Name: ____

Problem	Points	Score
1	40	
2	30	
3	30	
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. I will not give partial credit for ungrammatical sentences or fragmented answers. Please collect your thoughts and compose coherent answers.

(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.

(40 pts) Problem No. 1: Consider the following training data:

$$\omega_1 = \{(-2.0, 0.5), (-2.0, -0.5), (-2.0, 0.0)\}$$
$$\omega_2 = \{(2.0, 0.5), (2.0, -0.5), (2.0, 0.0)\}$$

(a) A classical Support Vector Machine is referred to as a large margin classifier because it finds weights that optimize a decision surface described by this equation:

$$\vec{y} = \vec{w} \cdot \vec{x} + b$$

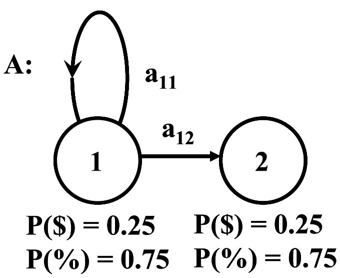
where $\vec{x}$ is a vector containing the input data, $\vec{w}$ is a weight vector and b is a scalar offset. For the training data provided, find the optimal values of $\vec{w}$ and b such that this classifier achieves zero empirical risk.

(b) Suppose you apply the k-nearest neighbor algorithm to this data set with $k = 2$. Draw the corresponding decision surface. Justify your answer and compare your solution to that in (a), describing why one or the other makes more sense.

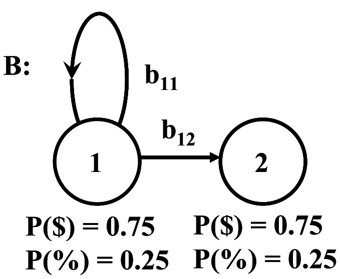
(c) Suppose you modeled each class with a Gaussian mixture model with 2 mixture components per class. Discuss what the optimal parameters of this model might be to achieve zero error on the training set.

(30 pts) Problem No. 2: For the discrete hidden Markov model shown to the right, you must start in state 1 and you must end in state 2.

Model A:



Model B:



(a) Suppose the training data for model A consists of the following sequences: "$$", "$$$", "$$$$", "%%", "%%%", "%%%%". What values of a_{11} and a_{12} maximize the probability that this model produced the training data?

(b) Suppose the training data for model B consists of sequences that are a minimum of 20 characters long (e.g., "$$$$$$$$$$%%%%%%%%%%", "$%$%$%$%$%$%$%$%$%$%"). What are the most likely values for b_{11} and b_{12}?

(c) Suppose all transition probabilities are equal. Which model was most likely to produce the sequence "$%$%"? Explain. Be as precise as possible.

(d) If you trained the parameters of these models with the Forward-Backward algorithm, and then repeated the training process with the Viterbi algorithm, would you expect to see a significant difference in the two results. Explain.

(30 pts) Problem No. 3: Consider a simple two-codeword binary code used in a digital communication system. The two codewords are "01" and "1010". Assume you decode the sequence "111" at the receiver. Obviously, this pattern has been corrupted by noise. Some of the received bits might be in error and you might have missed a bit or detected a bit that wasn't sent (you had synchronization problems for example). Using dynamic programming, determine which codeword was most likely sent. Assume the cost of incorrectly detecting a bit is 1.0. Also assume the cost of missing or inserting a bit is 2.0.

Be sure you document any assumptions you make in your solution and you explain how you have applied dynamic programming. Demonstrate your solution works by showing that you correctly decode the two codewords (consider this the zero-noise case – you receive exactly what was sent).