ME 466 Introduction to AI
Fall 2021
Programming Assignment 2
Due: 7 November 2021 @23:59

The solutions of this assignment should be prepared according to the guidelines in the course syllabus. Please read the syllabus first. Submit all evidence of your work; i.e., code, algorithms, flowcharts, MATLAB Command Window/command prompts, inputs, outputs, plots, results, error messages, etc. as an appendix or in the main body of your submission, with proper explanations. Also submit your code and a text file with instructions to run it. **Please fill in the page at the end of this assignment, sign it, and use it as the cover page of your assignment report.**

---

1. (2014 Midterm Question) You are an astronaut who has just landed on a distant and unknown planet, with a mission to determine if the sun rises there every day. Your observation capabilities on Earth were limited, but you were able to tell that with prior probability 0.6, the sun rises on that planet every day; or, with prior probability 0.4, it rises only 50% of the time.

(a) You stay on the planet for three days, and the sun has risen every day.

  (a) What is the probability that the sun always rises on that planet?

  (b) What is your decision? What is the probability of error for your decision?

(b) Each day you stay on the planet costs \$10000. If you decide that the sun will rise every day, your team will build a solar power plant that costs \$10 million. But if your decision is wrong, if the sun does not rise even for one day, the power plant will be destroyed and \$10 million will be wasted. How many days would you stay and see the sun rise before deciding to build the plant?

2. Consider a two-category classification problem in two dimensions. The CCPDFs are Gaussians given by $p(\mathbf{x}|w_1) \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $p(\mathbf{x}|w_2) \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ where

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$
$$\boldsymbol{\mu}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

(a) For equal prior probabilities, find the equation of the Bayes decision boundary, plot it, and define the decision regions. MATLAB's Symbolic Toolbox and *fimplicit()* function might be helpful.

(b) Repeat part (a) if the prior probabilities are $P(w_1) = 1/4$ and $P(w_2) = 3/4$.

(c) Repeat parts (a) and (b) for the following:

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix}$$
$$\boldsymbol{\mu}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$$

(d) Give one possible set of $\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2$ values for which the decision boundary is an ellipse.

3. In this problem, you will work on the Wisconsin Breast Cancer data set. To download the data, please follow these steps:

i. Go to the web site `https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)`. Read the information on the page. The data consist of 569 instances that contain features extracted from images of either benign (healthy) or malign (diseased) cells.

ii. Click the 'Data Folder' link and download the text file wdbc.data. Each row in this file corresponds to a sample. As explained in the "Attribute Information" section of the web page, the first attribute is the ID number, the second attribute is the class ('M' for malign and 'B' for benign). Attributes 3–32 are the features extracted from the image of the cell nucleus, details of which are explained in that section.

iii. Read the data file in MATLAB. You can open MATLAB, right-click on the file, and select 'Import Data.' Note that since there are both characters and real numbers in the file, MATLAB will probably put the data into a *cell* or *table* data type. You might need to manipulate the table a bit to represent the data appropriately. *table2array()* function might be useful.

After correctly reading the data, you should now have 569 samples with 30 features. You can either store the class information for the samples in a 0–1 vector (e.g., 0 for benign and 1 for malign), or work with 'B' and 'M' characters as in the original data file. The problem is to train a classifier to determine if a cell is benign or malign.

(a) Divide the data randomly into training and test sets. Use a 10-fold cross validation scheme. You may find MATLAB's *cvpartition()* function helpful. For each fold, the prior probabilities of benign/malign classes can be assigned as the fraction of benign/malign samples. Note that in reality, many other factors may contribute to the prior probability, such as the patient's age, medical history, etc.

(b) For each fold, assume Gaussian distribution for the classes. Estimate the means and covariances, and apply the MAP decision rule to classify the cell as benign or malign. Determine the confusion matrix.

(c) In this part, determine the probability distributions using kernel density estimation with a Gaussian kernel and apply the MAP rule for classification. Determine an appropriate window size h by trial and error. Determine the confusion matrix.

(d) If a classifier falsely diagnoses a healthy cell with cancer, we say that a "false alarm," a "false positive," or a "Type I error" has occurred. Conversely, if a diseased cell is falsely classified as healthy, we say that a "Type II error" or a "miss" has occurred. How many Type I and Type II errors does each method make?

(e) Apparently a Type II error is more risky. Suppose we would like to reduce this error (possibly at the expense of increasing Type I error). Suggest a method to reduce Type II error.

(f) Apply your method for parts (b) and (c) and demonstrate that the Type II error is reduced.

ME 466 Introduction to AI
Fall 2021
Programming Assignment 2
Submitted on: 7 November 2021

Name:

Student ID:

Grade:

1	
2	
3	
Σ	

I hereby declare that the paper I am submitting under this cover is product of my own efforts only. Even if I worked on some of the problems together with my classmates, I prepared this paper on my own, without looking at any other classmate's paper. I am knowledgeable about everything that is written under this cover, and I am prepared to explain any scientific/technical content written here if a short oral examination about this paper is conducted by the instructor. I am aware of the serious consequences of cheating.

Signature: ____________________________