INTRODUCTION TO STASTICAL LEARNING

HOMEWORK ASSIGNMENT - 1

ANUSHA MUPPALLA

Id: 16286311

Assignment 1

1)
a) $a = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 5 \\ -4 \end{bmatrix}$ $b = \begin{bmatrix} 3 \\ 1 \\ 0 \\ -2 \\ 1 \end{bmatrix}$ $c = \begin{bmatrix} 0 \\ -3 \\ 5 \\ -1 \\ 2 \end{bmatrix}$

$a+b \Rightarrow \begin{bmatrix} 2 \\ 0 \\ -1 \\ 5 \\ -4 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \\ 0 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ -1 \\ 3 \\ -3 \end{bmatrix}$

b) $b+c \Rightarrow \begin{bmatrix} 3 \\ 1 \\ 0 \\ -2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ -3 \\ 5 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \\ 5 \\ -1 \\ 3 \end{bmatrix}$

c) $3a+4b-2c \Rightarrow 3\begin{bmatrix} 2 \\ 0 \\ -1 \\ 5 \\ -4 \end{bmatrix} + 4\begin{bmatrix} 3 \\ 1 \\ 0 \\ -2 \\ 1 \end{bmatrix} - 2\begin{bmatrix} 0 \\ -3 \\ 5 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 18 \\ 10 \\ -13 \\ 9 \\ -12 \end{bmatrix}$

d) $\vec{a}.\vec{b} \Rightarrow \begin{bmatrix} 2 \\ 0 \\ -1 \\ 5 \\ -4 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 1 \\ 0 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} (2\times3)+(0\times1)+(-1\times0)+(5\times-2) \\ +(-4\times1) \end{bmatrix} = 6-10-4$
$= -8$

e) $\vec{a}.\vec{c} \Rightarrow \begin{bmatrix} 2 \\ 0 \\ -1 \\ 5 \\ -4 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -3 \\ 5 \\ -1 \\ 2 \end{bmatrix} = -5-5-8 = -18$

f) $\vec{b}.\vec{c} \Rightarrow \begin{bmatrix} 3 \\ 1 \\ 0 \\ -2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -3 \\ 5 \\ -1 \\ 2 \end{bmatrix} = -3+2+2 = 1$

2)

$$a = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \qquad b = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \qquad c = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$$

a) $\cos\theta = \dfrac{\vec{a}\cdot\vec{b}}{|\vec{a}|\cdot|\vec{b}|} = \dfrac{7}{\sqrt{6}\cdot\sqrt{15}} = \dfrac{7}{3\sqrt{10}} = 0.73 \quad \therefore \theta = 43^\circ$

$|\vec{a}| = \sqrt{1+1+4} = \sqrt{6}$

$|\vec{b}| = \sqrt{4+1+9} = \sqrt{15}$

b) $\cos\theta = \dfrac{\vec{a}\cdot\vec{c}}{|\vec{a}|\cdot|\vec{c}|} = \dfrac{3}{\sqrt{6}\cdot\sqrt{5}} = \dfrac{3}{\sqrt{30}}, \quad \theta = \cos^{-1}\left(\dfrac{3}{\sqrt{30}}\right) = 57^\circ$

c) $\cos\theta = \dfrac{\vec{b}\cdot\vec{c}}{|\vec{b}|\cdot|\vec{c}|} = \dfrac{4}{\sqrt{14}\sqrt{5}} = \dfrac{4}{\sqrt{70}}, \quad \theta = \cos^{-1}\left(\dfrac{4}{\sqrt{70}}\right) = 61^\circ$

d) point $[1,1,1]$, vector $a = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$

$\Rightarrow 1(x-1) - 1(y-1) + 2(z-1)$

$\Rightarrow x-1-y+1+2z-2 = x-y+2z-2$

e) point $[1,1,1]$, vector $b = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$

$= 2(x-1) + 1(y-1) + 3(z-1)$

$= 2x + y + 3z - 6$

f) point $[1,1,1]$, vector $c = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$

