CS419 Assignment 1 | Kunal Agarwal | 150110051

Q.1. probability to rain $\rightarrow r$

probability for student to come given raining $\rightarrow q$

(——— j students ——— $\rightarrow \sum_{i=j}^{N} {}^{N}C_i q^i (1-q)^{N-i}$

prob of having exam

$\sum_i$ prob. of i students showing up.

& we are taking $i \geq j$

$\therefore$ Prob of having exam $\rightarrow r \sum_{i=j}^{N} {}^{N}C_i \, q^i (1-q)^{N-i}$

probability if its not raining $\rightarrow 1-r$

prob. for student to come to school (given NOT raining) $\rightarrow$ ~~q~~ p.

prob for i student's to come $\rightarrow {}^{N}C_i p^i (1-p)^{N-i}$

prob of having exam (at least j students showup) $\Big\}$ $(1-r) \sum_{i=j}^{N} {}^{N}C_i \, p^i (1-p)^{N-i}$

$\therefore$ Probability of having exam (professor conducting exam)

$$\rightarrow r \sum_{i=j}^{N} {}^{N}C_i \, q^i (1-q)^{N-i} + (1-r) \sum_{i=j}^{N} {}^{N}C_i \, p^i (1-p)^{N-i}$$

Q.2. $P(X=i) = p(1-p)^{i-1}$, $i = 1, 2 \ldots$

~~Given X & Y are both independent, identically distributed, geometric random var. with parameter p.~~

$$P\left(\frac{X=i}{X+Y=n}\right) = P\left(\frac{A}{B}\right) = \frac{P(A\cap B)}{P(B)} \qquad i = 1 \ldots n-1$$

$$P(B) = P(X+Y=n) = \sum_{i=1}^{N-1} P(X=i)\,P(Y=n-i)$$

$$\sum_{i=1}^{N-1} \left[p(1-p)^{i-1}\right]\left[p(1-p)^{n-i-1}\right]$$

$$= \sum_{i=1}^{n-1} p^2(1-p)^{n-2}$$

$$= (n-1)\,p^2(1-p)^{n-2}$$

$$P(A\cap B) = P\left(\frac{X=i}{X+Y=n}\right)$$ since X & Y are independent

$$= P(A\cap B) = \cancel{P(X=P)}\ \underline{P(X=i \cap X+Y=n)}$$

$X + Y = n$, i, $n-i$

$$= P(X=i \wedge Y=n-i)$$

$$= P(X=i)\,P(Y=n-i)$$

$$= p(1-p)^{i-1}\; p(1-p)^{n-i-1}$$

$$= p^2(1-p)^{n-2}$$

$$\therefore P\left(\frac{X=i}{X+Y=n}\right) = \frac{P(A\cap B)}{P(B)}$$

$$= \frac{p^2(1-p)^{n-2}}{(n-1)p^2(1-p)^{n-2}} \qquad \text{for } i = 1, 2 \ldots n-1$$

$$= \frac{1}{n-1}$$

Hence, proved.

Q.3. $x_i \in R^{N\times d}$ $Y \in R^N$ $W \in R^d$

$$L_R(W) = (Y-XW)^T(Y-XW) + \lambda W^T x \qquad \lambda>0$$

$$= (Y^T - W^TX^T)(Y-XW) + \lambda W^TW$$

$$= Y^TY - W^TX^TY - Y^TXW + W^TX^TXW + \lambda W^TW$$

A $\nabla_W L_R(W) = D_W(Y^TY) - D_W(W^TX^TY) - D_W(Y^TXW) + D_W(W^TX^TXW) + D_W(\lambda W^TW)$

[using $D_x X^TAX = (A^T+A)X$]

$$D_W(Y^TY) = 0$$

$$D_W(W^T(X^TX)W) = [X^TX + (X^TX)^T]W = 2(X^TX)W$$

$$D_W(\lambda W^TW) = 2\lambda W$$

$$D_W(W^TX^TY) = D_W(\underbrace{(Y^TXW)^T}_{\text{scalar}}) = D_W(Y^TXW)^T$$

so dimensions of Y^T should be $1\times d$ as $W \to d\times 1$

$\to X^TY$ [from resources for calc^n of gradiant → google]

$$\Rightarrow D_W L_R(W) = 0 + 2(X^TX)W + 2\lambda W - 2X^TY$$

$$= 2[(X^TX)W - X^TY + \lambda W]$$

$$= 2[(X^TX+\lambda I)W - X^TY]$$

~~Adding scalar to matrix~~ so using Identity matrix

B: Minimize $L_R(W)$

find $\hat{w}$ s.t. $D_w L_R(w) = 0$

from A $\rightarrow$ $2\left[(X^T X + \lambda I)w - X^T y\right] = 0$

$$(X^T X + \lambda I)\,w = X^T y$$

$$\hat{w} = w = \underline{(X^T X + \lambda I)^{-1} X^T y}$$ soln.

C: to compare our solution with ridge regression loss.

$$L(w) = (y - Xw)^T (y - Xw)$$

we can see that $\lambda = 0$ — $L_R(w)$ & $L(w)$ same

$\therefore$ soln would be $\hat{w} = (X^T X)^{-1} X^T y$

what happens when $\lambda \to \infty$?

$$\lim_{\lambda \to \infty} \hat{w} = (X^T X + \lambda I)^{-1} X^T y$$

taking λ as common

$$\cancel{\lambda I}\left(\frac{1}{\lambda} X^T X + I\right)^{-1} (\lambda I)^{-1} X^T y$$

$\therefore \lim_{\lambda \to \infty} \hat{w}$ tends to 0.

D: Assume $X^T X = I$

& $\lambda W^T W = \underbrace{W^T D W}$ $\rightarrow$ diagonal matrix

[$D^T = D$ → Diagonal]

$$D_R(W^T X^T X W) = D_R(w^T w) = 2$$

$$D_R(\lambda w^T w) = D_R(w^T D w) = (D^T + D)\,w = 2Dw$$

$$D_w(L_R(w)) = 2(D+I)w - 2X^T y$$

for minimize put $D_w(L_R w) = 0$

$$\hat{w} = (D+I)^{-1} X^T y$$

$(D+I)^{-1}$ is diagonal $\left[\frac{1}{d_{ii}+1}\right]$ {each element is $\left[\frac{1}{1+d_{ii}}\right]$ — initial diagonal element

from B

After

$$\boxed{\hat{w} = (X^T X + \lambda I)^{-1} \quad (I + DI)^{-1} \Rightarrow (D+I)^{-1}}$$