MUSHFIQUE AHMED

ECE 4363 - PATTERN RECOGNITION

PROJECT 1

1. Dataset information
    a. Dataset has 3 classes: Setosa, Virginica and Versicolor
    b. There are 150 samples or feature vectors. 50 of each class
    c. The data represents feature vectors for three species of iris flowers called Setosa, Virginica and Versicolor. The features are sepal length, sepal width, petal length and petal width. The given features and their values for Setosa seem to be comfortably different from the other two species whereas the given attributes for Virginica and Versicolor appear to be similar and could possibly overlap. The performance of our classifier will depend on the specific classification task. It would possibly perform well if it were to separate Setosa from the others.
2.

	Sepal Length	Sepal Width	Petal Length	Petal Width
Minimum	4.3	2	1	0.1
Maximum	7.9	4.4	6.9	2.5
Mean	5.843333	3.057333	3.758000	1.199333
Variance	0.681122	0.188713	3.095503	0.577133
Within class variance	0.265	0.115	0.185	0.0419
Between class variance	0.421	0.0756	2.91	0.536

Observations: Petal length has the highest variance whereas sepal width has the lowest. Petal length also has an unusually high between class variance whereas sepal width has the lowest.

3. **Correlation coefficient map**

Petal length and petal width seem to be quite highly positively correlated to each other. There is also significant positive correlation between sepal length and both petal length and petal width. Sepal width doesn't seem to be strongly correlated with any of the features.

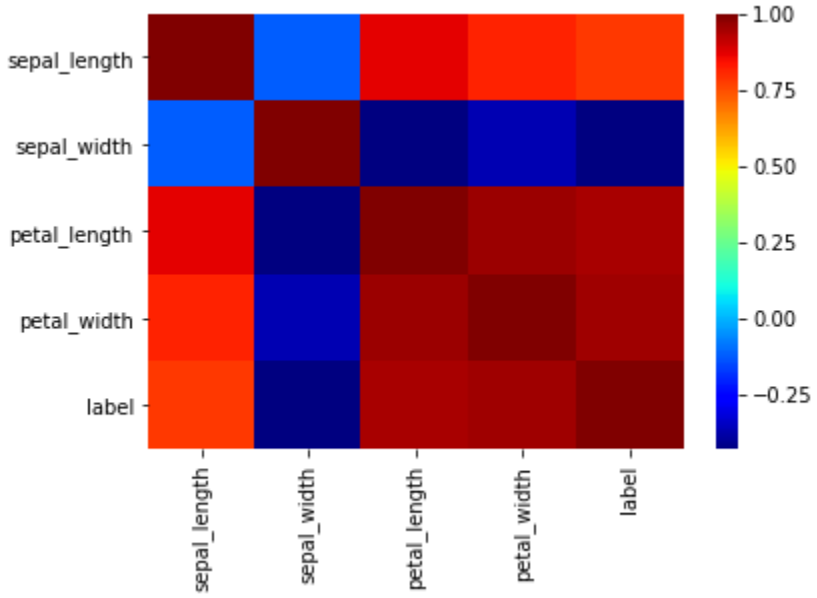

Figure 1: Heatmap of the correlation coefficients

4. **Possible feature performance in classification**

There seems to be quite significant overlap between the sepal length and sepal width features between all three classes suggesting a linear decision boundary is improbable. The petal length and petal width features seem to separate class 1 from the other two classes quite well with almost no overlap. This suggests if our classification task was to separate class 1 from the rest using these 2 features, then a linear decision boundary is very likely. However there is some overlap between class 2 and 3 for these two features.

Since there is significant correlation between sepal length and both petal length and petal width, including sepal length would not be a sensible choice.

