CIS 663 Biometrics

Assignment #4

Due: Before Live session 8

For this assignment it is expected you will write some code. You may use **any** language or even just an excel spreadsheet. But you must turn in your code and provide instructions on how to run.

Question 1) Using the iris (plant not eye) data set found at the UCI ML Database, using the just first four columns of data:
1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
   -- Iris Setosa
   -- Iris Versicolour
   -- Iris Virginica

Compute the 3 closest pairs of flowers using:
a) Euclidean distance

$$D(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

b) Manhattan distance

$$D(X,Y) = \sum_{i=1}^{n}|x_i - y_i|$$

c) Cosine similarity

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$$

Recall that
- $\mathbf{A} \cdot \mathbf{B} = A_x \times B_x + A_y \times B_y + \ldots$
- $\|A\| = \sqrt{A_x^2 + A_y^2 + \ldots}$

Question 2) Given the summarized following dataset use Naïve bayes to compute which class the following inputs belong to.

Fruit	Red	Has Leaves	Has Seeds	# Examples
Strawberry	300	250	200	350
Apple	400	100	300	600
Pear	100	50	200	250
Total	800	400	700	1200

There are three types of fruit, strawberries, apples and pears. We've analyzed 1200 total individual pieces of fruit and for each category recorded:
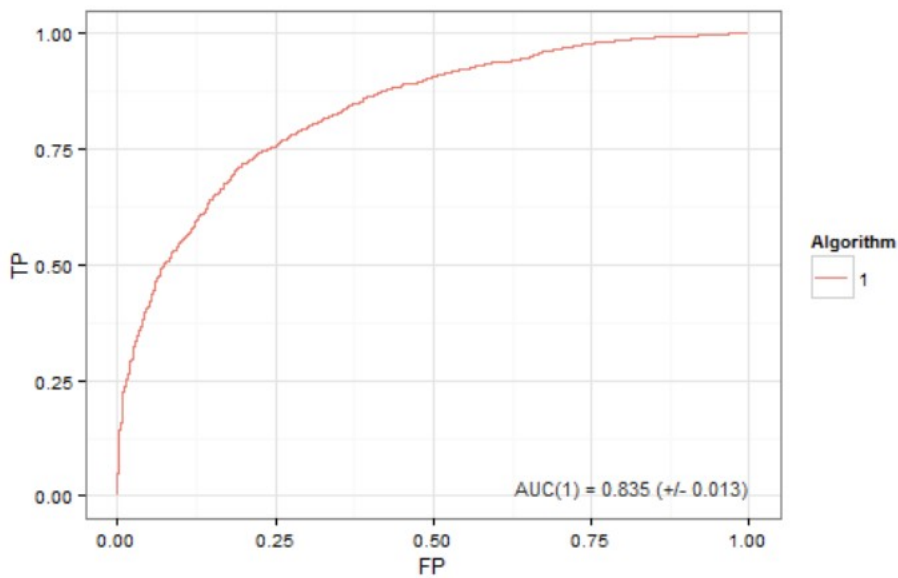1. If it is red
2. If it had leaves
3. If it had seeds

And recorded the counts for each category. The # Examples tells us the actual number of pieces of fruit of each type that were evaluated. For example, of the 350 strawberries examined, 300 were red.

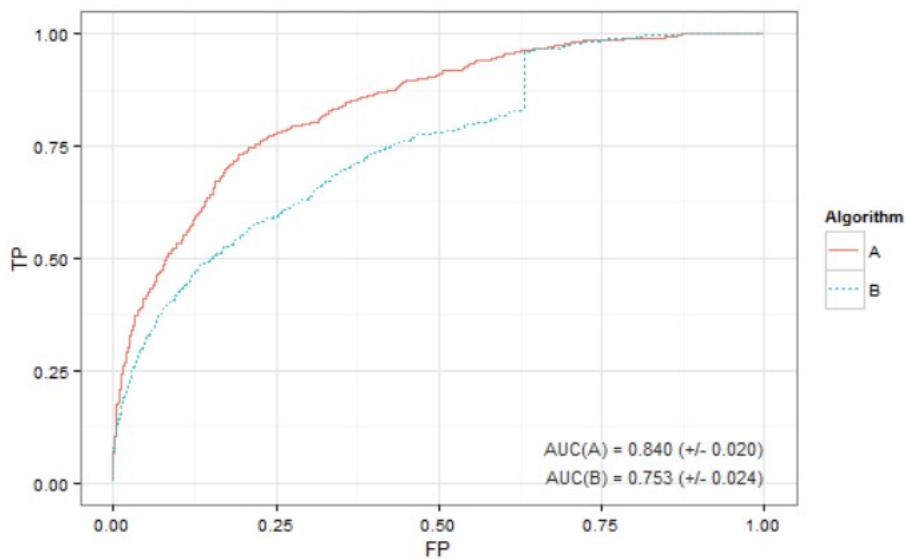Classify the following inputs using Naïve Bayes, <u>show the probability of each class for full credit</u>.

- **Input 1:** Not Red, No Leaves, Has Seeds
- **Input 2:** Red, No Leaves, No Seeds
- **Input 2:** Red, Has Leaves, Has Seeds

Question 3)

a) What you can say about this ROC curve? What does a point on the ROC curve correspond to?

ROC curve for Algorithm 1. Axes: TP (vertical) vs FP (horizontal). AUC(1) = 0.835 (+/- 0.013)

b) Compare two ROC curves. Which one is a better model and why?

