~~SOFT COMPUTING~~

~~type~~

FEATURE ENGG

- In machine learning:
  1. Data collection.
  2. Data cleaning.
  3. Feature engineering.
  4. ML model training.

- Types of Evaluation metrics:
  - Classification.
  - Regression.
  - Ranking & Recommendation

* Classification Metrics:

Confusion matrix:

	Pred 'no'	Pred 'yes'
Act 'no'	45 TN	5 FP
Act 'yes'	5 FN	95 TP.

Total 150; 100 yes, 50 no.

* Calc

1. Accuracy: $TP+TN/TP+TN+FP+FN = 45+95/150 = 140/150 = \underline{0.93}$
2. Misclassification rate: $1 -$ Accuracy $= 1-0.93 = 0.07$
3. True +ve rate: $TP/TP+FN = 95/100 = 0.95$
4. True -ve rate: $TN/TN+FP = 45/50 = 0.9$
5. Precision: $TP/TP+FP = 95/100 = 0.95$
6. Prevalency: $\frac{TP+FN}{TP+FP+TN+FN} = \frac{100}{150} = \underline{0.67}$.

__/__/__

Qn: Confusion matrix: Multiclassifier:

True class Predicted class.

	1	2	3
1	8	2	0
2	1	9	0
3	1	2	7

Multiclass → binary.

⇒ 3 binary matrices:

1 ⇒ 1 & 2-3

	1 (yes)	2&3 (no)
(yes) 1	8 TP	2 FN
(no) 2&3	2 FP	18

2 ⇒

	2 yes	1&3 no.
2 yes	9 TP	1 FN
1&3 no	4 FP	16

3 ⇒

	3 yes	1&2 no
3 yes	7 TP	3 FN
1&2 no	0 FP	20

Metrics		Binary 1	Binary 2	Binary 3.
1. Accuracy: $\frac{TP+TN}{Total}$	:	$\frac{26}{30} = 0.86.$	0.83	0.9
2. Precision: TP/TP+FP	:	0.8	0.69	1
- Misclassification	:	1-0.86 = 0.14	0.17	0.1
3. True +ve rate: TP/TP+FN	:	0.8	0.9	0.7
True -ve rate: TN/TN+FP	:	0.9	0.8	1
Prevalency: $\frac{TP+FN}{Total}$	:	0.33	0.33	0.33.
4. F1 Score: $\frac{2P \times R}{P+R}$	:	0.8	0.78	0.82

5. Weighted F1 Score:
$\Sigma w_i \times F1score_i$

$$\frac{10 \times 0.8 + 10 \times .78 + 10 \times .82}{30} = \underline{0.8}$$

6. Macro F1 Score: $\frac{F1_1 + F1_2 + F1_3}{3} = \frac{0.8 + 0.78 + 0.82}{3} = \underline{0.8}$

_/_/_

Bias: Difference b/w actual & predicted values.
Variance: How scattered the predicted values are.
Relationship b/w the predicted values.

Overfitting : High variance, low bias.
Underfitting : Low variance, high bias.

Qu. Regression matrix:

Training set:

x_i (items)	y_i (sales)
I_1	80
I_2	90
I_3	100
I_4	110
I_5	120

2 fresh items I_6 & I_7, with actual values 80 & 75.

Test Items	Actual value	Predicted value
I_6	80	75
I_7	75	85

Find MAE, MSE, RMSE, RelMSE & CV:

$$MAE = \frac{1}{n}\sum_{i=0}^{n-1} |(y_a - y_p)| = \frac{5+10}{2} = 15/2 = \underline{7.5}$$

$$MSE = \frac{1}{n}\sum (y_a - y_p)^2 = \frac{25+100}{2} = \underline{62.5}$$

$$RMSE = \sqrt{MSE} = \sqrt{62.5} = \underline{7.9}$$

$$RelMSE = \frac{\sum (y_a - y_p)^2}{\sum (y_a - y_{avg})^2} = \frac{\cancel{62.5}\ 125}{(80-100)^2+(75-100)^2} = \frac{\cancel{625}\ 125}{1025} = \underline{0.122}$$

__/__/__

$CV = R\,value = \frac{RMSE}{\bar{y}} = \frac{7.9}{100} = \underline{\underline{0.0791}}$.

31 Dec

* Cross Validation:
To ensure generalization to unseen data.
Split into subsets and perform training & testing.

Train set: Has validation sets as well.
Validation done to see if it is valid/correct.
Parameter Tuning is done here.
Test set: Using unseen data.

Train-test splitting using random split: Hold Out method.

- K-fold: Split dataset into k parts.
K iterations of the i^{th} part being test set.
  - Sometimes, a whole class would be in test set, so none for training.

- Stratified k-fold: K-fold, but take a certain percent of each class.

- Leave one out: Leave 1/N ~~out~~ for testing, rest training.
This is done for N samples.

- Leave p-out: Randomly choosing p samples for testing.
  - Disadvantage: samples can be repeatedly in test set.

• Out of fold estimation: To produce unbiased prediction.

_/_/_

* EDA:
- Understand the Data.
- Detect Missing or Incorrect Data.
- Summarize key characteristics (mean, median, std. dev, corr.)
- Identify patterns and relations.
- Formulate Hypotheses.
- Prepare data for modelling.

1 Jan 2025.

* Curse of Dimensionality:
Upto a certain threshold limit, increased features are helpful for the model; past this threshold, the features confuse the model.

Refers to the phenomenon where the efficiency & effectiveness of algorithms deteriorate as the dimensionality (of the model increases)?

Solution:
- Feature selection.
- Feature extraction: PCA, t-SNE: t-distributed Stochastic Neighbourhood Embedding.
- Data Preprocessing: Normalization, handling missing values.

Machine Learning Life Cycle
1. Data collection.
2. Data preparation.
3. Data Wrangling.
4. Data Modelling.
5. Model Training.
6. Model Testing.
7. Deployment.

8 Jan. (1 & 3 modules :)

* PCA :
Notes ?

* LDA :

- Dimensionality reduction technique.
- Minimize ~~means~~ b/w variance b/w class values.
- Maximizing means b/w classes.