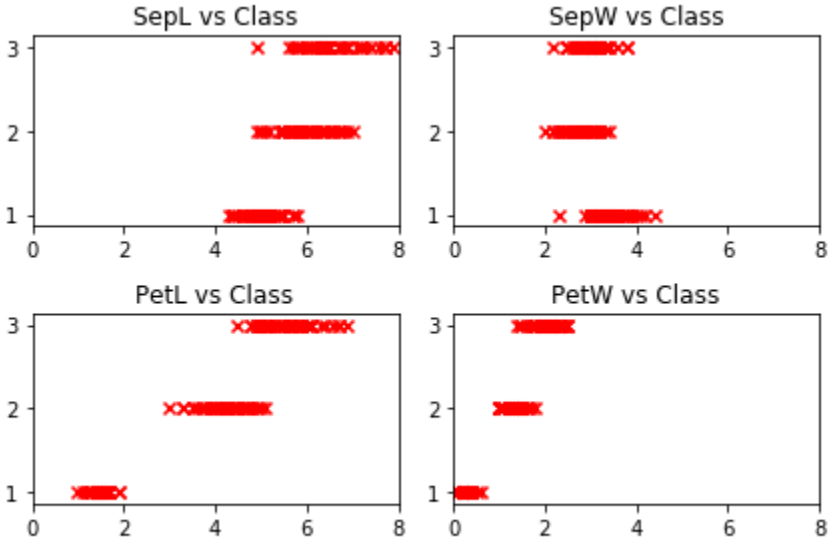

Figure 2: Features vs Class plot

5. Classification Results

	Converged	No.of epochs	Computed weight vector	No. of training misclassifications
Setosa vs Virgi+Versi (All features)	Yes	8	[0.48771415 2.6877357 -5.50931146 -1.21397014 0.26584663] for Batch_Perceptron [[0.06602977], [0.24284787], [-0.22465712], [-0.05747273], [0.11822289]] for LS	0 for LS
Setosa vs Virgi+Versi (Features 3 & 4)	Yes	7	[-0.75428585 -1.1992643 2.14768854] for Batch Perceptron [[-0.25132905] [0.00983426] [1.26603335]] for LS	1 for LS
Virgi vs Setosa+Versi (All features)	No		[[-0.04587608], [0.20276839], [0.00398791], [0.55177932], [-0.69528186]] for LS	11 for LS
Virgi vs Setosa+Versi (Features 3 & 4)	No		[[-0.07300611], [0.64025527], [-0.16018919]] for LS	8 for LS
Setosa vs Virgi vs Versi (Features 3 & 4)			[[-0.25132905, 0.32433516, -0.07300611], [0.00983426, -0.65008953,	34

			0.64025527], [1.26603335, -0.10584416, -0.16018919]]	

6. **<u>Decision Boundary plots</u> (Y-axis is petal_width and X-axis is petal_length)**
   a. Setosa vs Rest (Features 3 & 4 Least Squares)

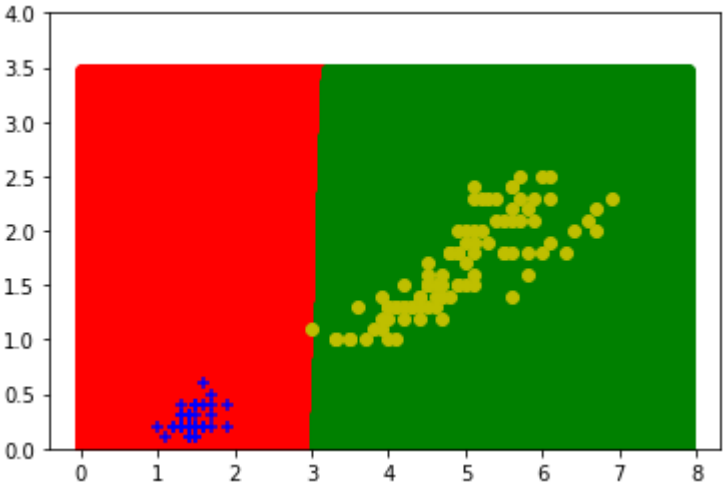

Figure 3: Setosa(blue) vs Rest(yellow)

   b. Setosa vs Rest (Features 3 & 4 Batch Perceptron)

