Mounica Subramani

a) $f(B) = \|X_1 B - Y_1\|^2$, where $B \in \mathbb{R}^p$

Show that the minimizer of $f(B)$, denoted by $\hat{\beta}_1$ is equal to $(X_1^T X_1)^{-1} X_1^T Y_1$.

$$f(B) = \|X_1 B - Y_1\|^2$$

$$\boxed{Y' = X_1 B}$$

$$\Rightarrow \|Y' - Y_1\|^2$$

$$\Rightarrow (Y' - Y_1)^T (\hat{Y}^1 - Y_1)$$

$$\Rightarrow (X_1 B - Y_1)^T (X_2 B - Y_1)$$

diff. w.r.to $B$

$$\Rightarrow \frac{\partial}{\partial B}\left(B^T X_1^T X_1 B - B^T X_1^T Y_1 - Y_1^T X_2 B + Y_1^T Y_1\right)$$

diff. w.r.to $x$.

$$\frac{\partial}{\partial x} x^T x = 2x \quad \& \quad \frac{\partial}{\partial x} Ax = A^T.$$

$$\therefore \frac{\partial}{\partial B}(f(B)) = X_1^T X B - 2X^T Y_1$$

$$X^T X B - \cancel{2} X^T Y_1 = 0$$

$$X^T X B = X^T Y_1$$

$$\boxed{B = (X^T X)^{-1} X^T Y_1}$$

c) Assume $\eta$ is infinetesimally small, $w_t$ is accurately captured by,

$$\frac{dw_t}{dt} = -X_2^T (X_2 w_t - Y_2) \quad \text{——— (1)}$$

To solve:

$w_t$ & loss curve $L(t) = \|w_t - \beta_2\|^2$.

General solution for first - ODE is,

$$w_t = \exp(A_1 \cdot t) \cdot A_2 + A_3 \quad \text{—— (2)}$$

Sub (2) in (1)

$$\frac{dw_t}{dt} = -X_2^T \left( X_2 \left( \exp(A_1 t) A_2 + A_3 \right) - Y_2 \right)$$

$$\Rightarrow -X_2^T \left( X_2 \exp(A_1 t) A_2 + X_2 A_3 \right) - Y_2 )$$

$$\Rightarrow -X_2^T X_2 \exp(A_1 t) A_2 - X_2^T X_2 A_3 + X_2^T Y_2$$

$$\boxed{A_1 = -X_2^T X_2} \quad \text{—— (3)}$$

$$X_2^T X_2 A_3 = X_2^T Y_2 . \quad \text{—— (4)}$$

$$A_3 = (X_2^T Y_2)(X_2^T X_2)^{-1} = \hat{\beta}_2 \quad \text{—— (5)}$$

$$\boxed{\therefore A_3 = \hat{\beta}_2}$$

also given that

initialization : $w_0 = \hat{\beta}_1$

So, $A_2 + A_3 = \hat{\beta}_1$

$A_2 = \hat{\beta}_1 - A_3$.

$$\boxed{A_2 = \hat{\beta}_1 - \hat{\beta}_2}$$

Substitute the values of $A_1$, $A_2$ and $A_3$ in eq (2)

$$w_t = \exp(-X_2^T X_2 t)(\hat{\beta}_1 - \hat{\beta}_2) + \hat{\beta}_2$$

Loss curve $L(t) = \|w_t - \hat{\beta}_2\|^2$.

$$\therefore L(t) = \| \exp(-X_2^T X_2 t)(\hat{\beta}_1 - \hat{\beta}_2) + \cancel{\hat{\beta}_2} - \cancel{\hat{\beta}_2} \|^2$$

$\therefore$ The closed form of $L(t)$ depends only on $X_2, \hat{\beta}_1, \hat{\beta}_2$, and $t$.

d) Assume, $X_2^T X_2 = I$

(i) use this assumption to rewrite $L(t)$ to a simpler form.

$$L(t) = \|(\exp(-It)(\hat{\beta}_1 - \hat{\beta}_2) + \hat{\beta}_2) - \beta_2\|^2$$

$$\Rightarrow \|e^{-t}(\hat{\beta}_1 - \hat{\beta}_2) + \hat{\beta}_2 - \beta_2\|^2$$

(ii) Write $\hat{\beta}_2 - \beta_2$ and $\hat{\beta}_1 - \beta_2$ in terms of $x_1, x_2, \varepsilon_1, \varepsilon_2$ using the fact that $\hat{\beta}_1, \hat{\beta}_2$ are the ordinary least square solutions to linear regression.

