NNSE784-3510 – Applications of Machine Learning in Data Science

Project 2

1 – A mother has a brother with down syndrome but her father does not have down syndrome. Assuming his father is not a carrier of the gene and the gene has come from her mother to her brother, there is 50% chance that the woman is also a carrier of the gene. Assume the Random Variable (R.V) α represents if she carries the gene or not:

(α = 0): Does not have the gene

(α = 1): Has the gene

(a) What is the prior distribution P(α)?

(b) If she is not a carrier her children cannot have down syndrome but if she is a carrier there is 50% chance that they are born with down syndrome. Assume probability of a child having down syndrome is independent from others and Xi = 0 represents the event of child number i does not have down syndrome and Xi = 1 is the event that the child has down syndrome. If the mother has three children and none of them has down syndrome what is the likelihood function?

(c) Find the Posterior distribution (i.e. P(α |X1,X2,X3 = 0,0,0)).

2- A Mechanic shop buys three brands of bearings to use for client repairs. Each bearing has a probability of failure θ which follows the distribution given in the first row of Table 1. The Mechanic shop uses the brands according to probabilities in the second row.

If two of the nine repairs done today are returned as broken due to bearing failure, what is the posterior distribution? (i.e. P(θ|X) where X = 2 of the 9 repairs are broken)

Table 1 - Bearing Failure and Use Distribution

	Bearing A	Bearing B	Bearing C
Probability of Failure θ	0.05	0.1	0.15
Probability of Using Brand	0.3	0.5	0.2

3 – A factory production line produces toys which are defective with probability θ.

(a) If θ follows a uniform distribution in the range of 0 <θ<1; what is the posterior distribution P (θ | D) where D is the random variable representing the item being defective (D=1) or not defective (D=0).

(b) Solve (a) if P(θ) = $\frac{1}{\theta(1-\theta)}$.

4- A Covid-19 test has resulted in two gaussian distribution of positive and negative groups with Probability Density Functions (PDF) as shown in Figure 1. Assume the positive group is distributed with G (μ=10, σ=5) and the negative group follows G (μ=1, σ=2).

