Question 1 (10 points):

Calculate city block, Euclidean and supremum distances for the below data.

(a). x = (1, -1, 10, 3, 4), y = (10, -1, 4, 5, 2)

(b). x1 = (5, 4), x2 = (-2, 3)

Question 2 (25 points):

Given the lattice structure in the below picture and the transactions given in the below table, label each node with the following letter(s): I if it is infrequent; F if it is frequent; M if the node is a maximal frequent itemset. Assume that the support threshold (minimum support) is 20%.

Table

Description automatically generated

Question 3 (25 points):

Consider the training examples show in the below table for a binary classification problem

(a). Compute the Gini index for the overall collection of training examples.

(b). Compute the Gini index for the Customer ID attribute.

(c). Compute the Gini index for the Gender attribute.

Table

Description automatically generated

Question 4 (20 poionts):

Consider the following dataset for a binary class problem.

Calculate the information gain when splitting on A and B. Which attribute would the decision tree induction algorithm choose?

A picture containing text, device

Description automatically generated

Question 5 (20 poionts):

Using the below data, compute the silhouette coefficient for each point, each of the two clusters, and the overall clustering.

Table

Description automatically generated

Extra points (10 points)

Given the below four faces, darkness or number of dots represents density. Lines are used only to distinguish regions and do not represent points.

For each figure, could you use partition, hierarchical, density, and other algorithms we learned in class to find the patterns represented by the nose, eyes, and mouth? Please list at least 3 different types of algorithms and explain the pros and cons of each.

Shape, circle

Description automatically generated