Machine learning

Homework week 3

September 2021

1 t = y(x,w) + noise $\rightarrow w = (X^TX)^{-1}X^Tt$

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

$$\rightarrow p(t \mid x, w, \beta) = \Pi_{n=1}^{N} N(t_n \mid y(x_n, w), \beta^{-1})$$

$$log\ p(t \mid x, w, \beta) = \sum_{n=1}^{N} log(N(t_n \mid y(x_n, w), \beta^{-1}))$$

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

$$maxlog\ p(t \mid x, w, \beta) = -max\frac{\beta}{2}\sum_{n=1}^{n}(y(x_n, w) - t_n)^2$$

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

$$We\ minimize\ P = min\frac{1}{2}\sum_{n=1}^{n}(y(x_n, w) - t_n)^2 to\ find\ w.Suppose$$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ ... & ... \\ 1 & x_n \end{bmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

$$\rightarrow P = \|Xw - t\|_2^2$$

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

$$\Leftrightarrow X^T = X^T Xw$$

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

2 Proof $\mathbf{X}^T X$ *invertable when X full rank.*

dim(X) = NxM, dim(X^T) $= MxN, dim(X^T X) = MxM(square\ matrix)$
$Suppose : X^T v = 0$
$XX^T v = 0$
$v^T XX^T v = 0$
$(X^T v)^T (X^T v) = 0$
$Hence,\ we\ have\ to\ proved\ X^T v = 0 if\ and\ only\ if\ v\ is\ in\ the\ nullspace\ of X^T X$
$But : X^T v = 0\ and\ v \neq 0\ if\ and\ only\ if\ X\ has\ linear\ dependent\ rows.$
$Thus, X^T X\ is\ invertible\ if\ and\ only\ if\ X\ has\ full\ row\ rank$