Steps :

1. Compute the class means of dependent variables:
$$\mu_i = \frac{1}{N_i} \Sigma x_i.$$

2. Derive covariance matrix of the class variable:
$$S_i = \Sigma (x-\mu_i) \cdot (x-\mu_i)^T.$$

3. Compute the within class scatter-matrix :
$$S_W = S_1 + S_2.$$

4. Compute the b/w class scatter matrix:
$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T.$$

5. Compute the eigen values & eigen vectors from S_B & S_W :
$$S_W^{-1} S_B w = \lambda w \Rightarrow |S_W^{-1} S_B - \lambda I| = 0.$$

6. Sort the values of eigen values and select the top k values.

7. Find the eigen vectors corresponding to the top k eigen vectors.
$$[S_W^{-1} S_B - \lambda I] \underbrace{\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}}_{\text{eigen vector}} = 0.$$

OR

$w = S_w^{-1} (\mu_1 - \mu_2)$.

- use when we need only for largest eigen value.

8. Obtain the LDA by taking the dot product of eigen vectors and original data.

Qu. Compute the LDA projection :

$x_1 : \{(4,2), (2,4), (2,3), (3,6), (4,4)\}$.

$x_2 : \{(9,10), (6,8), (9,5), (8,7), (10,8)\}$.

Find a new axis, to project the data values.

1. $\mu_1 = \frac{1}{5} \Sigma (x_1, x_2) = \frac{1}{5} (15, 19) = (3, 3.8)$.

$\mu_2 = \frac{1}{5} \Sigma (x_1, x_2) = \frac{1}{5} (42, 38) = (8.4, 7.6)$.

2. $S_1 = \Sigma (x - \mu_1)(x - \mu_2)^T$

$= \cancel{(4,2)+} \begin{bmatrix} 1 \\ -1.8 \end{bmatrix} \begin{bmatrix} 1 & -1.8 \end{bmatrix} + \begin{bmatrix} -1 \\ 0.2 \end{bmatrix} \begin{bmatrix} -1 & 0.2 \end{bmatrix} + \begin{bmatrix} -1 \\ -0.8 \end{bmatrix} \begin{bmatrix} -1 & -0.8 \end{bmatrix} +$

$\begin{bmatrix} 0 \\ 2.2 \end{bmatrix} \begin{bmatrix} 0 & 2.2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0.2 \end{bmatrix} \begin{bmatrix} 1 & 0.2 \end{bmatrix}$.

$S_1 = \begin{bmatrix} 1 & -1.8 \\ -1.8 & 3.24 \end{bmatrix} + \begin{bmatrix} 1 & -0.2 \\ -0.2 & 0.04 \end{bmatrix} + \begin{bmatrix} 1 & 0.8 \\ 0.8 & 0.64 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 4.84 \end{bmatrix} + \begin{bmatrix} 1 & 0.2 \\ 0.2 & 0.04 \end{bmatrix}$.

$S_1 = \cancel{\begin{bmatrix} 4 & -1 \\ -1 & 8.224 \end{bmatrix}} \cdot \begin{bmatrix} 4 & -1 \\ -1 & 8.8 \end{bmatrix}$.

$S_2 = \Sigma (x - \mu_2)(x - \mu_2)^T$.

$= \begin{bmatrix} 0.6 \\ 2.4 \end{bmatrix} \begin{bmatrix} 0.6 & 2.4 \end{bmatrix} + \begin{bmatrix} -2.4 \\ 0.4 \end{bmatrix} \begin{bmatrix} -2.4 & 0.4 \end{bmatrix} + \begin{bmatrix} 0.6 \\ -2.6 \end{bmatrix} \begin{bmatrix} 0.6 & -2.6 \end{bmatrix} + \begin{bmatrix} -0.4 \\ -0.6 \end{bmatrix} \begin{bmatrix} -0.4 & -0.6 \end{bmatrix} + \begin{bmatrix} 1.6 \\ 0.4 \end{bmatrix} \begin{bmatrix} 1.6 & 0.4 \end{bmatrix}$.

$= \cancel{\begin{bmatrix} 2.3 & -0.05 \\ -0.05 & 3.3 \end{bmatrix}} \begin{bmatrix} 9.2 & -0.2 \\ -0.2 & 13.2 \end{bmatrix}$.

__/__/__

$$S_W = S_1 + S_2 = \begin{bmatrix} 1 & -1 \\ -1 & 8.8 \end{bmatrix} + \begin{bmatrix} 9.2 & -0.2 \\ -0.2 & 13.2 \end{bmatrix}.$$

$$S_W = \begin{bmatrix} 13.2 & -1.2 \\ -1.2 & 22 \end{bmatrix}$$

$$w = S_W^{-1}(\mu_1 - \mu_2).$$

$=$

Incomplete..

Qn. Find the LD for given datasets:

$$x_1 = \{(4,2), (2,4), (2,3)\}.$$
$$x_2 = \{(3,6), (4,4), (5,5)\}.$$

1. $\mu_1 = \frac{1}{N}\sum x_i = 1/3 [8, 9] = (2.67, 3).$

$\mu_2 = \frac{1}{N}\sum x_i = 1/3 (12, 15) = (4, 5).$

2. $S_1 = \frac{1}{N-1} \sum (x-\mu_1)(x-\mu_1)^T.$

$$= \frac{1}{2}\left[\begin{bmatrix} 1.33 \\ -1 \end{bmatrix}\begin{bmatrix} 1.33 & -1 \end{bmatrix} + \begin{bmatrix} -0.67 \\ 1 \end{bmatrix}\begin{bmatrix} -0.67 & 1 \end{bmatrix} + \begin{bmatrix} -0.67 \\ 0 \end{bmatrix}\begin{bmatrix} -0.67 & 0 \end{bmatrix}\right]$$

$$= \frac{1}{2}\left[\begin{bmatrix} 1.77 & -1.33 \\ -1.33 & 1 \end{bmatrix} + \begin{bmatrix} 0.45 & -0.67 \\ -0.67 & 1 \end{bmatrix} + \begin{bmatrix} 0.45 & 0 \\ 0 & 0 \end{bmatrix}\right]$$

$$= \frac{1}{2}\begin{bmatrix} 2.67 & -2 \\ -2 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 1.335 & -1 \\ -1 & 1 \end{bmatrix}$$

$$S_2 = 1/2\left[\begin{bmatrix} -1 \\ 1 \end{bmatrix}\begin{bmatrix} -1 & 1 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix}\begin{bmatrix} 0 & -1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \end{bmatrix}\right].$$

$$= \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$$

_/_/_

3: $S_W = \begin{bmatrix} 1.335 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$.

