Machine Learning 1 - Homework week 3: Linear Regression

Mai Xuan Bach
ID 11200489

Ngày 5 tháng 10 năm 2022

1 Problem 1. Biến đổi:

We assume that the the target variable t is given by a deterministic function $y(x, w)$ with additive Gaussian noise so that

$$t = y(x, w) + \epsilon \tag{1}$$

Let

$$\epsilon = N(\mu, \sigma^2) \tag{2}$$

Set

$$\epsilon \backsim N(\mu, \sigma^2) \tag{3}$$

$$\Rightarrow t = y(x, w) + \epsilon \backsim N(y(x, w), \sigma^2) \tag{4}$$

$$\Rightarrow p(t) = N(t|y(x, w), \sigma^2) \tag{5}$$

We want:

$$t_n \approx y(x_n, w) \tag{6}$$

Then, $\boldsymbol{p(t_n)}$ could be maximized

Now consider a data set of inputs $X = \{x_1, ..., x_N\}$ with corresponding target values $t_1, \ldots, t_N$. We group the target variables $\{t_n\}$ into a column vector that we denote by $\boldsymbol{t}$ where the typeface is chosen to distinguish it from a single observation of a multivariate target, which would be denoted t. We obtain the following expression for the likelihood function, which is a function of the adjustable parameters w and β, in the form

$$p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^{n} \mathcal{N}(t_n|y(x_n, w), \beta^{-1}) \tag{7}$$

Let maximize (7) to find the model, i.e find the w:

$$\begin{aligned}
log\prod_{n=1}^{n}\mathcal{N}(t_n|y(x_n,w),\beta^{-1}) &= log\prod_{n=1}^{n}\frac{1}{\sqrt{2\pi\beta^{-1}}}e^{-(t_n-y(x_n,w))^2.\frac{\beta}{2}} && (8)\\
&= \sum_{n=1}^{n}log\left(\frac{1}{\sqrt{2\pi\beta^{-1}}}e^{-(t_n-y(x_n,w))^2.\frac{\beta}{2}}\right) && (9)\\
&= \sum_{n=1}^{n}\frac{-1}{2}log(2\pi\beta^{-1})-(t_n-y(x_n,w))^2.\frac{\beta}{2} && (10)\\
&= const-\sum_{n=1}^{n}(t_n-y(x_n,w))^2 && (11)
\end{aligned}$$

To maximize (11), we minimize:

$$\sum_{n=1}^{n}(t_n-y(x_n,w))^2 \tag{12}$$

In other words, we minimize the loss function:

$$L=\frac{1}{N}\sum_{n=1}^{N}(\hat{y}_i-y_i)^2 \tag{13}$$

with $\hat{y}_i = w_1x_i + w_0$

We have: $w=\begin{bmatrix} w_0\\ w_1\end{bmatrix}, X=\begin{bmatrix} 1 & x_1\\ 1 & x_2\\ ... & ...\\ 1 & x_n\end{bmatrix}, y=\begin{bmatrix} y_1\\ y_2\\ ...\\ y_n\end{bmatrix}, \hat{y}=\begin{bmatrix} \hat{y}_1\\ \hat{y}_2\\ ...\\ \hat{y}_n\end{bmatrix}=\begin{bmatrix} w_0+w_1x_1\\ w_0+w_1x_2\\ ...\\ w_0+w_1x_n\end{bmatrix}=Xw.$

So that, $L=\|\hat{y}-y\|^2=\|Xw-y\|^2=(Xw-y)^T(Xw-y)$

$$\begin{aligned}
\frac{\partial L}{\partial w} &= 2X^T(Xw-y)=0 && (14)\\
<=> &\quad X^TXw=X^Ty && (15)\\
<=> &\quad w=(X^TX)^{-1}X^Ty && (16)
\end{aligned}$$

2 Problem 4. Chứng minh:

$X^T X$ is invertible when X full rank.

<u>Answer</u>:

Assume X is an m x n size, with rank m (n $\geq m$).

Let $X^T Xv = 0$, for some v $\in R^M$

$$\Rightarrow v^T X^T Xv = 0 \tag{17}$$

$$\Rightarrow (Xv)^T Xv = 0 \tag{18}$$

$$\Rightarrow Xv = 0 \tag{19}$$

If the rank of X is m, this means that X is one-to-one when acting on R^M. Then, $X^T X$ is one-to-one too, and because it is a square matrix, it is invertible.