$= -1(x-1) + 0(y-1) + 2(z-1)$

$= -x + 2z - 1$

3) Given, $A = \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}$ $B = \begin{bmatrix} 1 & 3 & 0 \\ 0 & -2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

a) characterstic polynomial of A

$\det(A - \lambda I) = 0$

$\Rightarrow \left| \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = 0$

$\Rightarrow \left| \begin{bmatrix} 1-\lambda & 3 \\ 0 & 2-\lambda \end{bmatrix} \right| = 0$

$(2-\lambda)(1-\lambda) - (3 \times 0) = 0$

$2 + \lambda^2 - 2\lambda - \lambda = 0$

$\lambda^2 - 3\lambda + 2 = 0$

characterstic polynomial is

$\lambda^2 - 3\lambda + 2$

characterstic polynomial of B

$\det(B - \lambda I) = 0$

$\left| \begin{bmatrix} 1 & 3 & 0 \\ 0 & -2 & 1 \\ 1 & 1 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} \right| = 0$

$\begin{vmatrix} 1-\lambda & 3 & 0 \\ 0 & -2-\lambda & 0 \\ 1 & 1 & 2-\lambda \end{vmatrix} = 0$

$\Rightarrow 1-\lambda \left((-2-\lambda)(2-\lambda) - 0\right) + 3$

$(0-0) + 0\,(0 - 1(-2-\lambda))$

$\Rightarrow 1-\lambda\,(-4 + \cancel{2\lambda} - \cancel{2\lambda} + \lambda^2) = 0$

$\Rightarrow 1-\lambda\,(\lambda^2 - 4) = 0$

$\Rightarrow \lambda^2 - 4 - \lambda^3 + 5\lambda = 0$

characteristic polynomial equation

b) Eigen values of A

$\det(A - \lambda I) = 0$

$\left| \begin{bmatrix} 1-\lambda & 3 \\ 0 & 2-\lambda \end{bmatrix} \right| = 0$

$\lambda^2 - 3\lambda + 2 = 0$

$\lambda^2 - 2\lambda - \lambda + 2 = 0$

$(\lambda - 2) - 1(\lambda - 2) = 0$

$\lambda = 2, \lambda = 1$

Eigen values of B

$\det(B - \lambda I) = 0$

$\begin{vmatrix} 1-\lambda & 3 & 0 \\ 0 & -2-\lambda & 0 \\ 1 & 1 & 2-\lambda \end{vmatrix} = 0$

$-\lambda^3 + \lambda^2 + \cancel{4\lambda - 4} = 0$ (5, -2)

$\lambda^2(-\lambda + 1) - 4(-\lambda + 1) = 0$

$\cancel{\lambda = \pm 4 = +2}$ $\cancel{-\lambda + 1 = 0,\ \lambda = 1}$

$\cancel{\lambda = +2, -2}$

$\cancel{\lambda = -2, +2}$

c) Eigen vectors of A

Eigen values of matrix A are $\lambda = 2, 1$

$$(A-\lambda I)\begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\left(\begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix} - 2\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)\begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\begin{pmatrix} -1 & 3 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$-x+3y=0$

$x=3y$

let $y=1, x=3$ for $\lambda=1$

$$\begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Eigen values of B are

$-\lambda^3+\lambda^2+5\lambda-2=0$

$\lambda=-2$

$\lambda = \frac{3+\sqrt{5}}{2}, \frac{3-\sqrt{5}}{2}$

Eigen vectors.