$S_W = \begin{bmatrix} 2.335 & -1.5 \\ -1.5 & 2 \end{bmatrix}$. //

4. $S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$

$= \begin{bmatrix} -1.33 \\ -2 \end{bmatrix} \begin{bmatrix} -1.33 & -2 \end{bmatrix}$.

$= \begin{bmatrix} 1.77 & 2.66 \\ 2.66 & 4 \end{bmatrix}$. //

5. $|S_W^{-1} S_B - \lambda I| = 0$?

$\Rightarrow 0 = \begin{vmatrix} 3.1-\lambda & 4.7 \\ 3.7 & 5.5-\lambda \end{vmatrix} = (3.1-\lambda)(5.5-\lambda) - (4.7)(3.7)$
$= 17.05 - 3.1\lambda - 5.5\lambda + \lambda^2 - 17.39$

$\lambda^2 - \cancel{8.46}\lambda - 0.34 = 0$.

$b^2 - 4ac = (-8.6)^2 - 4(1)(-0.34) = 73.96 + 1.36 = 75.32$.

$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{8.6 \pm \sqrt{75.32}}{2} = \frac{8.6 \pm 8.7}{2}$

$x = 4.3 \pm 8.7 \Rightarrow \lambda_1 = 13$; $\lambda_2 = -4.4$. ← -ve ✗.

6: $\therefore \lambda = 13$ //.

Q7: $w = S_W^{-1}(\mu_1 - \mu_2)$.

$w = \begin{bmatrix} 0.83 & 0.62 \\ 0.62 & 0.96 \end{bmatrix}_{2\times2} \begin{bmatrix} -1.33 \\ -2 \end{bmatrix}_{2\times1}$.

$w = \begin{bmatrix} -2.34 \\ -2.75 \end{bmatrix}$ //.

Threshold by default = 0.7.

_/_/_

$$w = \begin{bmatrix} -2.34 \\ -2.45 \end{bmatrix}$$

8. $x_1 = \{(-9.36, -5.5), (-4.68, -11), (-4.68, -8.25)\}$.

$x_2 = \{(-7.02, -16.5), (-9.36, -11), (-11.7, -13.75)\}$.

* Feature Selection:

Supervised

① Filter method.
- ~~square~~ Chi square test.
- Correlation coeff.
- Mutual information.

② Wrapper method:
- Forward selection.
- Backward elimination.
- Recursive method.

③ Embedded method.
- Lasso regularization.

Unsupervised.

① Variance threshold.

② Dimensionality reduction (PCA).

③ Correlation Analysis.

Value $\geq$ 0.7 - strong correlation - keep feature
Value $\leq$ 0.3 - weak correlation.

__/__/____

15 Jan.

~~Covariance~~:

* Pearson's correlation coefficient:

$$r = \frac{Cov(x,y)}{\sigma_x \, \sigma_y}$$

$$= \frac{\frac{\sum_{i=1}^{n} (x_i - \mu_x) \cdot (y_i - \mu_y)}{n}}{\sum_{i=1}^{n} \sqrt{\frac{(x_i - \mu_x)^2}{n}} \cdot \sqrt{\frac{(y_i - \mu_y)^2}{n}}}$$

x	y	$x_i - \mu_x$	$y_i - \mu_y$
-6	6.7		
2	4.7		
0.2	8		
7	2		
-4	3.4		

$\mu_x = -0.16$. $r =$
$\mu_y = 4.9$.
$n = 5$.

$r = \underline{-0.47}$

Use ANOVA when a feature has 3 or more categories.
Use T-test for 2 categories.

Qn: A teacher wants to test whether 3 different teaching methods affect student test scores differently(?). The 3 methods are:

_/_/_

i. Traditional ~~tea~~ lecture (Group A).
ii. Online learning. (Group B).
iii. Blended learning. (Group C).

Group	Scores				
A	85	88	90	86	87.
B	78	~~88~~ 74	80	82	76.
C	92	94	89	91	95.

Steps:

1. H_0 & H_1: Hypothesis:

H_0: The mean of all groups are equal.
i.e. Teaching method does not affect scores.

H_1: Teaching method affects the scores.

2. Means: Group A: 87.2
B: 78.
C: 92.2.
Overall: 85.8

$SSB = \cancel{\Sigma (x_i - \mu_i)^2}$. $\Sigma(\text{class }\mu_i - \mu_g)^2 \times 5$ (rows).
$= [(87.2-85.8)^2 + (78-85.8)^2 + (92.2-85.8)^2] \times 5$.
$= (1.96 + 60.84 + 40.96) \times 5$
$= 518.8$.

$SSW: \Sigma\Sigma(x_i - \text{class }\mu_i)^2$.
$W_1 = (\cancel{4.84} + \cancel{4.84 + 0.04 + 17.64 + 0.04 + 1.44})$
$W_1 = 24$.

_/_/_

M: Overall: 85.8; A: 87.2; B = 78; C = 92.2.

$W_1 = (1.84 + 0.64 + 7.84 + 1.44 + 0.04)$
$= \underline{14.8}$

$W_2 = (16 + 4 + 16 + 4) = \underline{40}$

$W_3 = (0.04 + 3.24 + 10.24 + 1.44 + 7.84)$
$= \underline{22.8}$

$SSW = 14.8 + 40 + 22.8 = \underline{77.6}$

$Total = 518.8 + 77.6 = \underline{596.4}$

~~3. F statistic:~~

3. Degree of freedom: df_{1_b} = no. of features − 1 = 3 − 1 = 2

df_{2_w} = total elements − features = 15 − 3 = 12

4. Mean squared:

$MSB = \frac{SSB}{df_1} = \frac{518.8}{2} = \underline{259.4}$

$MSW = \frac{SSW}{df_2} = \frac{77.6}{12} = \underline{6.47}$

5. F-statistic: $\frac{MSB}{MSW} = \underline{40.1}$

6. F-critical value: from table = $\underline{3.89}$

7. Inference: Reject H_0, accept H_1.
∵ F-statistic >> F critical value.
$40.1 > 3.89$

16 Jan.

Qn: Find ICA for the given dataset:

	1	2	3	4	5	6	7	8	9	10
x_1	2.5	0.5	2.2	1.9	3.1	2.3	2.0	1.0	1.5	1.1
x_2	2.4	0.7	2.9	2.2	3.0	2.7	1.6	1.1	1.6	0.9

No of features = 2. $n = 2$.
No of samples = 10. $N = 10$.

