Name: _____

Problem	Points	Score
1	25	
2	25	
3	25	
4	25	
Total	100	

Notes:

(1) The exam is closed books and notes. You are allowed on double-sides 8.5x11 in. page of notes. If you don't have an equation that you think you need, please ask me during the exam.

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

(3) Note that ungrammatical sentences, incoherent statements, or general illegible scratches will get zero credit.

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

(5) Although each problem has multiple parts, your grade will be based on your overall response to the entire problem.

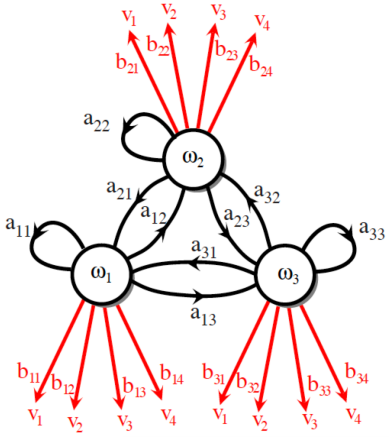
Problem No. 1 (25 points): The calculations we introduced for statistical significance ultimately hinge on the assumption that the probability distribution of potential estimates of a true parameter is Gaussian. Assume the probability distribution of the estimate of a parameter in an experiment, such as the error rate, is a uniform distribution over the range [-1,1], and that the true value of the parameter is 0. How would you modify our calculations for whether an estimate of the error rate is statistically significant?

Problem No. 2 (25 points): You are given a fully ergodic three-state HMM that outputs sequences of the characters "H" (first column of the matrix B) and "T" (second column of the matrix B):

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1c} \\ \vdots & \ddots & \vdots \\ a_{c1} & \cdots & a_{cc} \end{bmatrix} = \begin{bmatrix} 0.50 & 0.25 & 0.25 \\ 0.25 & 0.50 & 0.25 \\ 0.25 & 0.25 & 0.50 \end{bmatrix}$$

$$B = \begin{bmatrix} b_{11} & \cdots & b_{1M} \\ \vdots & \ddots & \vdots \\ b_{c1} & \cdots & b_{cM} \end{bmatrix} = \begin{bmatrix} 0.75 & 0.25 \\ 0.75 & 0.25 \\ 0.75 & 0.25 \end{bmatrix}$$

$$\pi^c = \{\pi_1, \pi_2, \ldots, \pi_c\} = \{0.5, 0.25, 0.25\}$$



(a) Using the Viterbi algorithm, compute the probability that this model output the sequence "HT".

(b) Using the Forward algorithm, compute the probability that this model output the sequence "HT".

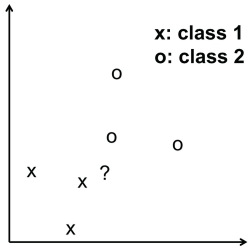
(c) Explain in terms of machine learning concepts why these probabilities are different and comment on the significance of this. Be very specific – don't simply repeat your calculations. For example, which probability is larger? Does this make sense?

Problem No. 3 (25 points): Given the data shown to the right, and assuming each class is equally probable, do the following using a Euclidean distance measure:

(a) Classify the data point shown using K-Nearest Neighbors (KNN). Use a value of $K = 2$. Draw the decision surface for this classifier. How would it label the data point marked '?'?

(b) If you were to implement a simple Support Vector Machine (SVM) with a hard margin (this means no soft margin classifiers), no kernel function, and one support vector per class, sketch the decision surface that would result. How would it label the data point marked '?'?

(c) Explain any similarities or differences between these classifiers. Why are they similar? Why are they different? Don't simply describe the pictures you have drawn. Analyze the results and justify your conclusions.



Problem No. 4 (25 points): 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 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).