Indian Institute of Technology, Jammu

Instructions

- All questions are compulsory.
- Show all calculations and justify your answers.
- Use appropriate formulas where necessary.
- **The bonus question is just to give you a chance if you are not correct with other questions. The total marks are 10 (including bonus question.).**

---

1. Given the following dataset:

   $$\{(1, 1, \text{Class A}), (3, 3, \text{Class B}), (2, 2, \text{Class A}), (5, 5, \text{Class B})\}$$

   Predict the class of the point $(4, 4)$ using $k = 3$ and the Euclidean distance metric. Show all computations and determine the majority class.[**2 marks**]

2. You are given the dataset:

   $$\{(2, 3), (3, 3), (6, 8), (7, 8), (4, 5)\}$$

   Perform the first iteration of the K-Means clustering algorithm with $k = 2$. Assume the initial centroids are:

   $$C_1 = (2, 3), \quad C_2 = (6, 8)$$

   Calculate the updated centroids after assigning points to their nearest cluster using the Euclidean distance.[**3 marks**]

3. Using the dataset:

   $$\{(1, 1), (2, 2), (4, 4), (6, 6)\}$$

   Perform minimum distance hierarchical clustering. Compute the initial distance matrix and show the first two merging steps. Represent your solution with a partial dendrogram.[**2 marks**]

4. A dataset consists of the following points in one-dimensional space:

   $$\{1, 2, 3, 6, 7, 8\}$$

   Using a Parzen window with width $h = 2$ and the uniform kernel function:

   $$K(u) = \begin{cases} 1 & \text{if } |u| \leq 1, \\ 0 & \text{otherwise} \end{cases}$$

   Estimate the density at $x = 4$. Show all calculations.[**3 marks**]

Bonus Question [2 marks]

Discuss how the choice of parameters (k in KNN, k in K-Means, and h in Parzen Window) impacts the performance and results of each method.