$\mu_{x_1} = 1.81$.
$\mu_{x_2} = 1.91$.

Covariance matrix:
Ordered pair: (x,x) (x,y) (y,x) (y,y).

$$cov(x,x) = \frac{1}{N-1} \Sigma (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j).$$

$$= \frac{1}{9}\left[(2.5-1.81)^2 + (0.5-1.81)^2 + (2.2-1.81)^2 + (1.9-1.81)^2 + (3.1-1.81)^2 + (2.3-1.81)^2 + (2-1.81)^2 + (1-1.81)^2 + (1.5-1.81)^2 + (1.1-1.81)^2\right].$$

$$= \frac{1}{9}\left[0.4761 + 1.7161 + 0.1521 + 0.0081 + 1.66414 + 0.24014 + 0.0361 + 0.6561 + 0.0961 + 0.5041\right].$$

$$= \frac{1}{9}(5.549) = \underline{0.6165} = \underline{0.62}$$

$$Cov(x,y) = \frac{1}{N-1} \Sigma (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j).$$

$$= \frac{1}{9}\left[0.34 + 1.6 + 0.386 + 0.03 + 1.41 + 0.387 - 0.059 + 0.656 + 0.096 + 0.714\right]$$

$$= \frac{1}{9}(5.563) = \underline{0.618} = \underline{0.62}$$

$Cov(y,x) = \underline{0.618 = 0.62}$,

__/__/____

$$Cov(y,y) = \frac{1}{9}\left[0.24 + 1.46 + 0.98 + 0.084 + 1.188 + 0.624 + 0.096 + 0.66 + 0.096 + 1.02\right]$$

$$= \underline{0.7164}.$$

$$Cov = \begin{bmatrix} 0.6165 & 0.62 \\ 0.62 & 0.7164 \end{bmatrix}.$$

Eigen values: $\begin{bmatrix} 0.62-\lambda & 0.62 \\ 0.62 & 0.72-\lambda \end{bmatrix}.$

$= (0.62-\lambda)(0.72-\lambda) - 0.3844.$

$= 0.4464 - 0.62\lambda - 0.72\lambda + \lambda^2 - 0.3844$

$= \lambda^2 \overset{-1.34\lambda}{\cancel{-0.1\lambda}} + \cancel{0.45}\ 0.062$

$0 = \lambda^2 - 0.1\lambda + 0.062.$ $\sqrt{b^2-4ac} = 0.01 - 4(1)(0.062)$

$x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}.$ $= -0.238$???

$\lambda = \frac{0.1 \pm \sqrt{-0.238}}{2} = 0.05 \pm 0.244i.$

$0 = \lambda^2 - 1.34\lambda + 0.062.$ $\sqrt{b^2-4ac} = 1.8 - 4(1)(0.062)$

$= \sqrt{1.552}$

$\lambda = \frac{1.34 \pm 1.246}{2} = 0.67 \pm 1.246.$

$\lambda_1 = \underline{1.9}$ $\lambda_2 = -\underline{0.576}.$

$1.28x - 0.62y = 0.$

$-1.28x + 0.62y = 0.$

$0.62x - 1.18y = 0.$

$(Cov - \lambda_1 I)U_1 = 0 \Rightarrow \begin{bmatrix} -1.28 & 0.62 \\ 0.62 & -1.18 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 0.$ $x = \frac{0.62y}{1.28}.$

$\begin{bmatrix} x \\ y \end{bmatrix} =$ $\cancel{x = \frac{0.62y}{1.28}}$ $= \cancel{0.48y}.$ $\cancel{x = 1.9y}.$ $1.28x = 0.62y.$ $x = \frac{0.62y}{1.28}.$

$x = 0.48y.$

$\cancel{1.28x = 0.62y}$ $1.28x - 0.62y = 0.62x - 1.18y. \Rightarrow x(1.28+1.18)$

$\Rightarrow \cancel{2.46x =}$

___/___/___

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cancel{0} & 0.48 \\ \cancel{0.24} & 1 \end{bmatrix}.$$

Normalize the eigen vector:

$$e_1 = \begin{bmatrix} 0.48 / \sqrt{0.48^2+1^2} \\ 1 / \sqrt{0.48^2+1^2} \end{bmatrix} = \begin{bmatrix} 0.43 \\ 0.9 \end{bmatrix}.$$

Derive new data:

$$PC1_1 = e_1^T \begin{bmatrix} 2.5 - 1.81 \\ 2.4 - 1.91 \end{bmatrix} = \begin{bmatrix} 0.43 & 0.9 \end{bmatrix} \begin{bmatrix} 0.69 \\ 0.49 \end{bmatrix} = 0.4377.$$

$$PC1_2 = \begin{bmatrix} 0.43 & 0.9 \end{bmatrix} \begin{bmatrix} -1.31 \\ -1.21 \end{bmatrix} = -1.65$$

$$PC1_3 = \begin{bmatrix} 0.43 & 0.9 \end{bmatrix} \begin{bmatrix} 0.39 \\ 0.99 \end{bmatrix} = 1.06.$$

$$PC1_4 = \begin{bmatrix} 0.43 & 0.9 \end{bmatrix} \begin{bmatrix} 0.09 \\ 0.29 \end{bmatrix} = 0.3.$$

$$PC1_5 = \begin{bmatrix} 0.43 & 0.9 \end{bmatrix} \begin{bmatrix} 1.29 \\ 1.09 \end{bmatrix} = 1.54.$$

$$PC1_6 = \begin{bmatrix} 0.43 & 0.9 \end{bmatrix} \begin{bmatrix} 0.49 \\ 0.79 \end{bmatrix} = 0.92$$

$$PC1_7 = \begin{bmatrix} 0.43 & 0.9 \end{bmatrix} \begin{bmatrix} 0.19 \\ -0.31 \end{bmatrix} = -0.2$$

$$PC1_8 = \begin{bmatrix} 0.43 & 0.9 \end{bmatrix} \begin{bmatrix} -0.81 \\ -0.81 \end{bmatrix} = -1.1$$

$$PC1_9 = \begin{bmatrix} 0.43 & 0.9 \end{bmatrix} \begin{bmatrix} -0.31 \\ -0.31 \end{bmatrix} = -0.41$$

$$PC1_{10} = \begin{bmatrix} 0.43 & 0.9 \end{bmatrix} \begin{bmatrix} -0.71 \\ -1.01 \end{bmatrix} = -1.2$$

$$\therefore PC = \begin{bmatrix} 0.4 & -1.65 & 1.06 & 0.3 & 1.54 & 0.92 & -0.2 & -1.1 & -0.41 & -1.2 \end{bmatrix}.$$

_/_/_

Qu. SVD : $A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$.

