Name:

|  |  |  |
| --- | --- | --- |
| Problem | Points | Score |
| 1(a) | 10 |  |
| 1(b) | 10 |  |
| 1(c) | 5 |  |
| 2(a) | 10 |  |
| 2(b) | 10 |  |
| 2(c) | 5 |  |
| 3(a) | 10 |  |
| 3(b) | 10 |  |
| 3(c) | 5 |  |
| 4(a) | 25 |  |
| 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. 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.

**Problem No. 1**: You are given a fully ergodic three-state HMM that outputs sequences of the characters “H” (first column of the matrix ) and “T” (second column of the matrix ):

(10 pts) (a) Using the Viterbi algorithm, compute the probability that this model output the sequence “HT”.

(10 pts) (b) Using the Forward algorithm, compute the probability that this model output the sequence “HT”.

(5 pts) (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. 2:** Given the data shown to the right, and assuming each class is equally probable, do the following using a Euclidean distance measure:



(10 pts) (a) Classify the data point shown using -Nearest Neighbors (KNN). Use a value of . Draw the decision surface for this classifier. How would it label the data point marked ‘?’?

(10 pts) (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 ‘?’?

(5 pts) (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. 3:** Given the data for problem no. 2, and again assuming equiprobable classes and a Euclidean distance measure:

(10 pts) (a) Cluster class 1 using K-Means and cluster class 2 using K-Means. Use in both cases. Sketch the resulting decision surface if you were to use these clusters to classify data.

(10 pts) (b) Explain how you would model the posterior for each class using a Parzen window approach with a Gaussian kernel.

(5 pts) (c) Explain the differences between (b) and a Gaussian mixture model that uses three mixture components for each distribution.

**Problem No. 4:** You have been hired by a famous Hollywood producer to write the next script for the Mission Impossible movie franchise (or whatever action/adventure movie series or television crime series you are familiar with ☺ Your script must be as realistic as possible.

(25 pts) (a) The producer wants you to include a scene in which you identify a potentially dangerous criminal using Facebook, Amazon, and Netflix data. This criminal is, of course, not stupid, so simple searches won’t work. This person has been largely invisible on the Internet for years. However, you do know that this person was born in the mythical country of Ruritania (*https://tvtropes.org/pmwiki/pmwiki.php/Main/Ruritania*) in 1960, has relatives in the United States, and collected 1965 Chevrolet Corvette cars.

Explain how you might find this mysterious person using concepts developed in this course. The better you make my course look, the higher your grade ;)