Name:

|  |  |  |
| --- | --- | --- |
| Problem | Points | Score |
| 1(a) | 15 |  |
| 1(b) | 15 |  |
| 1(c) | 10 |  |
| 2(a) | 10 |  |
| 2(b) | 10 |  |
| 2(c) | 10 |  |
| 3(a) | 10 |  |
| 3(b) | 10 |  |
| 3(c) | 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 a two-category two-dimensional classification problem in which you are given the following training data: [ω1: {(1,1), (0, 1), (1,0), (0,0}] and [ω2: {(0.5, 0), (0.5, 1), (1.5, 1), (1.5, 0)].

(a) Using maximum likelihood principles, design a classifier and compute the associated probability of error, P(E).

(b) Using Bayesian principles, design a classifier and compute the associated probability of error, P(E).

(c) Assume you are told class 2 is 10 times more likely than class 1, explain how your results for (a) and (b) would change.

**Problem No. 2**: Consider a two-state model of a coin toss: and .

1. Compute the probability that a sequence of two heads (e.g., HH) can be observed, or generated from this model.
2. What is the most likely state sequence that produced this sequence of “HH”?
3. Given a training sequence of “HTHT”, reestimate the transition probabilities. Does this result make sense? Explain.

**Problem No. 3**: Consider the problem of classifying two-dimensional data that is known not to be Gaussian and known not to have an identity covariance matrix.

1. Consider a linear transformation of the data using PCA: **.** Demonstrate that the covariance of **y** is an identity matrix if **A** is estimated using PCA.
2. Explain the difference between PCA and Linear Discriminant Analysis. How would you determine which transformation was more appropriate for your data set?
3. Describe the difference between class-dependent and class-independent PCA for a two-class classification problem. Explain the difference in decision surfaces that can be achieved by the two schemes.