Given:

the labels for the data,

$$Y_1 = X_1\beta_1 + \varepsilon_1 \quad \text{——— (1)}$$

$$Y_2 = X_2\beta_2 + \varepsilon_2 \quad \text{——— (2)}$$

Assuming, $\beta_2 = \beta_1 + \delta R$. ——— (3)

From a) we know, $\hat{\beta}_1 = (X_1^T X_1)^{-1} X_1^T Y_1$ ——— (4)

From c) we know, $\hat{\beta}_2 = (X_2^T X_2)^{-1}(X_2^T Y_2)$ ——— (5)

From (1), $\beta_1 = (Y_1 - \varepsilon_1) X_1^{-1}$

From (2), $\beta_2 = (Y_2 - \varepsilon_2) X_2^{-1}$

$$\hat{\beta}_1 - \beta_2 = \hat{\beta}_1 - (\beta_1 + \delta R)$$

$$\Rightarrow (X_1^T X_1)^{-1} X_1^T Y_1 - ((Y_1 - \varepsilon_1)X_1^{-1} + \delta R)$$

$$\Rightarrow (X_1^T X_1)^{-1} X_1^T Y_1 - Y_1 X_1^{-1} + \varepsilon_1 X_1^{-1} - \delta R.$$

$X_1^T X_1 = I$ (assuming)

$$\Rightarrow I X_1^T Y_1 - Y_1 X_1^{-1} + \varepsilon_1 X_1^{-1} - \delta R.$$

$$\boxed{\hat{\beta}_1 - \beta_2 \Rightarrow Y_1(X_1^T - X_1^{-1}) + \varepsilon_1 X_1^{-1} - \delta R.}$$

Scanned with CamScanner

$$\hat{\beta}_2 - \beta_2 = (X_2^T X_2)^{-1} X_2^T Y_2 - (Y_2 - \varepsilon_2) X_2^{-1}$$

$$\Rightarrow I\, X_2^T Y_2 - Y_2 X_2^{-1} + \varepsilon_2 X_2^{-1}$$

$$\boxed{\hat{\beta}_2 - \beta_2 \Rightarrow Y_2 (X_2^T - X_2^{-1}) + \varepsilon_2 X_2^{-1}}$$

Now, $\hat{\beta}_1 - \hat{\beta}_2$ is,

$$\Rightarrow \hat{\beta}_1 - \beta_2 - (\hat{\beta}_2 - \beta_2)$$

$$\Rightarrow Y_1 (X_1^T - X_1^{-1}) + \varepsilon_1 X_1^{-1} - \delta R - \left[ Y_2 (X_2^T - X_2^{-1}) + \varepsilon_2 X_2^{-1} \right]$$

$$\hat{\beta}_1 - \hat{\beta}_2 = Y_1 (X_1^T - X_1^{-1}) + \varepsilon_1 X_1^{-1} - \delta R - Y_2 (X_2^T - X_2^{-1}) - \varepsilon_2 X_2^{-1}.$$

Loss $L(t) = \| e^{-t} \cdot (\hat{\beta}_1 - \hat{\beta}_2) + \hat{\beta}_2 - \beta_2 \|^2$

$$L(t) = \| e^{t} \cdot Y_1 (X_1^T - X_1^{-1}) + \varepsilon_1 X_1^{-1} - \delta R - Y_2 (X_2^T X_2^{-1}) - \varepsilon_2 X_2^{-1}$$

$$+ Y_2 (X_2^T X_2^{-1}) + \varepsilon_2 X_2^{-1} \|^2$$

iii) Using results, derive $E[L(t)]$, expectation of $L(t)$ over $\varepsilon_1$ and $\varepsilon_2$.

$$L(t) = \| e^{-t} (\hat{\beta}_1 - \hat{\beta}_2) + \hat{\beta}_2 - \beta_2 \|^2$$

$\hat{\beta}_1$ and $\hat{\beta}_2$ are OLS solutions to linear regression.

$$\therefore \hat{\beta}_1 = \beta_1 + \varepsilon_1$$

$$\hat{\beta}_2 = \beta_2 + \varepsilon_2. \qquad [\varepsilon_2 = \hat{\beta}_2 - \beta_2]$$

$$L(t) = \| e^{-t} (\hat{\beta}_1 - (\beta_2 + \varepsilon_2)) + \varepsilon_2 \|^2$$

$$L(t) = \| e^{-t} (\hat{\beta}_1 - (\beta_1 + \delta R + \varepsilon_2))) + \varepsilon_2 \|^2$$

$$\Rightarrow \| e^{-t} (\hat{\beta}_1 - \beta_1 - \delta R - \varepsilon_2) + \varepsilon_2 \|^2$$

$$\Rightarrow \| e^{-t} (\varepsilon_1 - \delta R - \varepsilon_2) + \varepsilon_2 \|^2$$

Integrate $L(t)$ over $\varepsilon_1$ and $\varepsilon_2$ for $E[L(t)]$.

$$E[L(t)] = \frac{e^{-2t} (\varepsilon_1 + \varepsilon_2 (e^{t} - 1) - \delta R)}{12 (e^{t} - 1)} + \varepsilon_1 c_1 + c_2$$

(iv) Optimal stopping time, $t^* = \arg\min_t E[L(t)]$.

Differentiate $E[L(t)]$ w.r.t. $t$ and set it to zero and solve for $t$.

Scanned with CamScanner