Homework 3 Machine Learning

October 5, 2022

Exercise 1:

We have : $t = y(w, x) + \varepsilon$
while $\varepsilon \approx N(0, \sigma^2)$
If $\mu(\varepsilon) \neq 0$ we just adjust bias of y : $\mu(\varepsilon) = 0$
$\Rightarrow P(t) = N(t|y(w, x), \sigma^2)$
Suppose: $t_n = y(x_n, w) + \varepsilon$
$\Rightarrow P(t_n) = N(t_n|y(x_n, w), \varepsilon^2)$
Generality: Maximum for all point, we use maximum likelihood function:
$P(t|x, w, \beta) = \prod_{n=1}^{N} N(t_n|y(x, w), \beta^{-1})$
Simplize:
$\log(P(t|x, w, \beta)) = \sum_{n=1}^{N}(log(N(t_n|y(x, w), \beta^{-1})))$
$= \dfrac{-\beta}{2}\sum_{n=1}^{N}(y(x_n, w) - t_n)^2 + \dfrac{N}{2}\log(\beta) - \dfrac{N}{2}\log(2\pi)$
Maximum likelihood:
Max $\log(P(t|x, w, \beta))$ = -Max $\dfrac{\beta}{2}\sum_{n=1}^{N}(y(x_n, w) - t_n)^2$
$=$ Min $\dfrac{1}{2}\sum_{n=1}^{N}(y(x_n, w) - t_n)^2$
We minimize P $= \dfrac{1}{2}\sum_{n=1}^{N}(y(x_n, w) - t_n)^2$ to find w
Suppose:

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ ... & ... \\ 1 & x_n \end{pmatrix}; t = \begin{pmatrix} t_1 \\ t_2 \\ ... \\ t_n \end{pmatrix}; w = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}$$

By minimizing P, we can find w. P is called Mean Squared Error Loss(MSE):
$L = \dfrac{1}{N}\sum_{n=1}^{N}(t_n - y(x_n, w))^2$

we have:
$y(x_n, w) = w_1 x_v + w_0$

$$y = \begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{pmatrix} = \begin{pmatrix} w_1x_1 & w_0 \\ w_2x_2 & w_0 \\ ... & ... \\ w_nx_n & w_0 \end{pmatrix} = XW$$

$$t - y = \begin{pmatrix} t_1 - y_1 \\ t_2 - y_2 \\ ... \\ t_n - y_n \end{pmatrix}$$

$$\Rightarrow L = \|t - y\|_i^2 = \|t - Xw\|_i^2 = (t - Xw)^T(t - Xw)$$

$$\frac{\partial(L)}{\partial(w)} = 2X^T(t - Xw) = 0$$

$$\Leftrightarrow X^T t = X^T X w$$

$$\Leftrightarrow w = (X^T X)^{-1} X^T t$$