$A^TA = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 13 & 12 \\ 12 & 13 \end{bmatrix}$,

$(13-\lambda)(13-\lambda)$

$A^TA = \begin{vmatrix} 13-\lambda & 12 \\ 12 & 13-\lambda \end{vmatrix} = (13-\lambda)^2 - 144$
$= 169 - 144 - 13\lambda - 13\lambda + \lambda^2$
$= \cancel{25}\ \lambda^2 - 26\lambda + 25 = 0.$

$x = \frac{-b \pm \sqrt{b^2-4ac}}{2}$ ⇒ $b^2 - 4ac = 676 - 100 = 576.$

$\cancel{\lambda = \frac{676 \pm \sqrt{576}}{2} = 338 \pm 24 \Rightarrow \lambda_1 = 362;\ \lambda_2 = 314.}$

$x = \frac{+26 \pm 24}{2} \Rightarrow \lambda_2 = \underline{25};\ \lambda_1 = \underline{1}.$

$\therefore \lambda_1 = 1 : (A^TA - \lambda I) = \begin{bmatrix} 12 & 12 \\ 12 & 12 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 0$

$12x = -12y \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} -1/\sqrt{-1^2+1^2} \\ 1/\sqrt{-1^2+1^2} \end{bmatrix} = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}$.

$\begin{bmatrix} 12 & 12 \\ 12 & 12 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$

$\therefore \lambda_2 = 25 : (A^TA - \lambda I) = \begin{bmatrix} -13 & 12 \\ 12 & -13 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 0 \Rightarrow -13x + 12y = 0.$

$12y = 13x \Rightarrow y = \frac{13}{12}x.$

$y = 1.083x \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1.083 \end{bmatrix} \Rightarrow \cancel{\begin{bmatrix} 0.68 \\ 0.735 \end{bmatrix}} \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}$ ~~$3x + 12y = 0.$~~

$V = \cancel{\begin{bmatrix} -0.707 & 0.68 \\ 0.707 & 0.735 \end{bmatrix}}\begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}$ $\Sigma = \begin{bmatrix} \sqrt{25} & 0 \\ 0 & \ \end{bmatrix}$.

/ /

$A = U \Sigma V^T$

$V = \begin{bmatrix} 0.68 & -0.707 \\ 0.736 & 0.707 \end{bmatrix}$ $\Sigma = \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$

$AV\Sigma^{-1} = U$

$AV = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 0.68 & -0.707 \\ 0.735 & 0.707 \end{bmatrix} = \begin{bmatrix} 3.51 & -0.707 \\ 3.565 & 0.707 \end{bmatrix}$ //

$u_i = \frac{1}{\sigma_i} A v_i$

$V = \begin{bmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{bmatrix}$. $\Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$ //

$u_i = \frac{1}{\sigma_i} A v_i$.

$u_i = \frac{1}{5} \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 3.535 \\ 3.535 \end{bmatrix} = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}$ //

$u_1 = \frac{1}{1} \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix} = \frac{1}{1} \begin{bmatrix} -3.535 \\ 3.535 \end{bmatrix} = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}$ //

$\therefore \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{bmatrix}$

TOXIC 

Qn. Given dataset, apply different normalizations. __/__/____

21 Jan.

* Independent Component Analysis: ICA:

Cocktail party problem.

Chirping signals: Signals that change over time.

- Dimensionality reduction by identifying independent components.

$$\underset{\text{observed signal}}{x} = \overset{\text{mixing matrix}}{A}\underset{\text{source}}{s}.$$

Reconeeing original: $s = \overset{\text{demixing matrix}}{W}x$.

$$(W = A^{-1}).$$

* Data Normalization:

1. Min-max normalization [0,1].

$$x' = \frac{x - \min}{\max - \min}.$$

Data: $x' = \underset{\min}{200}\ ;\ 300\ ;\ 400\ ;\ 600\ ;\ \underset{\max}{1000}.$

$$x' = \frac{200-200}{1000-200}\quad \frac{300-200}{1000-200}\quad \frac{400-200}{1000-200}\quad \frac{600-200}{1000-200}\quad \frac{1000-200}{1000-200}.$$

$$x' = 0\ ;\ \frac{100}{800}\ ;\ \frac{200}{800}\ ;\ \frac{400}{800}\ ;\ 1$$

$$x' = 0\ ;\ 0.125\ ;\ 0.25\ ;\ 0.5\ ;\ 1.$$

/ /

2. z-score normalization Mean-standard deviation:

$$z = \frac{x-\mu}{\sigma}$$

$x^i = 200; 300; 400; 600; 1000 \quad \mu = 500$

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}} = \sqrt{\frac{90000+40000+10000+10000+250000}{5}} = \sqrt{80000}$$

$\sigma = \underline{283}$

$$z = \frac{200-500}{283}; \frac{300-500}{283}; \frac{400-500}{283}; \frac{600-500}{283}; \frac{1000-500}{283}$$

$z = -1.06; -0.71; -0.35; 0.35; \underline{1.77}$

3. z-score normalization: Mean-Absolute deviation:

$$z = \frac{x-\mu}{A}$$

$A = |200-500| + |300-500| + |400-500| + |600-500| + |1000-500| / 5$

$A = 300+200+100+100+500/5 = 1200/5 = \underline{240}$

$$z = \frac{200-500}{240}; \frac{300-500}{240}; \frac{400-500}{240}; \frac{600-500}{240}; \frac{1000-500}{240}$$

$z = -1.25; -0.83; -0.42; 0.42; 2.1$ //

4. Normalization by Decimal Scaling:

- Find value of j.
- Smallest integer j such that
  $Max\left(\frac{v_i}{10^j}\right) \leq 1$.

