Assignment 2

Question 1:

a.) After fitting the data, refer to the code, we get,

$B_0 = 1.6736$ and
$B_1 = 0.8744$ approx.

We predict and compare with the original data to get

R2 value = 0.8829

b.) After normalisation, we get

R2 value = 0.8829

Which is exactly the same as what we got before. Thus ordinary least squares method is invariant to data normalisation. However, if we had outliers in our data, Normalisation would be a good practice. Some algorithms run faster and better on normalised data.

c.) Outlier is the data that signifacantly differs from the rest of the data. They are usually more than 2 standard deviations away from the mean. One such outlier for our given data can be (-100, 100).

NEW INDIA PAGE NO.: DATE

d.)

Step1. Determine the smallest number of poin required, S.

Step2. Determine the iterations needed, N.

Step3. Determine threshhold to identify well fitting point.

Step4. Determine number of nearby points needed to assert a model fits well.

Step5. Repeat till N iterations.

Step6. Sample S points randomly from data.

Step7. Fit these S points.

Step8. For data points outside S

Step9. Find their distance from the fitted line and compare it with d value.

end.

Step10. If there are more than T such points, then it is a good fit.

Step11. Refit using all these points

end.

Step12. Use the best fit, by taking the one with the least error.

1(e) Outliers can be found out by rejecting the data that lies outside ±1.5 times the inter quartile distance. After rejecting the outliers, the data points reduce and the mean changes. We can replace the outliers by the median of the data as well since median stays fairly constant.

1f) Refer to the code.
The fitted parabola is,

$$0.09841x^2 + 0.29265x + 2.14882.$$

Question 2

a.) Data matrix has 5 features for each of the 5 samples.

5 Features of 3 samples are:

Sample 1: 12, -4, 0, 0, 0

Sample 2: 4, 12, 0, 0, 0

Sample 3: 0, 0, 2, 0, 0

b) feature covariance is given by $D^T D$ where D is the data matrix.
This gives us

$$D_a^T D_a = \begin{bmatrix} 160 & 0 & 0 & 0 & 0 \\ 0 & 160 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 2 & -2 \\ 0 & 0 & 0 & -2 & 2 \end{bmatrix}$$

c.) Eigen value decomposition of of $D^T D$ gives eigen values as :

4, 4.44×10^{-16}, 1.6×10^{2}, 1.6×10^{2} and 4

We get the eigen vector matrix as

$$E = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -0.707 & 0.707 & 0 & 0 & 0 \\ 0.707 & -0.707 & 0 & 0 & 0 \end{bmatrix}$$

Next we compute D_a diag(eigen values) D_a^T to get back the covarilance matrix.

d.) SVD of data matrix gives.

U : Its column's are eigen vectors of $D_a D_a^T$

V : Its columns are eigen vectors of $Da^T Da$

D : Square roots of non zero eigenvectors in U and V.

$$U = \begin{bmatrix} -0.3162 & -0.9486 & 0 & 0 & 0 \\ 0.9486 & -0.3162 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -0.707 & 0 & 0.707 \\ 0.707 & 0.707 & 0.707 & 0 & -0.707 \end{bmatrix}$$

$$V^T = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.707 & 0.707 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.707 & -0.707 \end{bmatrix}$$

This gives $UDV^T = Da$

(e) Scalar λ is the eigen value of a linear transformation. ~~of Da~~

If there exists v, such that $Av = \lambda v$ where A is the linear transformation.

v is called the eigenvector of λ

On the other hand. If singular value of A is σ then we have U and V such that.

$AV = \sigma U$ and $A^T U = \sigma V$.

Eigenvectors define the invariant directions of a matrix while singular vectors define the direction of maximum action. The eigen values and singular values represent the magnitude of the change or action.

NEW INDIA
PAGE NO.: DATE

Question 3

a.) Mean (M_1) = 48.833
Mean (M_2) = 55.833
Mean (M_3) = 50.833

Variance between these = 14.853

Overall mean = 51.8333

SSB = 156 $\quad$ k = 3
SSE = 96.5 $\quad$ n = 6
SST = 252.5

MSB = 78

S_x^2 = 13

b.) Data = Fit + Residue

$$S_p^2 = \frac{SSE}{k(n-1)} = 6.4333$$

c.) $$F \text{ ratio} = \frac{MSB}{MSE} = 12.125$$

Degree of freedom between group = k − 1 = 2
Degree of freedom within group = n(k − 1) = 15

d.) Null Hypothesis is that all means are the same.

p-value that we get is 0.00074

p-value provides statistical significance for the null hypothesis. For a threshnold value of α,

if p-value $< \alpha$
then strong evidence to reject null hypothesis

if p-value $> \alpha$
then null hypothesis is satisfied and gets accepted.

In our case p-value $< \alpha$ thus we reject the null hypothesis.

e.) In 2-way ANOVA test,

H_0: Means for each column is same and means for each row is same.

H_1: One mean of row-wise mean is different or one mean of column-wise means is different from the others.

For our data we have,

$$F_R = \frac{MSR}{MSE} = 4.047$$

$$F_C = \frac{MSC}{MSE} = 5.85$$

$$(F_{critical})_R = 3.71$$

$$(F_{critical})_C = 3.33$$

we see the F_R and F_C are lesser than their critical values and thus we reject the noll hypothesis.

