Name: _____

Problem	Points	Score
1(a)	20	
1(b)	10	
1(c)	10	
1(d)	10	
2(a)	20	
2(b)	10	
2(c)	10	
2(d)	10	
Total	100	

Notes:

(1) The exam is closed books and notes except for one double-sided sheet of notes.

(2) Please indicate clearly your answer to the problem.

(3) If I can't read or follow your solution, it is wrong and no partial credit will be awarded.

Problem No. 1: Consider 5 data points: (0,1), (-1,0), which belong to class 1, and (1,0), (0, -1), and (-1/2, 1/2), which belong to class 2. In this problem we are going to walk through the K-MEANS clustering process.

   (a) Assume your initial guesses for two cluster centers are (0,0) and (1/2,1/2). Execute an iteration of K-MEANS by computing the new cluster centers and assigning the data points to the correct cluster. Use averaging to compute the new cluster center.

   (b) Assign an identity to each cluster based on a majority-voting scheme and draw the maximum likelihood decision surface.

   (c) Consider two test data points: (-3/4,3/4), which belongs to class 1, and (1/2,1/2), which belongs to class 2. Compute the probability of error based on your K-MEANS clustering.
   (d) Compute the probability of error based on a k-nearest neighbor rule. How different should this result be from (c) for large k?

Problem No. 2: Consider the same 5 data points above.

   (a) Construct a dendogram for the data.

   (b) Construct a top-down clustering (e.g., LBG) clustering (you can also think of this as a crude decision tree).

   (c) If you were to use your dendogram to do unsupervised clustering of the data, what clusters would you create (specify them by the mean and the elements associated with the cluster).

   (d) Suppose (0,1) and (1,0) occur 5 times more often than the rest of the data points. How would you adjust your strategy for clustering the data? How would that impact your decision regions?