ECS 132 Spring 2024: Assignment 6

Due May 31, 2024 11:59pm

Instructions

1. You may in no circumstances upload your homework to private tutoring websites such as CourseHero or Chegg. Remember all material related to this course is a property of the University of California and posting them is a violation of the copyright laws.
2. If you refer to a source (either a book or the internet), you must cite it.
3. You are highly urged to work on these problems on your own at first. Not getting the answer correct will incur a low penalty, and trying it and then figuring out where you went wrong will really help you understand the material better and you will be much better prepared for the exams. If you do discuss with others, you must list their names.
4. You will submit a pdf file generated via latex (or a Jupyter notebook) to Gradescope. **Note, both your code and output of the code must be included in the pdf file submitted to Gradescope!** The file will contain your written work and your Rcode embedded in the pdf file. You will also submit either the R code file or a corresponding Jupyter notebook via Canvas.

There are 4 problems. This problem set covers:

- Problem 1: Binary classification and ROC curves
- Problem 2: Binary classification and the confusion matrix
- Problem 3: A 2-state Markov chain
- Problem 4: A 3-state Markov chain

1 Problem

In this problem we will plot the ROC curve for the example that we did class. We had defined the following two random variables

$$D = \begin{cases} 1 & \text{if sample is infected} \\ 0 & \text{if sample is not infected} \end{cases}$$

$$T = \begin{cases} 1 & \text{test classifies the sample as infected} \\ 0 & \text{test classifies the sample as not infected} \end{cases}$$

Assume that given D=0, the random variable follows $X|(D = 0) \sim Norm(50, 10)$ and that given D=1 then $X|(D = 1) \sim Norm(70, 15)$.

1. Plot the pdf and the cdf of $X|(D = 1) \sim Norm(70, 15)$ for values 20 to 120.
2. Let x^* denote the cutoff value such that if $X > x^*$ then the test classifies the sample as positive, otherwise it is considered be negative. Plot the ROC curve for values of x^* ranging from $52 < x \leq 65$.
3. If the False Positive Rate and the False Negative Rates are equally bad, determine the value of x^*. (Hint, plot FRP and FNR both as a function of x^* and see where they intersect.)

2 Problem – confusion matrix

Consider this confusion matrix associated with a binary classifier that determines whether or not a sample is infected:

$$\begin{pmatrix} TP & FP \\ FN & TN \end{pmatrix} = \begin{pmatrix} 15 & 10 \\ 10 & 100 \end{pmatrix}$$

1. What is the specificity of the test, θ?
2. What is the sensitivity of the test, η?
3. What is the prevalence of the disease, π?

3 Problem – 2-state Markov Chain

In a Binary Symmetric (Communication) Channel (BSC) data is sent data is sent using bits 0 and 1. When the source and the destination are far apart, there are repeaters that decode the bit and then retransmit the signal. However due to noise, there is decoding error, i.e., there is a probability α that a 0 bit will be decoded as (and hence retransmitted) as 1. Similarly, there is a probability β that a 1 bit will be decoded as (and hence retransmitted as) 0. Let X_0 be the bit's initial parity (i.e., 0 or 1) and and let X_n be the bits parity after the nth repeater. Since this is a 2-state Markov Chain it can be solved exactly. Express the answers to the questions below in terms of mathematical equations.

1. Write down the one-step transition matrix for this Markov Chain.
2. Suppose the input stream to this communication channel consists of 80% 0s and 20% 1s. Determine the proportion of 0s and 1s after the first repeater.
3. Under the same input values as in (b) determine the proportions of 0s and 1s exiting the 5th relay.
4. Determine the proportions of 0s and 1s as the number of relays goes to infinity.

4 Problem – 3-state Markov Chain

Consider a sensor that has three states: (0) awake; (1) asleep; (2) deep sleep. At the start of each second the sensor can change its state. If it is awake it stays awake with probability 0.5, otherwise it goes asleep. If it is asleep it stays asleep with probability 0.5, it goes to awake with probability 0.25, and it goes to deep sleep with probability 0.25. If it is in deep sleep it stays in deep sleep with probability 0.7 otherwise it goes to the asleep state.

1. Write out the state transition matrix (in right-stochastic form).
2. Draw the state diagram including the arrows for all the non-zero state transition probabilities. (You can draw this by hand and take a picture to embed in your pdf.)
3. The sensor's activity has reached the steady-state distribution. What is the probability the sensor is awake? (You can use R to solve this problem.)