ROC curves for Algorithm A and B. Axes: TP (vertical) vs FP (horizontal). AUC(A) = 0.840 (+/- 0.020), AUC(B) = 0.753 (+/- 0.024)

c) Draw or describe what a ROC curve would look like for a random guess.
   What would be varying over the set of random guess algorithms?
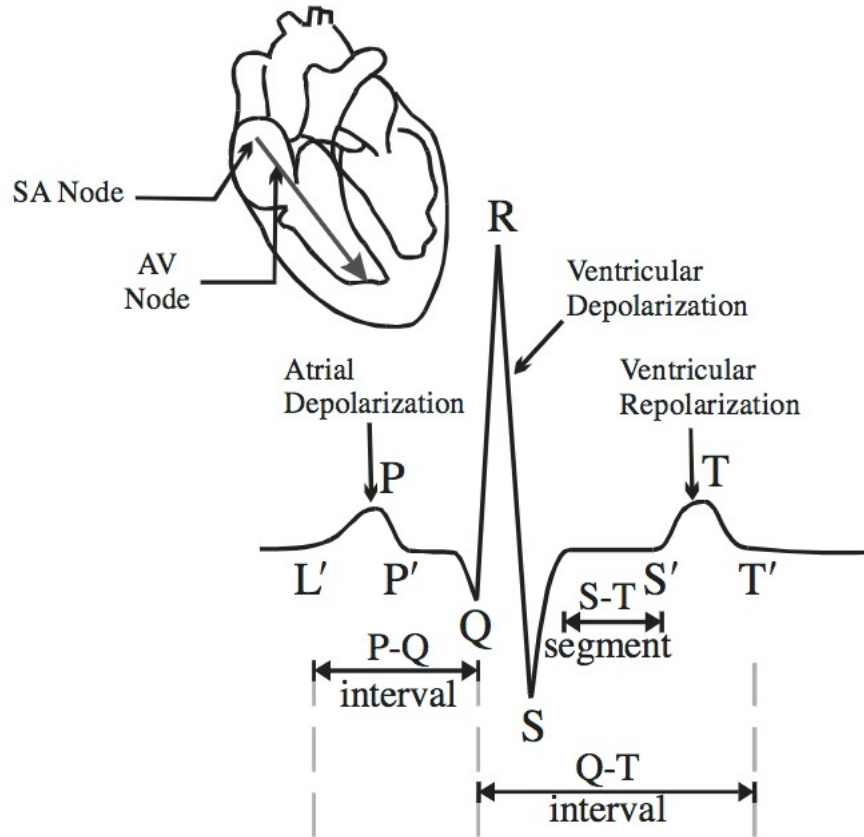   <u>Explain why it looks the way it does.</u>

Question 4) You are developing a finger authentication system for parking garage. The system works as follows.

- User approaches the gate and punch in the ID
- Puts his/her finger on the screen
- Once a fingerprint matches one of the registered users, the system charges the user $10.
- You know that about 50% of the requests come from an imposter hoping to park their car for free.
- When a transaction is made, the system sends the user a receipt via email.
- User has a quick option to say that this was not a valid transaction, then the system needs to refund the money to the user.
- You have found out that if you reject a legitimate user, they immediately leave and find a different parking garage that they will be using regularly, and your business loses $100 of potential revenue.

which of the following algorithms you would like to use and describe how you will set a threshold for the system. **Why?**
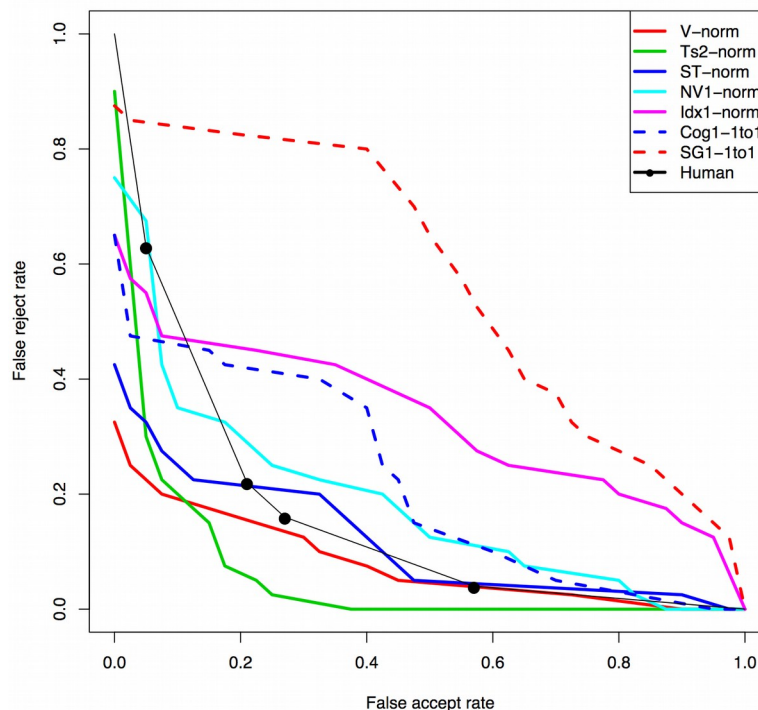
Question 5) You have a ECG data given in the table with all detected fiducial points. Extract features RQ, RP', RP, RL', RS, RS', RT, RT'.

Fiducial points	t
L'	8.6
P	8.9
P'	9.1
Q	9.3
R	9.5
S	9.6
S'	9.9
T	10.4
T'	10.6

Using R = 9.5 as the reference point, each feature is the time difference from R:

- RQ = |9.5 − 9.3| = 0.2
- RP' = |9.5 − 9.1| = 0.4
- RP = |9.5 − 8.9| = 0.6
- RL' = |9.5 − 8.6| = 0.9
- RS = |9.5 − 9.6| = 0.1
- RS' = |9.5 − 9.9| = 0.4
- RT = |9.5 − 10.4| = 0.9
- RT' = |9.5 − 10.6| = 1.1