cost function :→ $J = \sum (t_i - y_i)^2$

Turn it into matrix form ↪

$J = (t-y)^T(t-y)$

where $t$ = N×1 matrix of target

$y$ = N×1 matrix of target

$y = Xw$

where

$X$ = N×D matrix

$w$ = D×1 matrix

so, $y$ = N×1 vector

Now we want to calculate derivative of $\frac{\partial J}{\partial w}$

first expand $J$! →

Camlin Page
Date / /

$J = (t-y)^T(t-y) = (t^T - y^T)(t-y)$

$J = t^Tt - y^Tt - t^Ty + y^Ty$ $\quad \because y = Xw$

$J = t^Tt - (Xw)^Tt - t^T(Xw) + (Xw)^T(Xw)$

~~$J = t^Tt - (Xw)^Tt -$~~ $J = t^Tt - w^TX^Tt - t^TXw + w^TX^TXw$

$\frac{\partial J}{\partial w} = -2X^Tt + 2X^TXw = 0$

$\frac{\partial J}{\partial w} = X^Tt = X^TXw$

$= \quad w = (X^TX)^{-1}X^Tt$

From matrix cookbook :→ $\frac{\partial x^Ta}{\partial x} = \frac{\partial a^Tx}{\partial x} = a$

chain rule

$\frac{\partial J}{\partial w} = 0 - X^Tt - (t^TX)^T + \boxed{X^TXw + (w^TX^TX)^T}$

$\frac{\partial J}{\partial w} = - X^Tt - X^Tt + X^TXw + X^TXw$

equate to zero →

$\frac{\partial J}{\partial w} = -2X^Tt + 2X^TXw = 0$

$\frac{\partial J}{\partial w} = X^TXw = X^Tt$

$= \boxed{w = (X^TX)^{-1}X^Tt}$