CS 419M - Introduction to Machine Learning

Assignment - 1

Ans 1

If it rains- r * nCr * (p)^j

If it doesn't rain- (1 - r) * nCr * (q)^j

Total probability = r * nCr * (p)^j + (1 - r) * nCr * (q)^j

Ans 2

From Bayes' theorem for Conditional probability,

P (A | B) = P (B | A) * P(A)/P(B)

P (X=i | X+Y=n) = P(X+Y=n | X=i) * P(X=i) / P(X+Y = n)

= P (Y = n-1) * P(X=i) / P(X+Y = n)

Given,

$P(X = i) = p*(1-p)^{i-1}$

We have,

$P(Y = n-1) = p* (1-p)^{n-i-1}$

Also,

P(X+Y = n) = P (X=1 and Y=n-1) + P(X=2 and Y=n-2) ….. + P(X=n-1 and Y=1)

$= p\ (1-p)^0\ p\ (1-p)^{n-2} + p\ (1-p)^1\ p\ (1-p)^{n-3}\ \ldots.. + p\ (1-p)^{n-2}\ p\ (1-p)^0$

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

P (X=i | X+Y=n) = P(X+Y=n | X=i) * P(X=i) / P(X+Y = n)

= P (Y = n-1) * P(X=i) / P(X+Y = n)

$= p* (1-p)^{n-i-1} * p*(1-p)^{i-1} / [(n-1) * p^2\ (1-p)^{n-2}]$

= 1/(n-1)

Ans 3

Dimensions-

Y- Nx1

X- Nxd

W- dx1

A)

$$L_R(W) = (Y - XW)^T(Y - XW) + \lambda W^T W$$
$$= (Y^T - W^T X^T)(Y - XW) + \lambda W^T W$$
$$= Y^T Y - W^T X^T Y - Y^T XW + W^T X^T XW + \lambda W^T W$$

$$\because W^T X^T Y \ \ is\ 1x1 \Rightarrow W^T X^T Y = (W^T X^T Y)^T$$
$$\Rightarrow W^T X^T Y = Y^T XW$$

$$\frac{\delta L_R(W)}{\delta W} = \frac{\delta\,[Y^T Y - 2Y^T XW + W^T X^T XW + \lambda W^T W]}{\delta W}$$

$$\because \nabla X^T AX = (A + A^T)X$$

$$\frac{\delta L_R(W)}{\delta W} = -2Y^T X + 2(X^T X)W + \lambda(2I)W$$

Correcting dimensions,

$$\frac{\delta L_R(W)}{\delta W} = -2X^T Y + 2(X^T X)W + 2\lambda W$$

B)

Setting gradient of the loss function to 0,

$$(X^T X + \lambda I)W = X^T Y$$
$$\Rightarrow W = (X^T X + \lambda I)^{-1} X^T Y$$

C)

Closed form solution for ordinary squared-error loss regression is-

$$W = (X^T X)^{-1} X^T Y$$

Closed form solution for ridge regression loss is-

$$W = (X^T X + \lambda I)^{-1} X^T Y$$

For λ=0, both solutions are same

$$\Rightarrow W = (X^T X)^{-1} X^T Y$$

For λ=∞, the solution for ridge loss regression results to

$$W = 0$$

Since λ effectively regularizes the linear predictors W, having a very large λ as hyperparameter would result in very small values for W.

D)

$$\frac{\delta L_R(W)}{\delta W} = \frac{\delta\,[Y^T Y - 2Y^T XW + W^T X^T XW + \lambda W^T W]}{\delta W}$$

$$\frac{\delta L_R(W)}{\delta W} = \frac{\delta\,[Y^T Y - 2Y^T XW + W^T IW + W^T DW]}{\delta W}$$

$$\frac{\delta L_R(W)}{\delta W} = -2Y^T X + 2(I)W + (2D)W$$

$$\frac{\delta L_R(W)}{\delta W} = -2X^T Y + 2W + 2DW$$

Setting gradient of the loss function to 0,

$$(I + D)W = X^T Y$$

$$W = (I + D)^{-1} X^T Y$$