$\Rightarrow$ $x = 200; 300; 400; 600; 1000$

$\frac{200}{10^3} \leq 1 = 0.2$. $\frac{400}{10^3} \leq 1 = 0.4$ $\frac{1000}{10^3} = 1$.

$\frac{300}{10^3} \leq 1 = 0.3$ $\frac{600}{10^3} \leq 1 = 0.6$

$x' = 0.2; 0.3; 0.4; 0.6; 1.$

* Binning:
- A form of quantization.
- Putting into buckets.
- Reduces overfitting.
- Smoothening data.

1- Equal frequency binning:
Input: [5, 10, 11, 13, 15, 35, 50, 55].
↳ Bin 1: [5, 10, 11, 13]
Bin 2: [15, 35, 50, 55].

2. Equal width binning: $w = \frac{max - min}{no. of bins}$

Bin 1: [min + w].
Bin 2: [min + 2w]
Bin n: [min + nw].

__/__/__

3. Customized binning:
Custom.

* Outliers:
Deviated datapoints.

Graphically: Scatter plot; box plot.

\+ IQR: Inter Quartile Range.

1. Sort the data from low to high.
2. Identify the first Quartile (Q_1), the median, third Quartile (Q_3).
3. Calculate $IQR = Q_3 - Q_1$.
4. Calculate upper fence = $Q_3 + (1.5 * IQR)$.
5. Calculate lower fence = $Q_1 - (1.5 * IQR)$.
6. Use fences to highlight any outliers, all values that fall outside the fences.
   The outliers are those values that are greater than the upper fence and lesser than the lower fence.

Qn. 1. ~~Sorted data: 22 24 26~~

26 37 24 28 35 22 31 53 41 64 29

1. Sorted: 22 24 26 28 ~~31~~ 29 31 35 37 41 53 64
2. $Q_1 = 26$; $Q_3 = \frac{41+37}{2} = 39$
3. $IQR = 39 - 26 = 13$.
4/5. Upper = $39 + (1.5 \times 13) = 58.5$; lower = $26 - (1.5 \times 13) = 6.5$

6. Outliers = 64

_/_/_

29 jan.

* Z-score based Outliers or Anomaly Detection.

$$Z = \frac{x-\mu}{\sigma}.$$

Z threshold = ±3, ±2, ±3.5, ±4.

Qu.

100	150	120	125	140	130	110	135	130	150
140	100	95	80	120	125	130	100	140	135
130	145	110	120	130	135	140	125	130	120.

$\mu = 124.67.$ $\sigma = 16.96. = \sqrt{\frac{\Sigma(x_i-\mu)}{N}}.$

$Z_{100} = \frac{100-124.67}{16.96} = \underline{-1.45}.$ $Z_{125} = \underline{0.02}.$

$Z_{150} = \underline{1.49}.$ $Z_{120} = \underline{-0.275}.$ $Z_{140} = \underline{0.904}.$

$Z_{130} = \underline{0.314}.$ $Z_{110} = \underline{-0.865}.$ $Z_{135} = \underline{0.61}.$

$Z_{95} = \underline{-1.45}.$ $Z_{80} = \underline{-2.639}.$ $Z_{145} = \underline{1.198}.$

Threshold ±3 : No outliers.
±2 : 80 is the outlier.

* LOF :

Qu. K = 2 nearest neighbours.

A(1,2) B(2,3) C(3,4) D(10,10).

① Euclidean distance formula: $\sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}$.

A(1,2) B(2,3) C(3,4) D(10,10).

$AB = \sqrt{(2-1)^2+(3-2)^2} = \sqrt{1+1} = \sqrt{2} \approx 1.4.$

$BC = \sqrt{(3-2)^2+(4-3)^2} = \sqrt{1+1} = \sqrt{2} = 1.4$

$CD = \sqrt{(10-3)^2+(10-4)^2} = \sqrt{49+36} = 9.22.$

$AC = \sqrt{(3-1)^2+(4-2)^2} = \sqrt{4+4} = 2.83.$

$AD = \sqrt{(10-1)^2+(10-2)^2} = \sqrt{81+64} = 12.04.$

$BD = \sqrt{(10-2)^2+(10-3)^2} = \sqrt{64+49} = 10.63.$

① Finding nearest neighbours:

A : B & C. C: B & A.

B : C & A. D: C & B.

② Reachability: reach dist (p,q) = max (dist (p,q), k distance).
neighbour distance

reach-dist (A,B) = max(1.4, 2.83) = 2.83.

reach-dist (A,C) = max(2.83, 1.4) = 2.83

~~reach-dist (A,D) = max(12.04,~~

reach-dist (B,A) = max (1.4, 2.83) = 2.83.

reach-dist (B,C) = max (1.4, ~~2.83~~ 1.4) = ~~2.83~~ 1.4

reach-dist (C,B) = ~~2.83~~ 1.4; reach-dist (C,A) = 2.83.

reach-dist (D,C) = max (9.22, 10.63) = ~~9.22~~ 10.63

reach-dist (D,B) = max (10.63, 1.4) = 10.63.

?

AD or CD?

$$\text{(4)} \quad LRD(p) = \frac{1}{\frac{1}{k} \sum_{q \in N_k(p)} \text{reach-dist}(p,q)}$$

$$LRD(A) = \frac{1}{\frac{1}{2}[1.4 + 2.83]} = \underline{0.47}$$

$$LRD(B) = \frac{1}{\frac{1}{2}[1.4 + 2.83]} = \underline{0.47}.$$

$$LRD(C) = \frac{1}{\frac{1}{2}[2.83 + 2.83]} = \underline{0.35}$$

$$LRD(D) = \frac{1}{\frac{1}{2}[9.22 + 10.63]} = \underline{0.1}.$$

$$\text{(5)} \quad LOF(p) = \frac{\frac{1}{k} \sum_{q \in N_k(p)} LRD(q)}{LRD(P)}.$$

* Modified z-score:

$$M_i = \frac{0.6745 \times (X_i - \text{Median})}{MAD}$$

where x_i - data point.