The six different operators have different work operation.

NEW INDIA
PAGE NO.: DATE

Question 4

a. Expanding the equation of circle, we get,

$$(x-x_0)^2 + (y-y_0)^2 = r^2$$

$$\Rightarrow x^2 - 2xx_0 + x_0^2 + y^2 - 2yy_0 + y_0^2 = r^2$$

$$\Rightarrow x^2 + y^2 - 2x_0x - 2y_0y + x_0^2 + y_0^2 - r^2 = 0$$

Comparing this equation with.

$$ax^2 + \cancel{by^2}\ bxy + cy^2 + dx + ey + f = 0,$$

we get,

$a=1$ $b=0$ $c=1$
$d=-2x_0$ $e=-2y_0$
$f = x_0^2 + y_0^2 - r^2$

b. Let X denote the x- coordinates and Y denote the y-coordinates of the data.

$$A = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X^2 & XY & Y^2 & X & Y & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}_{n\times 6}$$

$$X = \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix}_{6\times 1} \qquad b = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n\times 1}$$

For least squares, we find X such that $\| AX - b \|$ is minimum.

$UDV^T = SVD(A)$

$\| AX - b \| = 0$

$\| UDV^T X - b \| = 0$

$\| DV^T X - U^T b \| = 0$

$U^T b = 0 \Rightarrow Y = V^T X$

Last column in V^T will contain the values of X that minimize $\| AX - b \| = 0$.

c. We have, $ax^2 + bxy + cy^2 + dx + ey + f = 0$

Let $X = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}_{3\times1}$ and $C = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix}_{3\times3}$

Thus, $X^T C X = 0$

$\Rightarrow \begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} ax + by/2 + d/2 \\ bx/2 + cy + e/2 \\ dx/2 + ey/2 + f \end{bmatrix} = 0$

$\Rightarrow ax^2 + \frac{bxy}{2} + \frac{dx}{2} + \frac{bxy}{2} + cy^2 + \frac{ey}{2}$
$+ \frac{dx}{2} + \frac{ey}{2} + f = 0$

Using linear transform, $AX = 0$. Thus,

$(AX)^T AX = 0$

$\Rightarrow X^T (A^T A) X = 0$ and $X^T C X = 0$

A is the design matrix, A^TA is the scatter matrix and C is the constraint matrix.

For circles, $b^2 - 4ac < 0$
We need to find X such that
$$\underset{X}{\operatorname{argmin}} \left[X^T (A^TA) X \mid X^TCX = -\phi \right]$$

Using Lagrangian multiplier,
$$\mathcal{L}(X) = X^T(A^TA)X - \lambda(X^TCX + \phi)$$

$$\Rightarrow \mathcal{L}'(X) = 0$$
$$\Rightarrow X^T(A^TA)X = \lambda X^TCX = \lambda\phi$$
$$\Rightarrow (A^TA)X = \lambda CX$$

$$\Rightarrow \frac{X}{\lambda} = (A^TA)^{-1}CX$$

Using Eigen value decomposition on $(A^TA)^{-1}C$, we can find λ^{-1}, which is the required parameter.

Question 5.

a) If all n observations, $X_1, X_2, \ldots X_n$ are independent. Likelihood function is the product of their probability functions

$$L(\theta; X_1, X_2 \ldots X_n) = \prod_{i=1}^{n} P(X_i; \theta)$$

$$= \prod_{i=1}^{n} \frac{e^{(-\theta)}\theta^{x_i}}{x_i!}$$

Taking Natural Log on both sides gives Log-likelihood function.

$$\ell(\theta; X_1, X_2 \ldots X_n) = \ln\left(\prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!}\right)$$

$$= \sum_{i=1}^{n} \ln\left(\frac{e^{-\theta}\theta^{x_i}}{x_i!}\right)$$

$$= \sum_{i=1}^{n}\left(\ln(e^{-\theta}) + \ln(\theta^{x_i}) - \ln(x_i!)\right)$$

$$= -n\theta - \sum_{i=1}^{n} \ln(x_i!) + \ln(\theta)\sum_{j=1}^{n} x_j$$

Maximum likelihood estimator for θ is

$$\hat{\theta} = \arg\max\left(\ell(\theta, x_1, x_2, \ldots x_n)\right)$$

$$\therefore \frac{d}{d\theta} l(\theta, x_1, x_2 \ldots x_n) = 0$$

$$\therefore \frac{d}{d\theta}\left(-n\theta - \sum_{i=1}^{n} \ln(x_i!) + \ln(\theta)\sum_{i=1}^{n} x_i\right)$$

$$= -n + \frac{1}{\theta}\sum_{i=1}^{n} x_i$$

$$\Rightarrow \theta = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$\hat{\theta}$ is the mean of $\{x_i\}$. This is because in poisson distribution, the expected value is eaual to the parameter θ. Thus sample moan is an unbiased estimator of the expected value.

(b) Taking log of the likelihood function simplified the mathematical calculations that followed. Otherwise taking the derivative using chain rule would be very stenous. The max value of log and original function occur at same point.

(c) $\theta =$ mean $\{15, 8, 13, 11, 7, 16, 25, 30\}$

$= 15.625$