Week 3 - Linear Regression

NGUYEN NHU QUYNH

October 2021

Exercise 1

- **Question:**
$$t = y(x, w) + noise \Rightarrow w = (X^T X)^{-1} X^T t$$

- **We have:**

  - A observation: $x = (x_1, x_2, x_3, ..., x_N)^T$

  - Total observation N

  - Target values: $t = (t_1, t_2, t_3, ..., t_N)^T$

$$\Rightarrow t = y(x, w) + N(0, \beta^{-1})t$$

$$= N(y(x, w), \beta^{-1})$$

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

$$\Rightarrow p(t|x, w, \beta) = N(t|y(x, w), \beta^{-1})$$

Using maximum likelihood to observe data for model parameters w, β^{-1}, we have:

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

$$\Rightarrow p(t|x, w, \beta) = (\frac{1}{\sqrt{2\pi\beta^{-1}}})^N.exp(\frac{-\beta}{2}(t_n - y(x, w)))$$

Because log: $R^+ \mapsto R$ is an increasing function, we can instead maximize the log of the likelihood, which results in a simplier mathematical expression.

$$LL(t|x,w,\beta) = \frac{-N}{2}log(2\pi\beta^{-1}) - \frac{1}{2\beta^{-1}}\prod_{i=1}^{N}(y(x,w)-t)^2$$

We would like to find w that maximizes the log-likelihood. Alternatively, we can minimize the negative log-likelihood.

$$NLL(t|x,w,\beta) = \frac{1}{2}\beta\prod_{i=1}^{N}(y(x,w)-t)^2 + \frac{N}{2}log(2\pi\beta^{-1})$$

Minimize NLL $(t|x,w,\beta) \leftrightarrow Minimize$ S $= \prod_{i=1}^{N}(y(x,w)-t)^2$

$$\Rightarrow S_{min} = \|Xw - t\|^2$$

$$S_{min} = t^Tt - (Xw)^Tt - t^T(Xw) + (Xw)^T(Xw)$$

$$S_{min} = t^Tt - 2(Xw)^Tt + (Xw)^T(Xw)$$

$$S_{min} = t^Tt - 2X^Tw^Tt + w^TX^TXw$$

$$With\frac{dS}{dw} = -2X^Tt + 2X^TXw = 0$$

$$\Rightarrow 2X^TXw = 2X^Tt$$

$$\Leftrightarrow w = (X^TX)^{-1}X^Tt$$

Exercise 4

- Because X is full rank $\Rightarrow$ R(A) = Number of Rows = Number of Cols

$\Rightarrow$ X is square matrix

$$\Rightarrow \left|X^T\right| = |X| \Leftrightarrow |X|^2 \neq 0 \Leftrightarrow \left|X^T\right||X| \neq 0 \Leftrightarrow \left|X^TX\right| \neq 0$$

$\Rightarrow X^TX$ is invertible.