machine learning homework 3

Nguyen Ba Dang Khoi

September 2021

Linear Regression

1 Exercise 1

Transform

$$t = y(x, w) + noise -> w = (X^T X) - 1X^T t$$

Suppose that the observations are drawn independently from Gaussian distribution:

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

With

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

t: observed value
y: predicted value /model
x: training data

We now use the training data x, t to determine the values of the unknown parameters w by maximum likelihood. If the data are assumed to be drawn independently from the distribution then the likelihood function:

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

It is convenient to maximize the logarithm of the likelihood function

$$
\begin{aligned}
log\, p(\mathbf{t}|\mathbf{x,w,}\beta) &= \sum_{n=1}^{N} log(N(t_n|y(x_n,w),\beta^{-1})) \\
&= \sum_{n=1}^{N} log(\frac{1}{(2\pi\beta^{-1}})exp^{-\frac{(t-y(x_n,w))^2\beta)}{2}} \\
&= \sum_{n=1}^{N} [-\frac{1}{2}log(2\pi\beta^{-1}) - \frac{(t_n - y(x_n,w))^2\beta}{2}]
\end{aligned}
$$

We only need to deal with w so we can ignore the constants

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

=> we need to minimize $t_n - y(x_n,w))^2$ to find w.

Suppose

$$
L = \frac{1}{N}\sum_{n=1}^{N}(t_n - y(x_n,w))^2
$$

x = $\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ 1 & x_n \end{bmatrix}$, w = $\begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$, y = $\begin{bmatrix} 1 & y_1 \\ 1 & y_2 \\ . & . \\ . & . \\ 1 & y_n \end{bmatrix} = \begin{bmatrix} w_0 + x_1 w_1 \\ w_0 + x_2 w_1 \\ . \\ . \\ w_0 + x_n w_1 \end{bmatrix}$ = xw

$$
\frac{\partial L}{\partial w} = \begin{bmatrix} \frac{\partial L}{\partial w_0} \\ \frac{\partial L}{\partial w_1} \end{bmatrix} = \begin{bmatrix} t - xw \\ x(t - xw) \end{bmatrix} = x^\intercal (t - xw) = 0
$$

$$
\begin{aligned}
&=> x^\intercal t = x^\intercal x w \\
&<=> w = (x^\intercal x)^{-1} x^\intercal t
\end{aligned}
$$