$\lambda=-2$

$$(A-\lambda I): \begin{pmatrix} 1 & 3 & 0 \\ 0 & -2 & 1 \\ 1 & 1 & 2 \end{pmatrix} - (-2)\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 3 & 3 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 4 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 3 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 4 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

by reducing matrix to $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$

$x+y=0, z=0$

assume $x=-1, y=1$ then $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$

Eigen vectors for $\lambda = \frac{3 \pm \sqrt{5}}{2}$

$$(A-\lambda I): \begin{pmatrix} 1 & 3 & 0 \\ 0 & -2 & 1 \\ 1 & 1 & 2 \end{pmatrix} - \frac{3+\sqrt{5}}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{-1-\sqrt{5}}{2} & 3 & 0 \\ 0 & \frac{-7-\sqrt{5}}{2} & 1 \\ 1 & 0 & \frac{1-\sqrt{5}}{2} \end{pmatrix}$$ by reducing

$$\begin{pmatrix} 1 & 0 & -\frac{3(2\sqrt{5}-3)}{11} \\ 0 & 1 & -\frac{7+\sqrt{5}}{22} \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$x = \frac{3(2\sqrt{5}-3)}{11} z$, $y = \frac{-7+\sqrt{5}}{22} z$

assume $z=22$, $x = 12\sqrt{5}-18$, $y = 7-$

$\lambda = \frac{3-\sqrt{5}}{2}$

$$(A-\lambda I): \begin{pmatrix} 1 & 3 & 0 \\ 0 & -2 & 1 \\ 1 & 1 & 2 \end{pmatrix} - \frac{3-\sqrt{5}}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{5}-1}{2} & 3 & 0 \\ 0 & \frac{\sqrt{5}-7}{2} & 1 \\ 1 & 1 & \frac{1+\sqrt{5}}{2} \end{pmatrix}$$

reduced to

$$\begin{pmatrix} \sqrt{5}-1 & 3 & 0 \\ 0 & \frac{\sqrt{5}-7}{2} & 1 \\ 1 & 1 & \frac{1+\sqrt{5}}{2} \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & \frac{(3)(3+2\sqrt{5})}{11} \\ 0 & 1 & \frac{-\sqrt{5}+7}{22} \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$x = \frac{-3(3+2\sqrt{5})}{11} z$, $y = \frac{\sqrt{5}+7}{22} z$, let $z=22$,

$x = -18-12\sqrt{5}$

$y = \sqrt{5}+7$

$$B = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 12\sqrt{5}-18 \\ 7-\sqrt{5} \\ 22 \end{bmatrix}, \begin{bmatrix} -18-12\sqrt{5} \\ \sqrt{5}+7 \\ 22 \end{bmatrix}$$

4)

a) It is a 'regression' problem because the observations are in numerical values. It is an inference based because here we dont need accuracy rather than predictors analysis.

Here $n = 10,000$, $p = 8$ [no of floors, age, total no of rooms, no of bedrooms, no of bathrooms, basement, lot size & school dst]

b)

It is a 'regression' problem because the % change is a numerical value. It is a prediction based because we need accurate model.

Here $n = 52$, $p=3$ [% change in British market, % change in US market & % change in German]

