Safwan Mahmood
ES14BTECH11017

Assignment - 1

AML

1) (a) $P(A=x) = {}^{60}C_x \left(\frac{3}{7}\right)^x \left(\frac{4}{7}\right)^{60-x} = \dfrac{{}^{60}C_x (3^x)(4)^{60-x}}{7^{60}}$

$P(B=x) = {}^{40}C_x \left(\frac{3}{7}\right)^x \left(\frac{4}{7}\right)^{40-x} = \dfrac{{}^{40}C_x (3^x)(4)^{40-x}}{7^{40}}$

$P(C=x) = {}^{100}C_x \left(\frac{3}{7}\right)^x \left(\frac{4}{7}\right)^{100-x} = \dfrac{{}^{100}C_x\, 3^x\, 4^{100-x}}{7^{100}}$

(b) A and B are not dependent as outcome of A doesn't impact the outcome of B as we are examining sampling with replacement.

(c) Using Baye's ~~theorem~~ :

$$P(A=s \mid C=q) = \frac{P((C=q) \mid (A=s))\, P(C=q)}{P(A=s)}$$

$$= \frac{P(C=q)\,P(B=q-s)}{P(A=s)} = \frac{{}^{100}C_q \left(\frac{3}{7}\right)^q \left(\frac{4}{7}\right)^{100-q} \times {}^{40}C_{q-s} \left(\frac{3}{7}\right)^{q-s} \left(\frac{4}{7}\right)^{40-}}{{}^{60}C_s \left(\frac{3}{7}\right)^s \left(\frac{4}{7}\right)^{60-s}}$$

$$= \frac{{}^{100}C_q \left({}^{40}C_{q-s}\right)}{{}^{60}C_s} \; \frac{\left(3^{2(q-s)} \times 4^{80-2(q-s)}\right)}{7^{80}}$$

d) A and C are not independent.

We can see this by given $A = a$,

$\forall\, c \leq a$, we get $P(c) = 0$.

2) (a) We can show by $I = +ve$

$$I = \int_0^{\infty} \frac{y}{b^2} e^{-\frac{y^2}{2b^2}} dy = -\left(e^{-\frac{y^2}{2b^2}} \Big|_0^{\infty} \right)$$

$t = \frac{-y^2}{2b^2} \Rightarrow \frac{dt}{dy} = \frac{-y}{b^2}$ $\qquad = -(0-1) = 1$

using this

(b) Likelihood is same PDF.

$$\therefore L = \frac{y}{b^2} e^{-\frac{y^2}{2b^2}}$$

We get log-likelihood by taking $\log(L)$

$$\therefore \log(L) = \log y - 2\log b - \frac{y^2}{2b^2}$$

(c) Maximum likelihood estimate is max of L.

$$\therefore \frac{d(L)}{db} = \frac{d}{db}\left(\frac{y}{b^2} e^{-\frac{y^2}{2b^2}}\right) = 0$$

$\therefore$ we get $MLE(y) = \frac{y}{\sqrt{2}}$

Scanned by CamScanner

3) (a) It is a vector space $\therefore$ $k\vec{0} = \vec{0} \in \vec{V}$ and there are no two verlices so that we can use linear combination Its own vector space is $\vec{0} + \vec{0} = \vec{0}$.

(b) From Sherman_Morrison formula.

$$(A + wv^T)^{-1} = A^{-1} - \left(\frac{A^{-1}uv^TA^{-1}}{1 + v^TA^{-1}u}\right)$$

We substitute $A = A^{-1} = I$

we get = the inverse as

$$= \frac{uv^T}{1 + v^Tu}$$

(c) $\therefore$ From (b) we get $\alpha = \dfrac{1}{1 + v^Tu}$

(d) M is singular when α's denominator is zero

$1 + v^Tu = 0$ $\therefore$ $\boxed{v^Tu = -1}$

(e) M $\Rightarrow$ singular implies that $\boxed{v^Tu = -1}$

Suppose $M = (I + uv^T)u = u + uv^Tu = u + u(v^Tu) = u - u$
$= 0$

Scanned by CamScanner

Since M has rank $n-1$, nullspace of M is unidimensional vector with basis $\{u\}$.

4) (a) $\text{eigen}(M) = \begin{pmatrix} -2-\lambda & 2 \\ -6 & 5-\lambda \end{pmatrix} \bar{e} = 0$

$\therefore \quad (5-\lambda)(-2-\lambda) + 12 = 0 \Rightarrow -10 - 3\lambda + \lambda^2 + 12 = 0$

$\Rightarrow \lambda = 1, 2.$

$\therefore$ Substituting $\lambda = 1$: $\begin{pmatrix} -3 & 2 \\ -6 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$

$\therefore$ eigen vector one $= k\begin{pmatrix} 2 \\ 3 \end{pmatrix}$

eigen vector two $= k\begin{pmatrix} 1 \\ 2 \end{pmatrix}$

(b) $U = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ by taking any column permutation (p, q are arbitrary constants)

$\therefore \begin{pmatrix} p & q \\ \frac{3}{2}p & 2q \end{pmatrix} \quad \therefore \quad U = \begin{pmatrix} 1 & 1 \\ \frac{3}{2} & 2 \end{pmatrix}$

(c) $U^{-1} = \begin{pmatrix} 4 & -2 \\ -3/2 & 1 \end{pmatrix}$

$\therefore \begin{pmatrix} 1 & 1 \\ \frac{3}{2} & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 4 & -2 \\ -3/2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 3/2 & 4 \end{pmatrix} \begin{pmatrix} 4 & -2 \\ -3/2 & 1 \end{pmatrix}$

$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

5)(a) When k is value n. We know that when we make a random pick then it will have 50-50 error rate as there will be $n/2$ correct class and $\frac{n}{2}$ of incorrect class. Here error is very high.

→ Error is zero for k=1, as we get all predictions right as the point is nearest to itself.

→ As k varies from N to 1, training error decreases and finally goes to zero.

(b) For k = 1, we have high error rate with low accuracy because of overlapping of close by classes.

For k = N to 1, the error rate decreases and then doesn't decreases and keeps increasing, suggesting ~~overfitting~~ misclassification.

For k = N, error is high cause the pick is random.

5(c) Choosing k, we find the region where the error is minimal, generalization can be used for this. We generally give priority to test error to be low while selecting k.

5(d) We have the following factors :

(a) For high dimensions, calculating distances becomes a costly task.

(b) As the dimensions increase, the datapoints go sparse at a radial distance from origin leaving void at centre.