

# Ahmedabad Institute of Technology, Ahmedabad

## GUJCOST Sponsored Online National Level STTP on “Machine Learning”

Date: 07/11/2020

#	Task Description																								
1	<p>Compute Euclidean distance between vectors (A, B) and (A, C):</p> <p>A = [1, 2, 4, 7, 4, 4, 7, 3, 6, 5]                      B = [3, 1, 4, 6, 7, 3, 5, 4, 2, 5]                      C = [9, 8, 1, 1, 2, 8, 2, 9, 1, 2]</p> <p>Also find Pearson Coefficient and Manhattan distance between vectors A and B</p> $d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$ $r(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$																								
2	<p>Find within class and between class similarity for the vectors A, B and C</p> $S_B = \sum_{i=1}^c n_i (m_i - m) (m_i - m)^T$ $S_W = \sum_{i=1}^c \sum_{x \in w_i} (x - m_i) (x - m_i)^T$																								
3	<p>Partition following data using k-Means algorithm for k = 2, with random seeds 1 and 4</p> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="width: 5%;">i</th> <th style="width: 10%;">1</th> <th style="width: 10%;">2</th> <th style="width: 10%;">3</th> <th style="width: 10%;">4</th> <th style="width: 10%;">5</th> <th style="width: 10%;">6</th> <th style="width: 10%;">7</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>1.0</td> <td>1.5</td> <td>3.0</td> <td>5.0</td> <td>3.5</td> <td>4.5</td> <td>3.5</td> </tr> <tr> <td>B</td> <td>1.0</td> <td>2.0</td> <td>4.0</td> <td>7.0</td> <td>5.0</td> <td>5.0</td> <td>4.5</td> </tr> </tbody> </table>	i	1	2	3	4	5	6	7	A	1.0	1.5	3.0	5.0	3.5	4.5	3.5	B	1.0	2.0	4.0	7.0	5.0	5.0	4.5
i	1	2	3	4	5	6	7																		
A	1.0	1.5	3.0	5.0	3.5	4.5	3.5																		
B	1.0	2.0	4.0	7.0	5.0	5.0	4.5																		
4	<p>Given a data set of five objects characterized by a single continuous feature, assume that there are two clusters: C1: {a, b} and C2: {c, d, e}</p> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse; text-align: center;"> <tbody> <tr> <td style="width: 50px;"></td> <td>a</td> <td>b</td> <td>c</td> <td>d</td> <td>e</td> </tr> <tr> <td>Features</td> <td>1</td> <td>2</td> <td>4</td> <td>5</td> <td>6</td> </tr> </tbody> </table> <p>1. Calculate the distance matrix</p> <p>2. Calculate single link, complete link and average cluster distances between C1 and C2.</p> $D(c_1, c_2) = \min_{x_i \in c_1, x_j \in c_2} D(x_i, x_j)$ <p style="text-align: center;">Single Link</p> $D(c_1, c_2) = \max_{x_i \in c_1, x_j \in c_2} D(x_i, x_j)$ <p style="text-align: center;">Complete Link</p> $D(c_1, c_2) = \frac{1}{ c_1 } \frac{1}{ c_2 } \sum_{x_i \in c_1} \sum_{x_j \in c_2} D(x_i, x_j)$ <p style="text-align: center;">Average Link</p>		a	b	c	d	e	Features	1	2	4	5	6												
	a	b	c	d	e																				
Features	1	2	4	5	6																				