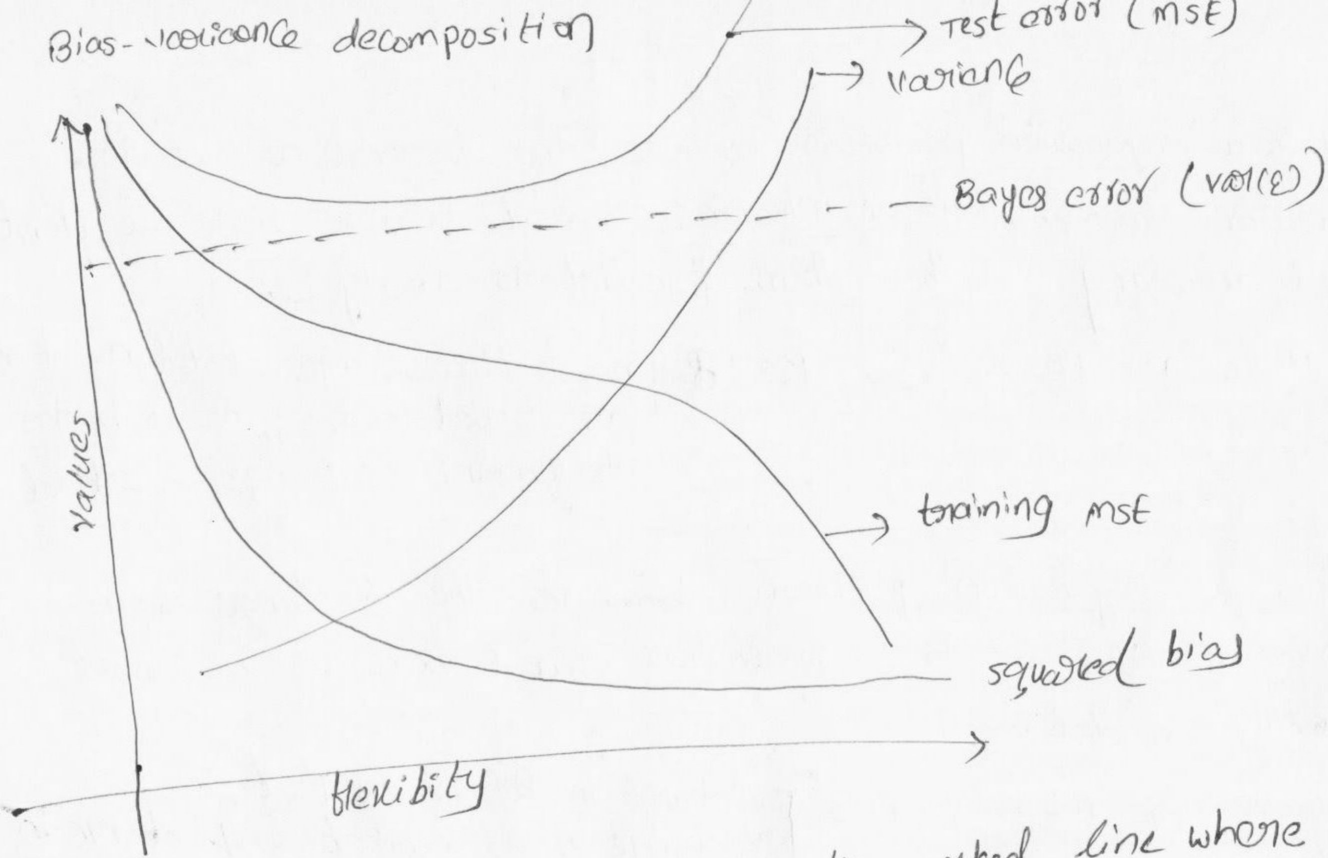
↓

because 52weeks/year.

c) It is a 'classification' problem because it is having either (High) or low risk. It is a prediction because we need accurate results of chemicals acting on a person per day

Here $n=100$, $p=30$ (chemicals)

5) Bias-variance decomposition



→ The test error begins decreasing until marked line where flexibilty starts to overfit the training data and the error starts to increase.

→ The Bayes-error is constant because it is partiular from the dataset and doesnot alter by modelling

→ the training error decreases gradually as flexibility increases and the function overfits the data and error would be minimum

→ The variance starts to increase slowly after higher levels of flexibility, the function becomes less robust and variance increases rapidly

→ The Bias decreases as flexibility increases because it is forming more complex function.

6)

a) if 'n' value is large then a flexible method would fit the data, there is a chance of reducing bias because we can train model with enough data.

b) Here, n is small and p is large, that means less data so inflexible method (linear regression) is better because we may know the trend of a graph easily.

c) same as question (a). i,e
if n is very large then we can train model with enough data and there is a less chance of bias so flexible model fit the data.

d) Flexible method is better for non linear ~~be~~ relationship between $y = f(x)$ i,e predictors and response. so that we can get less test error

e) inflexible model is better because flexible model will take errors too to fit the data and here noise is large so we can avoid flexible model

7)

X	Y	XY	X^2	Y^2
1.00	0.8	0.8	1	0.64
2.00	1.4	2.8	4	~~2.744~~ 1.96
3.00	3.8	11.4	9	14.44
4.00	3.7	14.8	16	13.69
5.00	6.2	31	25	38.44
6.00	7.0	42	36	49
7.00	9.2	64.4	49	84.64
8.00	9.3	74.4	64	86.49
9.00	10.1	90.9	81	102.01
10.00	12.2	122	100	148.84

