac{\beta}{2}}$$

$$\Longrightarrow p(T|X, W, \beta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\beta^{-1}}} e^{-(t_i - y(x_i,w))^2 \frac{\beta}{2}}$$

$$\begin{aligned} log(p(T|X, W, \beta)) &= \sum_{i=1}^{N} log(\frac{1}{\sqrt{2\pi\beta^{-1}}} e^{-(t_i - y(x_i,w))^2 \frac{\beta}{2}}) \\ &= \sum_{i=1}^{N} (\frac{-1}{2} log(2\pi\beta^{-1}) - (t_i - y(x_i, w))^2) \end{aligned}$$

The goal is to maximize $log(p(t|x, w, \beta))$ so we have to minimize:

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

Minimize loss function:

$$L = \frac{1}{N}\sum_{i=1}^{N}(t_i - y(x_i, w))^2$$

Let X = $\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ ... & ... \\ 1 & x_n \end{bmatrix}$; T = $\begin{bmatrix} t_1 \\ t_2 \\ ... \\ t_n \end{bmatrix}$; W = $\begin{bmatrix} w_0 \\ w_1 \\ ... \\ w_d \end{bmatrix}$; Y = $\begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix} = \begin{bmatrix} w_0 + w_1 x_{11} + ... + w_d x_{1d} \\ w_0 + w_1 x_{21} + ... + w_d x_{2d} \\ ... \\ w_0 + w_1 x_{n1} + ... + w_d x_{nd} \end{bmatrix}$

$\Longrightarrow T - Y = \begin{bmatrix} t_1 - y_1 \\ t_2 - y_2 \\ ... \\ t_n - y_n \end{bmatrix} \Longrightarrow ||T - Y||_2^2 = \sum_{i=1}^{n}(t_i - y_i)^2 = L$

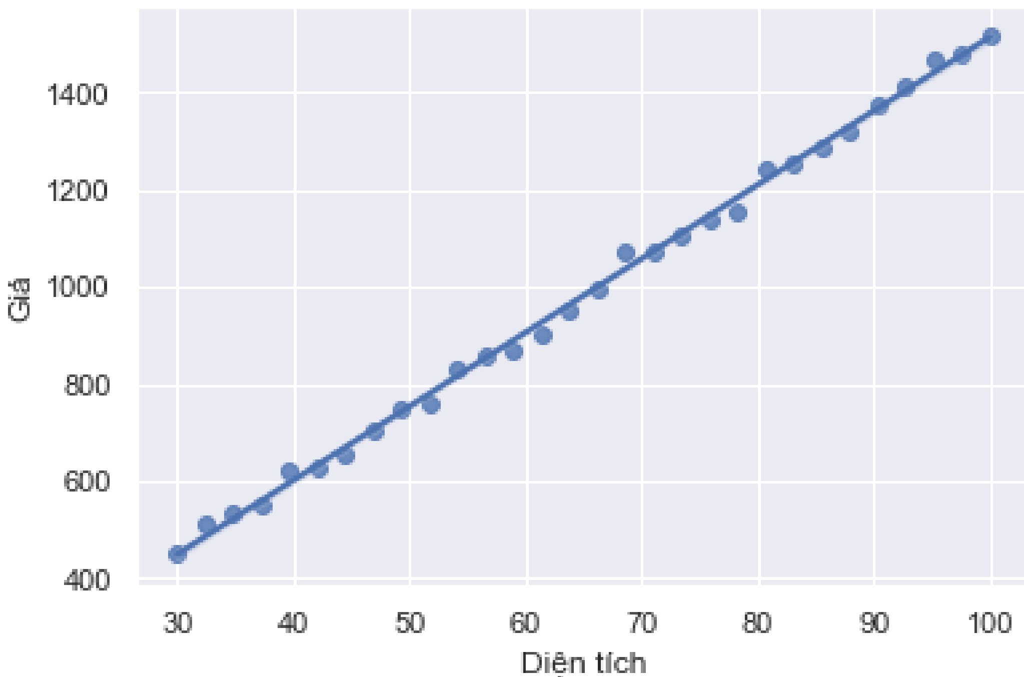
$$\Longrightarrow L = ||T - Y||_2^2 = ||T - XW||_2^2$$

$$\begin{aligned} \frac{\partial L}{\partial W} &= 2X^T(T - XW) = 0 \\ \Longleftrightarrow X^T T &= X^T X W \\ \Longleftrightarrow W &= (X^t X)^{-1} X^T T \end{aligned}$$

2 Question 2:

a) Linear Regression model for house price prediction:

$$\text{HousePrice} = -7.064 + 15.211 * \text{HouseArea}$$



b) Prediction:

Area	Price
50	753.49
100	1514.05
150	2274.59

3 Question 3:

Linear Regression model for house price prediction in Boston:

$$\begin{aligned}
\text{Price} = 36.459 &- 0.108011 * \text{CRIM} + 0.046420 * \text{ZN} + 0.020559 * \text{INDUS} \\
&+ 2.686734 * \text{CHAS} - 17.766611 * \text{NOX} + 3.809865 * \text{RM} + 0.000692 * \text{AGE} \\
&- 1.475567 * \text{DIS} + 0.306049 * \text{RAD} - 0.012335 * \text{TAX} \\
&- 0.952747 * \text{PTRATIO} + 0.009312 * \text{B} - 0.524758 * \text{LSTAT}
\end{aligned}$$

4 Question 4:

X is a m*n matrix $X^T X$ is a n*n matrix
If X in linearly indepedence when $X\vec{v} = 0$ have only trivial solution $\vec{v} = \vec{0}$,

and $\vec{v} \in N(X)$. We have:

$$\begin{aligned} X\vec{v} &= 0 \\ \implies X^T X\vec{v} &= 0 \\ \implies \vec{v} &\in N(X^T X) \end{aligned}$$

Hence $N(X) \subseteq N(X^T X)$

If $X^T X$ in linearly indepedence when $X^T X\vec{v} = 0$ have only trivial solution $\vec{v} = \vec{0}$, and $\vec{v} \in N(X^T X)$. We have:

$$\begin{aligned} X^T X\vec{v} &= 0 \\ \to \vec{v}^T X^T X\vec{v} &= \vec{v}^T 0 \\ \to (X\vec{v})^T (X\vec{v}) &= 0 \\ \to ||X\vec{v}||_2^2 &= 0 \\ \to X\vec{v} &= 0 \\ \to \vec{v} &\in N(X) \end{aligned}$$

Hence $N(X^T X) \subseteq N(X)$

Therefore $N(X) = N(X^T X)$, so $X^T X$ invertible $\leftrightarrow$ $X^T X$ linearly indepedence $\leftrightarrow$ X linearly indepedence $\leftrightarrow$ X full rank