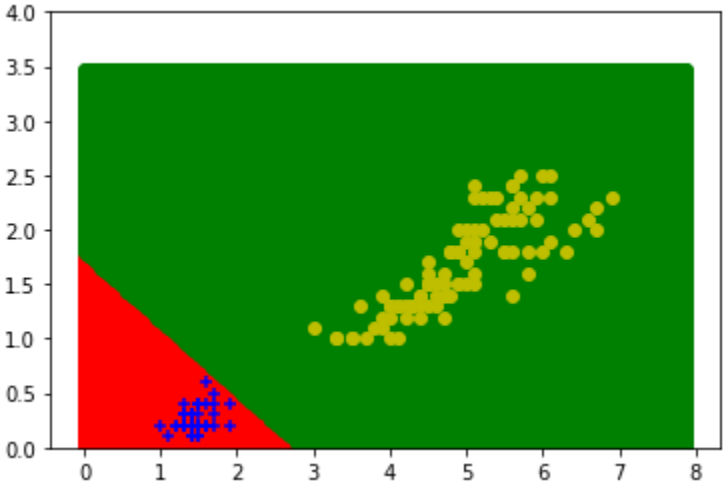

Figure 4: Setosa(blue) vs Rest(yellow)

c. Virginica vs Rest (Features 3 & 4 Least Squares)

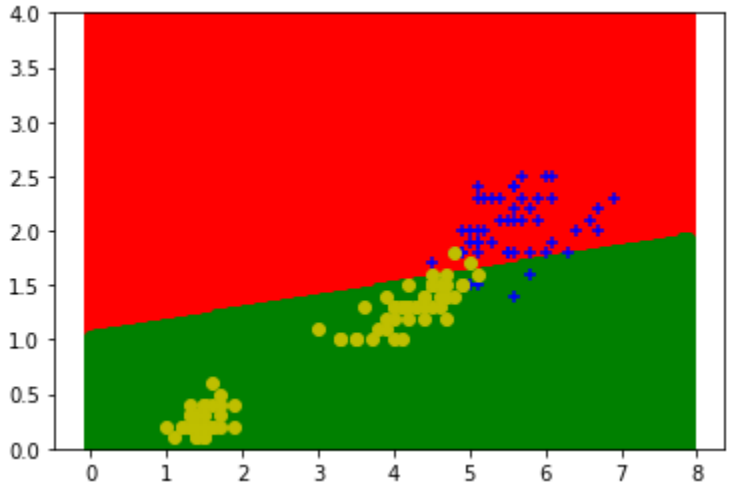

Figure 5: Virginca (blue) vs Rest(yellow)

d. Multiclass (Features 3 & 4 Least Squares)

Green region predicts Virginica, blue for Setosa and red for Versicolor

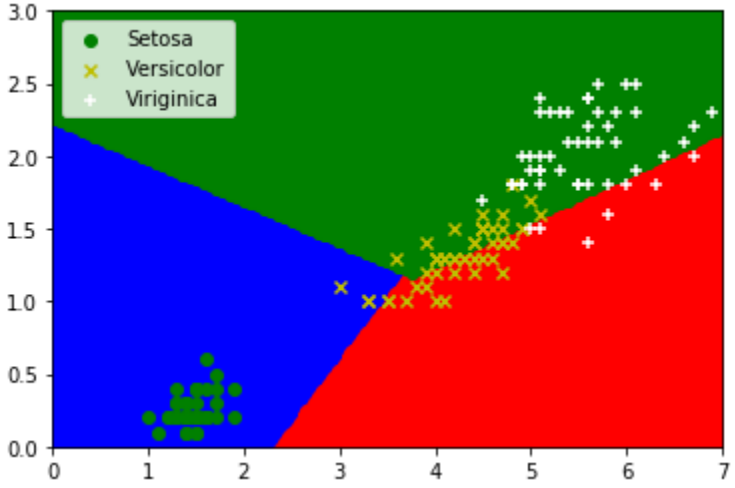

Figure 6: Multiclass classification