King Fahd University of Petroleum & Minerals
Information and Computer Science Department

Assignment 3

ICS485 (Machine learning)
Term: 191

For
Dr. Ahmad Irfan

Team 5

Student Name	ID#	Involvement
HATEM ALZAHRANI	201467840	PartC, PartD, 10% of partA (the test_module code and the final two questions)
ABDULLAH ALHABIB	201454740	
HUSSAIN ALMAHDI	201418600	Rest of PartA, PartB

Part A :

Q2. In the function **density_plot, the code for plotting the Gaussian density focuses on the region within 3 standard deviations of the mean. Do you see where this happens? Why do you think we make this choice?**

yes, in verifying x axis, this will give us the best destitution covering the plot, and going to zero

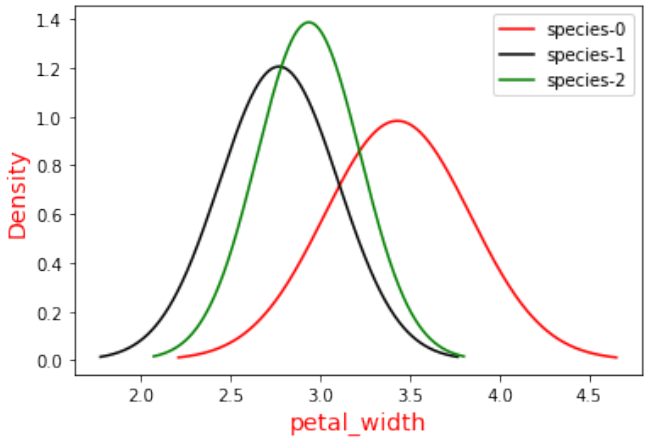
Q3. Here's something for you to figure out: for which feature (0-3) does the distribution of (training set) values for species-2 have the *smallest* standard deviation? what is the value?

with 3 feature we will have the smallest std with mean +3standard deviation equal approximately to 0.28

3

1. For which feature (0-3) do the densities for classes 0 and 2 *overlap* the most?

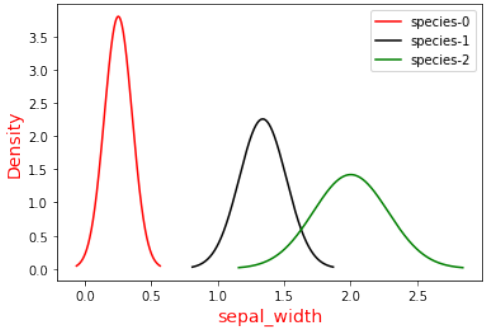
The most overlap will happen with feature 1

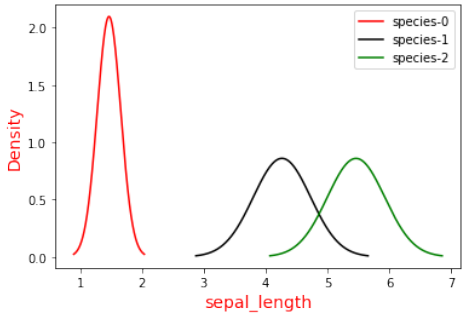


2. For which feature (0-3) is class 2 the most spread out relative to the other two classes?

Feature 2 is the most spread out

3. For which feature (0-3) do the three classes seem the most *separated* (this is somewhat subjective at present)?





In feature 3 and 2

- Which two features have the lowest training error? List them in order (best first).

All the features share the same training error, 1/45

- Which two features have the lowest test error? List them in order (best first).

Fetatures 0, 2, and 3 shared the same test error which is 28/45 while 1 is 33/45

Part B :

Part C :

All the questions in part C were solved in the ,ipyn file of part C

Part D :

Exercise 1: What happens if you do not regularize the covariance matrices?

If I did not regularize the matrix I will not be able to compute it PDF value because the resulting covariance matrices will not be positive semidefinite

Exercise 2: What happens if you set the value of c too high, for instance to one billion? Do you understand why this happens?

I did increase the c value up to 100 billion and nothing really seemed to change, however, the more you increase c the larger your sigma values will end up

Exercise 3: What value of c did you end up using? How many errors did your model make on the validation set?

I ended up putting c equal to 100 billion and the reason for that is when ever I performed the PDF function would get an error that says your sigma matrix is not positive semidefinite (which means I need to regularize my sigmas) so I had to keep increasing until that error went away

Exercise 4: How many errors did your model make on the test set?

I was not able to compute the errors due to an error that yields “singular matrix” whenever the PDF is computed, unfortunately I did not have enough time to figure out the error

Optional:

1. We have talked about using the same regularization constant `c` for all ten classes. What about using a different value of `c` for each class? How would you go about choosing these? Can you get better performance in this way?

By setting different c’s for the classes the actual results will be skewed due to the different values of c and we will end up with not accurate results

2. Try applying multivariate Gaussian classifier on other datasets.