ML dseb62 w3 - linear regression

Nguyen Tuan Duy

September 2022

1 Problem 1

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

Suppose that the observations are drawn independently from a Gaussian distribution. Then we wish to find

$$p(t_n) \approx y(x_n, w)$$
$$p(t_n) \approx N(t_n | y(x_n, w), \epsilon^2)$$

Let $T = \begin{bmatrix} t_1 \\ . \\ . \\ t_n \end{bmatrix}, W = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, X = \begin{bmatrix} 1 & x_1 \\ . & . \\ . & . \\ 1 & x_n \end{bmatrix},$

$Y = \begin{bmatrix} y_1 \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 w_1 + w_0 \\ . \\ . \\ x_n w_1 + w_0 \end{bmatrix} = XW, \quad \beta = \frac{1}{\epsilon^2}$, then

$$\begin{aligned} p(T|X, W, \beta) &= \prod_{i=1}^{N} N(t_i | y(x_i, w), \beta^{-1}) \\ &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\beta^{-1}}} e^{-(t_i - y(x_i, w))^2 \frac{\beta}{2}} \\ log(p(T|X, W, \beta)) &= \sum_{i=1}^{N} log(\frac{1}{\sqrt{2\pi\beta^{-1}}}) - (t_i - y(x_i, w))^2 \frac{\beta}{2} \end{aligned}$$

because the goal is to maximize $log(p(T|X, W, \beta))$ with respect to w, we only consider maximizing

$$\sum_{i=1}^{N} -(t_i - y(x_i, w))^2 \tag{2}$$

and the problem can also be written as minimizing:

$$
\begin{aligned}
L &= \sum_{i=1}^{N} (t_i - y(x_i, w))^2 && (3) \\
&= \sum_{i=1}^{N} (t_i - y_i)^2 && (4) \\
&= ||T - Y||_2^2 && (5) \\
&= ||T - XW||_2^2 && (6)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\delta L}{\delta W} &= 2X^T(T - XW) = 0 \\
\iff 2X^T T - 2X^T XW &= 0 \\
\iff X^T XW &= X^T T \\
\iff W &= (X^T X)^{-1} X^T T
\end{aligned}
$$

2 Problem 4

Because $x^T x$ is symmetric, if $x^T x$ is also positive definite then it is also invertible. $x^T x$ is positive definite when for all $z \in R^n \setminus \{0\}$:

$$z^T (x^T x) z > 0 \tag{7}$$

let $\bar{z} = xz$ then $z^T (x^T x) z = \bar{z}^T \bar{z} = ||\bar{z}||_2^2 \geq 0$
because x is linearly independent (full rank) and $z \neq 0 \rightarrow xz > 0 \rightarrow ||\bar{z}||_2^2 > 0$.
this proves (7) holds, thus, $x^T x$ is invertible