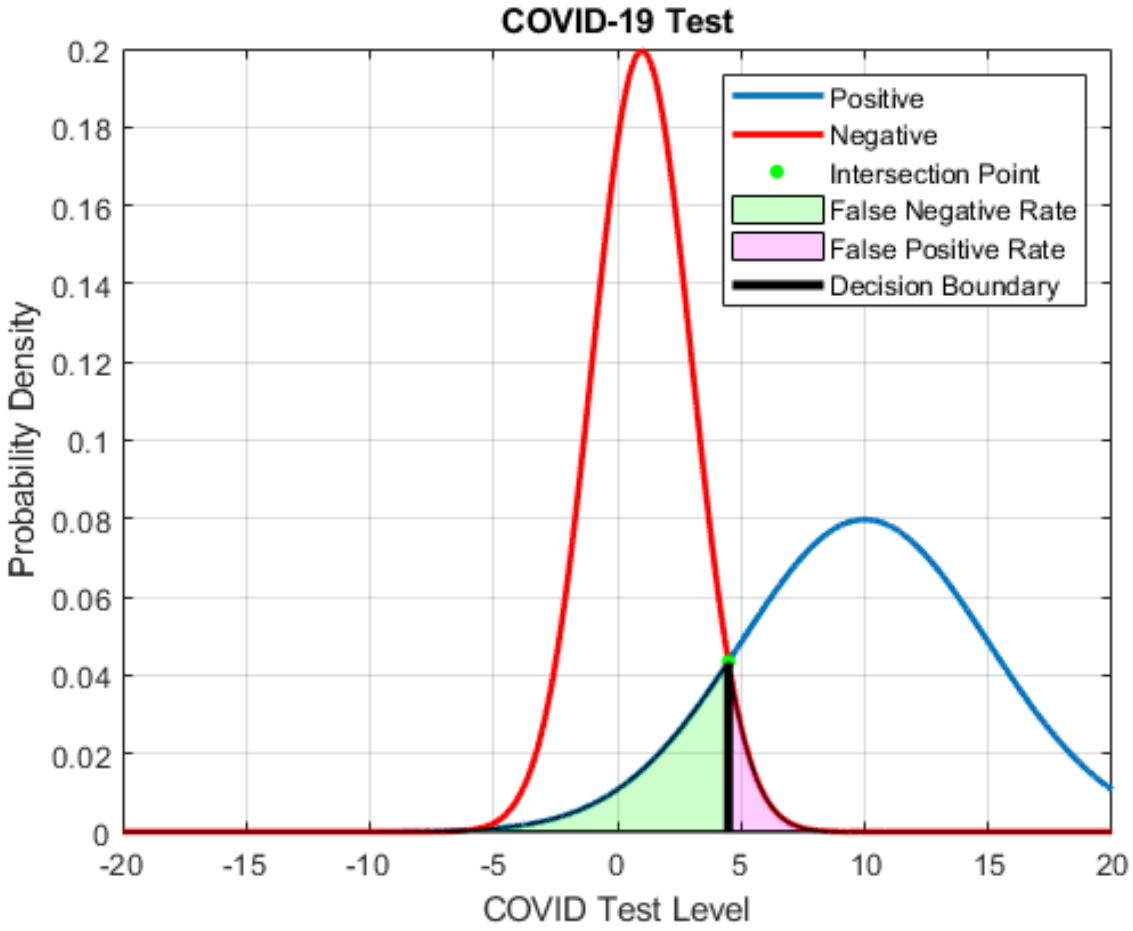

Figure 1 - PDF for Positive and Negative Populations

(a) Plot the PDF of both distributions and color the False Positive Rate (FPR) and False Negative Rate (NPR) in the curves. Use MATLAB or Python.

(b) If the cross point between the two curves is selected as the decision boundary, calculate the FPR and NPR. Use MATLAB or Python.

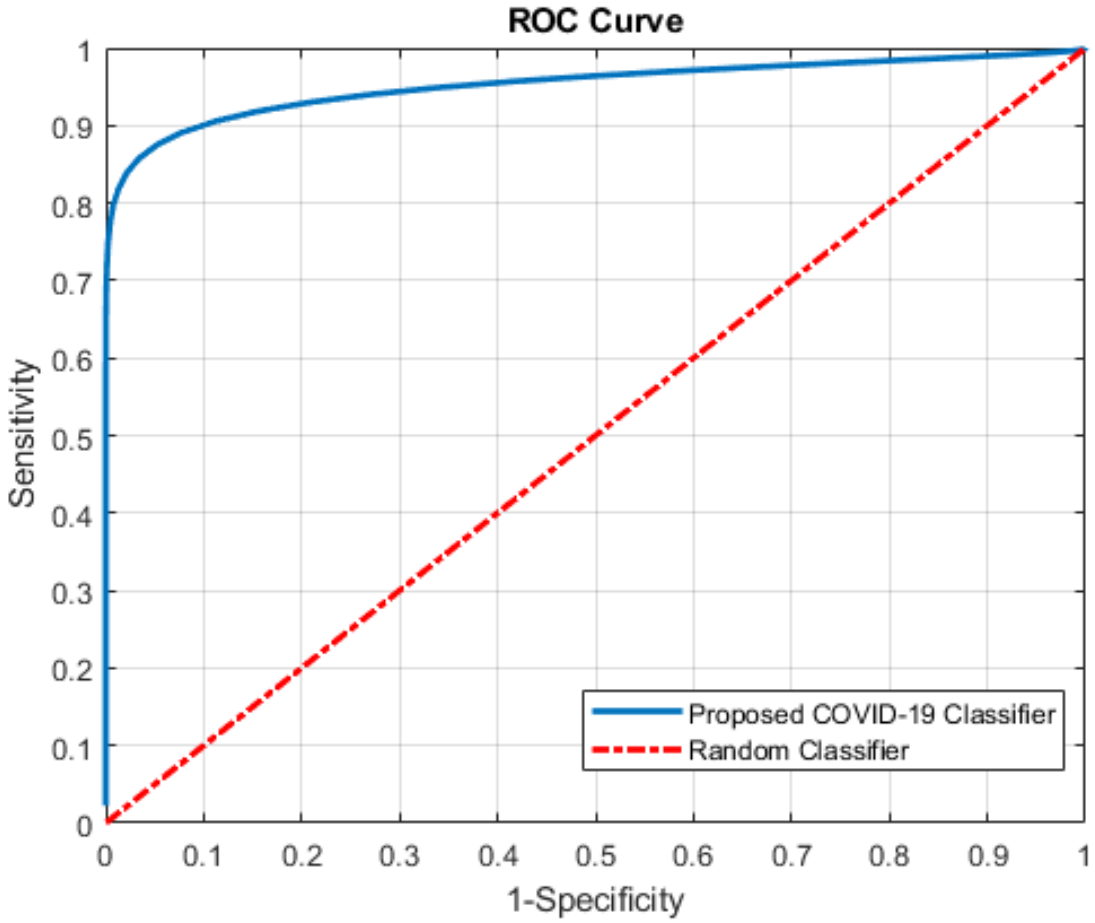

Figure 2 - ROC Curve for COVID-19 Test

(c) Move the decision boundary to generate the ROC curve shown in **Figure 2** for the test and plot it using MATLAB or Python. Is this a good COVID-19 test? Explain your answer by quantifying the Area Under the Curve (AUC) of the ROC and compare it to a random classifier.