MAD - Median ($|X_i$ - Median$|$).

- Median Absolute Deviation.

0.6745 is a scaling factor.

/ /

Qn. Data points = 10 12 14 15 100.

Median = 14.

MAD = Median (|10-14|, |12-14|, |14-14|, |15-14|, |100-14|)

MAD = Median (4, 2, 0, 1, 86) ⇒ Median (0, 1, 2, 4, 86).

MAD = 2.

M_1 score = $\frac{0.6745 \times (10-14)}{2}$ = -1.349.

M_2 = $\frac{0.6745 \times (12-14)}{2}$ = -0.6745.

M_3 = 0. M_4 = $\frac{0.6745 \times (15-14)}{2}$ = 0.33725.

M_5 = $\frac{0.6745 \times (100-14)}{2}$ = 29.0035.

Qn. n=7; 2D plane. k=2.

X = A(1,1) B(2,2) C(2,3) D(3,3)

E(3,4) F(8,8) G(100,100).

(i). Euclidean distance:

AB = $\sqrt{(2-1)^2+(2-1)^2}$ = 1.414.

AC = $\sqrt{(2-1)^2+(3-1)^2}$ = 2.24.

AD = $\sqrt{(3-1)^2+(3-1)^2}$ = 2.83.

AE = $\sqrt{(3-1)^2+(4-1)^2}$ = 3.6.

AF = $\sqrt{(8-1)^2+(8-1)^2}$ = 9.9.

AG = $\sqrt{(100-1)^2+(100-1)^2}$ = 140.007.

BC = $\sqrt{0+1^2}$ = $\sqrt{1}$ = 1. BD = $\sqrt{1^2+1^2}$ = 1.414. BE = $\sqrt{1^2+2^2}$ = 2.24.

BF = $\sqrt{6^2+6^2}$ = 8.5. BG = 138.6

CD = $\sqrt{1^2+0}$ = 1. CE = $\sqrt{1^2+1^2}$ = 1.414.

CF = $\sqrt{6^2+5^2}$ = 7.81. CG = 137.88.

$DE = \sqrt{0+1^2} = 1.$ $DF = \sqrt{5^2+5^2} = 7.1.$ $DG = 137.2.$

$EF = \sqrt{5^2+4^2} = \underline{6.4}.$ $EG = 136.47.$ $FG = \underline{130.1}.$

② Nearest neighbours:

$A = B, C = [1.414, 2.24].$
$B = A, C = [1.414, 1].$
$C = B, D = [1, 1].$
$D = C, E = [1, 1].$
$E = C, D = [1.414, 1].$
$F = D, E = [7.1, 6.4]$
$G = E, F = [\cancel{130.1}, \cancel{136.47}]. [136.47, 130.1].$

③ Reach distance: $(p, q) = \max(\text{dist}(p,q), \text{q's neighbour})$
[first occurrence]

$A \to B = \max(1.414, 1) = \underline{1.414}.$
$A \to C = \max(2.24, 2.24) = \underline{2.24}.$

$B \to A = 1.414$; $B \to C = \max(1, 2.24) = \underline{2.24}$

$C \to B = 1.414$; $C \to D = \max(1, 1) = \underline{1}.$

$D \to C = 2.24$; $D \to E = \max(1, 1) = \underline{1}.$

$E \to C = \max(1.414, 2.24) = \underline{2.24}$; $E \to D = \underline{1}.$

$F \to D = \max(7.1, 1) = \underline{7.1}$; $F \to E = \max(6.4, 1) = \underline{6.4}.$

$G \to E = \max(136.47, 1) = 136.47.$; $G \to F = \max(130.1, 130.1) = \underline{130.1}$

4 Feb.

(4) $LRD = \dfrac{1}{1/k \sum \text{reachdist}(p,q)}$

A: $1 / \dfrac{1.414 + 2.24}{2} = 1/1.827 = \underline{0.547}$

B: $1 / \dfrac{1.414 + 2.24}{2} = 1/1.827 = \underline{0.547}$.

C: $1 / \dfrac{1.414 + 1}{2} = 1/1.207 = \underline{0.8285}$.

D: $1 / \dfrac{2.24 + 1}{2} = 1/1.62 = \underline{0.6173}$

E: $1 / \dfrac{2.24 + 1}{2} = 1/1.62 = \underline{0.6173}$.

F: $1 / \dfrac{7.1 + 6.4}{2} = 1/6.75 = \underline{0.148}$.

G: $1 / \dfrac{136.47 + 130.1}{2} = 1/133.285 = \underline{0.0075}$.

(5) $LOF = \dfrac{1/k \sum LRD(q)}{LRD(P)}$.

A: $1/2 \dfrac{[LRD(B) + LRD(C)]}{LRD(A)} = \dfrac{1/2(0.547 + 0.83)}{0.547} = \underline{1.26}$.

B: $1/2 \dfrac{(A + C)}{B} = \dfrac{1/2(0.547 + 0.83)}{0.547} = \underline{1.26}$.

C: $1/2 \dfrac{(B + D)}{C} = \dfrac{1/2(0.547 + 0.6173)}{0.8285} = \underline{0.7}$.

D: $1/2 \dfrac{(C + E)}{D} = \dfrac{1/2(0.8285 + 0.6173)}{0.6173} = \underline{1.171}$.

E: $1/2 \dfrac{(C + D)}{E} = \dfrac{1/2(0.8285 + 0.6173)}{0.6173} = \underline{1.171}$.

F: $1/2(D, E)/F = 1/2(0.6173 + 0.6173)/0.148 = \underline{4.171}$.

G: $1/2(E, F)/G = 1/2(0.6173 + 0.148)/0.0075 = \underline{51.05}$.

on. LDA, PCA, performance evaluations, overfitting/underfitting-theory..

_/_/_

∴ The outliers are F & G. (mostly G).

* ~~Types of Kernels:~~

Lower dimension → higher dimension, and the linearly separate it. Used for classification.

- Types of Kernels:

(1) Linear Kernel: $K(x,y) = x^T y$.

$K(x,y) = \phi(x) \cdot \phi(y) = x^T y$.

(2) Polynomial kernel: $K(x,y) = (x^T y)^q$.

q - degree of polynomial.