$\Sigma x = 55.00$, $\Sigma y = 63.7$, $\Sigma xy = 454.5$, $\Sigma x^2 = 385$, $\Sigma y^2 = 540.15$ $n = 10$ (sample size)

slope $\beta_0 = \dfrac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2} \Rightarrow \dfrac{(63.7)(385) - (55.00)(454.5)}{10(385) - (55.00)^2}$

intercept $\beta_1 = \dfrac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} \Rightarrow \dfrac{10(454.5) - (55)(63.7)}{10(385) - (55)^2}$

$\beta_0 = -0.5733$

$\beta_1 = 1.26242$

$\hat{y} = \beta_0 + \beta_1 x = -0.5733 + 1.2624x$

2) $y_{15} = \beta_0 + \beta_1 (x_{15})$
$= (-0.573) + (1.2624)(15.00) \Rightarrow 18.3628$

3) $y_{18} = \beta_0 + \beta_1 (x_{18})$
$= (-0.573) + (1.2624)(18.00)$
$= 22.15$

8) A very flexible approach is great to use when the relationship $y = F(x) + \varepsilon_r$ is non linear because we have lot of datapoints to find a pattern and irreducible error is probably low

A very flexible approach would be preferred to a less flexible approach when data is in prediction format and not the interpretability of results.

A very inflexible approach is better when relationship is linear and irreducible error is high. It would be preferred to a high flexible approach when data need inference and interpretable results occur

9) In parametric approach, we assume the functional form (linear or non linear) i.e $y = f(x) + \varepsilon_r$. so that we can calculate the parameters. In nonparametric approach we dont make any assumption for function but we estimate which fits closely to the data like splines ...

Advantages:-

→ The advantages of parametric approach to regression or classification only few parameters are to be calculated for 'f' than non parametric approach.

Disadvantages:-

→ parametric approach with lots of parameters overfits the training data which leads to test errors

→ ~~Non~~ parametric models requires large data to train the model

10) Given, $X_1 = X_2 = X_3 = 0$ using k neighbours

a) Euclidean distance b/n each observation & test point.

$\sqrt{0^2+3^2+0^2} = 3$, $\sqrt{2^2+0^2+0^2} = 2$

$\sqrt{3^2+1^2+1^2} = 3.76$, $\sqrt{0^2+1^2+2^2} = 2.24$

$\sqrt{(-1)^2+0^2+1^2} = 1.414$, $\sqrt{1^2+1^2+1^2} = 1.732$

$\sqrt{0^2+1^2+2^2} = 2.24$, $\sqrt{-2^2+0^2+2^2} = 2.82$

$\sqrt{1^2+2^2+2^2} = 3$, $\sqrt{1^2+2^2+1^2} = 2.449$.

b) The point $(-1,0,1)$ is having euclidian distance of 1.414 and it has green which is closest neighbour for k=1.

c) the points $(0,3,0)$, $(1,2,2)$, $(3,1,1)$ of having Euclidian distances approximately to 3 and observations 1, 3, 9 are red as response so, we predict test point will be red

d) using Manhattan distance $\Sigma|x_i - y_i|$

observations	M.distance
1	$(0-0)+(3-0)+(0-0) = 3$
2	$(2-0)+(0-0)+(0-0) = 2$
3	$(3-0)+(1-0)+(1-0) = 5$
4	
5	$(0-0)+(1-0)+(2-0) = 3$
6	$(-1-0)+(0-0)+(1-0) = 0$
7	$(1-0)+(1-0)+(1-0) = 3$
8	$(0-0)+(1-0)+(2-0) = 3$
9	$(-2-0)+(0-0)+(2-0) = 0$
10	$(1-0)+(2-0)+(2-0) = 5$
	$(1-0)+(2-0)+(1-0) = 4$

e) As k value is large then it would fit for linear boundary. k value is small then it goes under flexible trend and it would be nonlinear boundary