

# Government Engineering College, Modasa

## M.E. – Computer Engineering (Semester – I)

### 3710216 – Machine Learning

#### LIST OF PRACTICALS

#	Description																																																
1	<p>Given the following vectors:            A = [1, 2, 3, 4, 5, 6, 7, 8, 9 10]            B = [4, 8, 12, 16, 20, 24, 28, 32, 36, 40]            C = [10, 9, 8, 7, 6, 5, 4, 3, 2, 1]</p> <p><b>Ex. 1:</b> Find the arithmetic mean of vector A, B and C  <b>Ex. 2:</b> Find the variance of the vector A, B and C  <b>Ex. 3:</b> Find the euclidean distance between vector A and B  <b>Ex. 4:</b> Find the correlation between vectors A &amp; B and A &amp; C</p>																																																
2	Load breast cancer dataset and perform classification using Euclidean distance. Use 70% data as training and 30% for testing.																																																
3	Repeat the above experiment with 10-fold cross validation and find the standard deviation in accuracy.																																																
4	Repeat the experiment 2 and build the confusion matrix. Also derive Precision, Recall and Specificity of the algorithm.																																																
5	<p>Predict the class for <b>X = &lt; Sunny, Cool, High, Strong &gt;</b> using Naïve Bayes Classifier for given data</p> $P(C   X) = \frac{P(X   C) \cdot P(C)}{P(X)}$ <table border="1"> <thead> <tr> <th>#</th> <th>Outlook</th> <th>Temp.</th> <th>Humidity</th> <th>Windy</th> <th>Play</th> </tr> </thead> <tbody> <tr> <td>D1</td> <td>Sunny</td> <td>Hot</td> <td>High</td> <td>False</td> <td>No</td> </tr> <tr> <td>D2</td> <td>Sunny</td> <td>Hot</td> <td>High</td> <td>True</td> <td>No</td> </tr> <tr> <td>D3</td> <td>Overcast</td> <td>Hot</td> <td>High</td> <td>False</td> <td>Yes</td> </tr> <tr> <td>D4</td> <td>Rainy</td> <td>Mild</td> <td>High</td> <td>False</td> <td>Yes</td> </tr> <tr> <td>D5</td> <td>Rainy</td> <td>Cool</td> <td>Normal</td> <td>False</td> <td>Yes</td> </tr> <tr> <td>D6</td> <td>Rainy</td> <td>Cool</td> <td>Normal</td> <td>True</td> <td>No</td> </tr> <tr> <td>D7</td> <td>Overcast</td> <td>Cool</td> <td>Normal</td> <td>True</td> <td>Yes</td> </tr> </tbody> </table>	#	Outlook	Temp.	Humidity	Windy	Play	D1	Sunny	Hot	High	False	No	D2	Sunny	Hot	High	True	No	D3	Overcast	Hot	High	False	Yes	D4	Rainy	Mild	High	False	Yes	D5	Rainy	Cool	Normal	False	Yes	D6	Rainy	Cool	Normal	True	No	D7	Overcast	Cool	Normal	True	Yes
#	Outlook	Temp.	Humidity	Windy	Play																																												
D1	Sunny	Hot	High	False	No																																												
D2	Sunny	Hot	High	True	No																																												
D3	Overcast	Hot	High	False	Yes																																												
D4	Rainy	Mild	High	False	Yes																																												
D5	Rainy	Cool	Normal	False	Yes																																												
D6	Rainy	Cool	Normal	True	No																																												
D7	Overcast	Cool	Normal	True	Yes																																												

D8	Sunny	Mild	High	False	No
D9	Sunny	Cool	Normal	False	Yes
D10	Rainy	Mild	Normal	False	Yes
D11	Sunny	Mild	Normal	True	Yes
D12	Overcast	Mild	High	True	Yes
D13	Overcast	Hot	Normal	False	Yes
D14	Rainy	Mild	High	True	No

**Ans:** Label = NO

**6** For the data given in Exercise 5, find the splitting attribute at first level:

$$\text{Information Gain: } I(P, N) = -\frac{P}{S} \log_2 \frac{P}{S} - \frac{N}{S} \log_2 \frac{N}{S} = 0.940$$

$$\text{Entropy: } E(\text{Outlook}) = \sum_{i=1}^v \frac{P_i + N_i}{P + N} I(P_i, N_i) = 0.694$$

$$\text{Gain (Outlook)} = I(P, N) - E(\text{Outlook}) = 0.246$$

**Ans:**

Attribute	Gain
Outlook	0.246
Temperature	0.029
Humidity	0.151
Windy	0.048

**7** Generate and test decision tree for the dataset in exercise 5

**8** Find the clusters for following data with k = 2: Start with points 1 and 4 as two separate clusters.

i	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

**Ans:**

i	Point
C <sub>1</sub>	1, 2
C <sub>2</sub>	3, 4, 5, 6, 7

<p><b>9</b></p>	<p>Find following statistics for the data given in Exercise 1</p> <p><i>Within Class Scatter:</i> <math>S_W = \sum_{i=1}^C \sum_{x \in w_i} (x - m_i) (x - m_i)^T</math></p> <p><i>Between Class Scatter:</i> <math>S_B = \sum_{i=1}^C n_i (m_i - m) (m_i - m)^T</math></p> <p><i>Total Scatter:</i> <math>S_T = \sum_{i=1}^M (x_i - m) (x_i - m)^T</math></p>
<p><b>10</b></p>	<p>Given the following vectors:  X = [340, 230, 405, 325, 280, 195, 265, 300, 350, 310]; %sale  Y = [71, 65, 83, 74, 67, 56, 57, 78, 84, 65];</p> <p><b>Ex. 1:</b> Find the Linear Regression model for independent variable X and dependent variable Y.  <b>Ex. 2:</b> Predict the value of y for x = 250. Also find the residual for y<sub>4</sub>.</p>

### Additional Tasks:

- Mini Project in a group of max. 3 students
- Writing a research paper on selected topic from content with latest research issues in that topic