Homogeneous kernel.

For inhomogeneous kernel:

$K(x,y) = (c + x^T y)^q$.

c - constant.

q - degree of polynomial.

(3) Gaussian / RBF kernel: Radial Basis Functions.

$K(x,y) = e^{-(x-y)^2 / 2\sigma^2}$.

σ - constant.

If y = small, RBF similar to SVM;

y = large, kernel is influenced by more support vectors.

(4) Sigmoid Kernel: $K(x_i, y_i) = \tanh(k x_i y_i - \sigma)$

Qn. Consider the 2 points $x(1,2)$, $y(2,3)$; $\sigma = 1$
Apply RBF kernel, find the value of RBF kernel for these points.

$$K(x,y) = e^{-(x-y)^2/2\sigma^2}.$$

$$= \cancel{e^{-((1,2)-(2,3))^2/2}} = \cancel{e^{-(1+1)/2}} = \cancel{e^{-(1/2,1/2)}}$$

$$= 2/2 = 1.$$

$$(1-2)^2 + (2-3)^2 = 2. \qquad e^{-2/2} = \underline{\underline{0.36}}.$$

Qn. $x(1,2)$ $y(2,3)$. $c=1$. $q=2$
Linear, homo and inhomo

Linear $= x^T y = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 2 & 3 \end{bmatrix} = \underline{\underline{8}}$.

Homogeneous $= (x^T y)^q = 8^2 = \underline{64}$.

Inhomogeneous $= (c + x^T y)^q = (1+8)^2 = 9^2 = \underline{\underline{81}}$.

__/__/__

6 Job.

Qn. i Consider:

y	3	5	7	2	6.
$\hat{y}$	2.8	5.2	6.9	2.1	5.7.

Compute the evaluation metrics including MAE, MSE, RMSE & RelMSE.

ii. Apply LDA on:

$x_1 = \{(2,3)\ (3,4)\ (4,5)\ (5,6)\}$

$x_2 = \{(7,8)\ (8,9)\ (9,10)\ (10,11)\}$.

iii. y true label: 1 0 1 0 1 1 0 1 0 1

y. ~~predict~~ pred: 1 1 1 0 0 1 0 1 0 1.

Evaluate confusion matrix, accuracy, precision, recall and F1 score.

→ i. $MAE = \frac{1}{n} \sum_{i=0}^{n-1} |y_a - y_p| = \frac{1}{5}[0.2 + (0.2) + 0.1 + 0.1 + 0.3]$

$= 0.18.$

$MSE = \frac{1}{n} \sum_{i=0}^{n-1} (y_a - y_p)^2 = \frac{1}{5}(0.2^2 + (-0.2)^2 + 0.1^2 + (-0.1) + 0.3^2)$

$= \frac{1}{5}(0.19) = 0.038.$

$RMSE = \sqrt{0.038} = 0.195$

Avg = 4.6.

$RelMSE = \frac{\sum (y_a - y_p)^2}{\sum (y_a - y_{avg})^2} = \frac{0.19}{(2.56 + 0.16 + 5.76 + 6.76 + 1.96)}$

$= \frac{0.19}{17.2} = 0.011.$

_/_/_

ii. LDA : $x_1 = \{(2,3)\ (3,4)\ (4,5)\ (5,6)\}$ N=4.
$x_2 = \{(7,8)\ (8,9)\ (9,10)\ (10,11)\}$.

① $\mu_1 = 1/N \sum x_i = 1/4(14, 18) = (3.5, 4.5)$.
$\mu_2 = 1/N \sum x_i = 1/4(34, 38) = (8.5, 9.5)$.

② $S_W = S_1 + S_2$.

$$S_1 = \frac{1}{N-1} \sum (x-\mu_1)(x-\mu_1)^T.$$

$$= 1/3 \left[\begin{bmatrix} -1.5 \\ -1.5 \end{bmatrix} \begin{bmatrix} -1.5 & -1.5 \end{bmatrix} + \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix} \begin{bmatrix} -0.5 & -0.5 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} + \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix} \begin{bmatrix} 1.5 & 1.5 \end{bmatrix} \right].$$

$$= \frac{1}{3} \left[\begin{bmatrix} 2.25 & 2.25 \\ 2.25 & 2.25 \end{bmatrix} + \begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix} + \begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix} + \begin{bmatrix} 2.25 & 2.25 \\ 2.25 & 2.25 \end{bmatrix} \right].$$

$$= \frac{1}{3} \begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix} = \begin{bmatrix} 1.67 & 1.67 \\ 1.67 & 1.67 \end{bmatrix}.$$

$$S_2 = \frac{1}{N-1} \sum (x-\mu_2)(x-\mu_2)^T.$$

$$= \begin{bmatrix} 1.67 & 1.67 \\ 1.67 & 1.67 \end{bmatrix}.$$

$$S_W = S_1 + S_2 = 2 \begin{bmatrix} 1.67 & 1.67 \\ 1.67 & 1.67 \end{bmatrix} = \begin{bmatrix} 3.34 & 3.34 \\ 3.34 & 3.34 \end{bmatrix}.$$

③ $S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T = \begin{bmatrix} -5 \\ -5 \end{bmatrix} \begin{bmatrix} -5 & -5 \end{bmatrix}$.

$$= \begin{bmatrix} 25 & 25 \\ 25 & 25 \end{bmatrix}.$$

④ Eigen values: $|S_W^{-1} S_B - \lambda I| = 0$.

$|S_w^{-1} S_B - \lambda I| \Rightarrow$ Since ~~Sw~~ det $(S_w) = 0$, S_w^{-1} does not exist.

iii. y-true : 1 0 1 0 1 1 0 1 0 1.

y-pred : 1 1 1 0 0 1 0 1 0 1

Confusion matrix :

		True		
		0	1	
Predicted	0	3 TN	1 FN	4
	1	1 FP	5 TP	6
		4	6	10

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{3+5}{10} = \frac{8}{10} = 0.8 = 80\%.$$

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{5}{5+1} = \frac{5}{6} = 0.83 = 83\%.$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{5}{5+1} = \frac{5}{6} = 0.83 = 83\%.$$

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \cdot 0.83 \cdot 0.83}{0.83 + 0.83} = 0.83 = 83\%.$$