Machine Learning week 3

Kieu Son Tung

5. Oktober 2021

1 Prove the normal equation

$$w = (X^T X)^{-1} X^T y$$

Given that

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_m \end{pmatrix} \qquad w = \begin{pmatrix} w_1 \\ w_2 \\ \cdots \\ w_n \end{pmatrix} \qquad x = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ \cdots \\ x_{m1} & x_{m2} & x_{m3} & \cdots & x_{mn} \end{pmatrix} \tag{1}$$

We have

$$Error = \begin{pmatrix} y_1 - (w_1 x_{11} + w_2 x_{12} + \cdots + w_n x_{1n}) \\ y_2 - (w_1 x_{21} + w_2 x_{22} + \cdots + w_n x_{2n}) \\ \cdots \\ y_m - (w_1 x_{m1} + w_2 x_{m2} + \cdots + w_n x_{mn}) \end{pmatrix} = y - xm \tag{2}$$

So we have the lost function

$$\begin{aligned} L &= (y - xm)^T (y - xm) \\ &= (y^T - w^T x^T)(y - xw) \\ &= y^T y - y^T xw - w^T x^T y + w^T x^T xw \end{aligned} \tag{3}$$

We have that $\nabla_w (w^T a) = \nabla_w (a^T w) = a$. So from (3)

$$\begin{aligned} \frac{\partial L}{\partial w} &= 0 - x^T y - x^T y + x^T xw + (w^T x^T x)^T \\ &= -2x^T y + 2x^T xw = 0 \end{aligned} \tag{4}$$

From that we have

$$x^T y = x^T x w$$
$$w = (x^T x)^{-1} x^T y$$

2 Prove that $A^T A$ is invertible when A is full rank

We have

$$A_{n\times k} = \begin{pmatrix} \vec{a_1} & \vec{a_2} & \cdots \vec{a_k} \end{pmatrix} \qquad x = \begin{pmatrix} x_1 & x_2 & \cdots x_k \end{pmatrix}$$

If A is full rank, then $\vec{a_1}, \vec{a_2}, \cdots, \vec{a_k}$ are linear independent $\implies x_i$ is 0

$$x_1\vec{a_1} + x_2\vec{a_2} + \cdots + x_k\vec{a_k} = 0$$

We have the shape of matrix $A^T A$ is $k \times n \times n \times k = k \times k \implies A^T A$ is a square matrix.
Suppose we have $\vec{v} \in N(A^T A)$ (null space of the matrix $A^T A$), we have

$$\begin{aligned} A^T A\vec{v} &= \vec{0} \\ \Rightarrow v^T A^T A\vec{v} &= v^T\vec{0} = \vec{v}\cdot\vec{0} = 0 \\ \Rightarrow (A\vec{v})^T A\vec{v} &= 0 \\ \Rightarrow (A\vec{v})(A\vec{v}) &= 0 \\ \Rightarrow \|A\vec{v}\|^2 &= 0 \\ \Rightarrow A\vec{v} &= \vec{0} \end{aligned} \tag{5}$$

Then $\vec{v}$ is also $\vec{v} \in N(A)$

$$\begin{aligned} \Rightarrow \vec{v} &= \vec{0} \\ \Rightarrow (A^T A)\vec{v} &= \vec{0} \end{aligned} \tag{6}$$

We have $A^T A$ is linear independent, and $A^T A$ is also a square matrix, so $A^T A$ is invertible.