Vishwakarma Government Engineering College, Chandkheda

FDP on Applied Machine Learning: A Practical Based Appraoch

Date: 17/10/2018

<table>
<tr><th>#</th><th>Task Description</th></tr>
<tr><td>1</td><td>Compute Euclidean distance between vectors (A, B) and (A, C):
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]
Also find Pearson Coefficient and Manhattan distance between vectors A and B
$d(x, y) = \sqrt[L]{\sum_{i=1}^{n}(x_i - y_i)^L}$ $r(x, y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$</td></tr>
<tr><td>2</td><td>Find within class and between class similarity for the vectors A, B and C
$S_B = \sum_{i=1}^{C} n_i(m_i - m)(m_i - m)^T$ $S_W = \sum_{i=1}^{C}\sum_{x \in w_i}(x - m_i)(x - m_i)^T$</td></tr>
<tr><td>3</td><td>Partition following data using k-Means algorithm for k = 2, with random seeds 1 and 4
<table>
<tr><td>i</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr>
<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>
</table></td></tr>
<tr><td>4</td><td>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}
<table>
<tr><td></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>
</table>
1. Calculate the distance matrix
2. Calculate single link, complete link and average cluster distances between C1 and C2.
$D(c_1, c_2) = \min_{x_i \in c_1,\, x_j \in c_2} D(x_i, x_j)$ Single Link
$D(c_1, c_2) = \max_{x_i \in c_1,\, x_j \in c_2} D(x_i, x_j)$ Complete Link
$D(c_1, c_2) = \frac{1}{|c_1|}\frac{1}{|c_2|}\sum_{x_i \in c_1}\sum_{x_j \in c_2} D(x_i, x_j)$ Average Link</td></tr>
</table>

Prof. (Dr.) Mahesh M. Goyani