Practice Problems 2

Saturday, 5 April 2025 7:22 PM

1. Assume that you have ten datapoints $P_1, P_2, \ldots P_{10}$ lying at positions $1, 2, 3, \ldots 10$ on the number line respectively. Apply the k-means algorithm on this dataset for $k=3$ and $k=4$ by hand calculation. For $k=3$ assume initial cluster centers to be P_1, P_7, P_9 and for $k=4$, assume initial cluster centers as P_3, P_5, P_8, P_{10}. Run the algorithm (by hand) for 3 iterations. Also calculate the cost value for k-means at each iteration.

2. Coding Question :-

Load the breast cancer dataset we had used for the classification task. Remove the label column. Apply k-means on the unlabeled dataset now for $k = 2, 3, 4, \ldots 10$ and plot the graph of cost v/s k.

ii) Next transform the data using PCA and project the data on the first two principal components Apply k-means on this projected data with k-values $k = 2 \ldots 10$. Plot the graph of cost v/s k.