Name: ____

Problem	Points	Score
1	50	
2	50	
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) Please try to make your solution legible and easy to follow. The better I can understand your thought process, the more generous I can be about partial credit.

(4) If you aren't sure how to work the details of a problem, at the very least write an outline of your solution indicating the step by step process that you think is needed to solve the problem.

(50 pts) Problem No. 1: A binary machine has been designed to output random sequences of 0's and 1's (e.g., "0", "1", "01"). You are given the training sequences "00", "01", "00", "10". You are also given an evaluation set that consists of two sequences, "000" and "111".

(a) Design a one-state fully ergodic observable discrete Markov model that is capable of generating these test sequences. You can use whatever approach you prefer to design this model provided you maximize the posterior probability of the evaluation data given the model.

(b) Compute the probability that the model produced the sequence given evaluation sequences.

(c) Explain why the model can output a three-symbol sequence given that such a sequence never occurs in the training data.

(d) Design a new discrete Markov model that achieves a higher posterior probability for the training data. Estimate its parameters and compute the posterior. Explain why it makes sense that this model has a higher posterior.

(50 pts) Problem No. 2: You are given a training set that consists of random samples from two uniform distributions. The first distribution, associated with Class 1, is a two-dimensional uniform distribution with a mean of (0,0) and extends from $-1 \leq x \leq 1$ and $-1 \leq y \leq 1$. The second distribution has a mean of (0.5,0.5) and extends from $0 \leq x \leq 1$ and $0 \leq y \leq 1$.

(a) Compute the probability of error that will be achieved on the training set (closed-set testing) by a classifier designed using a nearest-neighbor algorithm (e.g., kNN) based on a majority voting scheme (as we discussed in class).

(b) Suppose Class 1 was actually perfectly represented by a Gaussian distribution of mean (0,0) and an identity covariance matrix. Supposed Class 2 was perfectly represented by a Gaussian distribution of mean (0.5,0.5) and an identity covariance matrix. Write an equation for the probability of error if you applied the classifier trained in (a).

(c) Explain how your answers to (a) and (b) would change if you used a support vector machine rather than a kNN algorithm. Would the error rate increase or decrease if the system were trained properly?