Name:

|  |  |  |
| --- | --- | --- |
| Problem | Points | Score |
| 1(a) | 15 |  |
| 1(b) | 10 |  |
| 1(c) | 15 |  |
| 1(d) | 10 |  |
| 2(a) | 10 |  |
| 2(b) | 15 |  |
| 2(c) | 15 |  |
| 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 the data shown to the right. We are going to answer a few qualitative questions about classifiers trained on this data. Assume a Euclidean distance is used to compute probabilities and that priors and variances are ignored. Use maximum likelihood classification.



1. Draw the decision surface that would be computed using a kNN classifier where k → ∞ (in other words, all the data is used to form the decision surface). Explain.
2. What value of k is optimal (achieves a minimum error rate)? Explain.
3. Draw the decision surface that would be achieved by a Support Vector Machine (SVM) (with no slack variables). Explain.
4. Explain any differences between the SVM and kNN decision surfaces and what aspects of the data influence these differences.

**Problem No. 2**: Let’s assume that we train and classify the data to the right using a linear classifier. Again, ignore priors and variances. Use a Euclidean distance.



1. What is the minimum theoretical error rate that can be achieved?
2. There are four points in this data set. Using a kNN approach and leave-one-out cross validation, what is the minimum error rate that can be achieved?
3. Suppose a linear classifier was trained using maximum likelihood parameter estimation, and a decision surface was found that corresponds to a vertical line at the origin (y = 0). Next, suppose several iterations of discriminative training were run. How would you expect the decision surface to change? Note that since a linear classifier was used, the surface must always remain a line.
4. If you trained a decision tree on this data, what might the tree look like? Clearly explain your assumptions.