Rishikesh

R11643260

February 11, 2024

ECE 5363 – 001: Pattern Recognition

Project I

Dataset: **fisheriris** – The dataset contains a set of 150 records under five attributes: sepal length, sepal width, petal length, petal width and species.

Number of classes: 3 – Setosa, Versicolor, and Virginica

Number of features: 4 – Sepal Length, Sepal Width, Pedal Length, and Pedal Width.

The fisheriris dataset contains measurements for 150 flowers from 3 iris species: Setosa, Versicolor, and Virginica. The 4 features represent sepal length, sepal width, pedal length, and pedal width. Figure 1 and 2 shows scatter plots of sepal length vs sepal width and pedal length vs pedal width for all three species and it can be concluded from those figures that Setosa is linearly separable from the other 2 species based on petal measurements alone. But, Versicolor and Virginica substantially overlap in both sepal and petal features. This means that a linear classifier would struggle to perfectly separate these 2 species. With 3 classes that are not all linearly separable, a single perceptron will fail to converge on a solution because no linear boundary exists that can completely separate the classes.

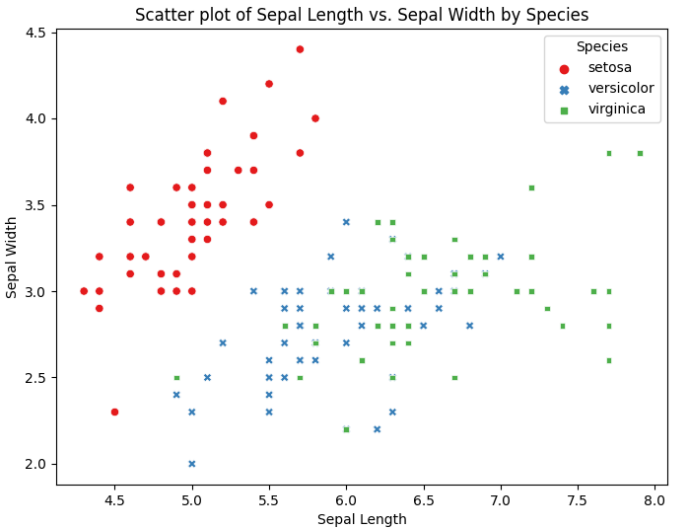

Figure 1: Scatter Plot of Sepal Length vs Sepal Width

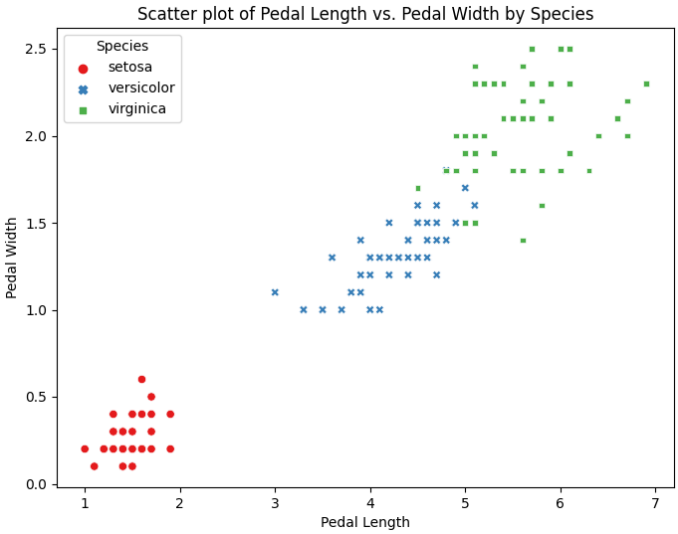

Figure 2: Scatter Plot of Pedal Length vs Pedal Width

Statistical Analysis:

	Sepal Length	**Sepal Width**	**Pedal Length**	**Pedal Width**
Minimum	4.3	2.0	1.0	0.1
Maximum	7.9	4.4	6.9	2.5
Mean	5.84	3.06	3.76	1.20
Variance	0.68	0.19	3.09	0.58
Within-Class Variance	0.265	0.115	0.185	0.042
Between-Class Variance	0.421	0.076	2.914	0.540

Table I: Min, Max, Mean, and Variance quantity for each feature.

The pedal length and pedal width features have much higher variance and between-class variance than the sepal length and sepal width features. The petal measurements vary more significantly between the different iris species, and it will likely be more useful for discriminating between classes. The sepal features have relatively low variance, indicating more consistent measurements across all samples. The sepal dimensions may not provide as much separating power. Overall, the statistics highlight that the petal measurements better capture the differences between iris species while the sepals are more uniform.

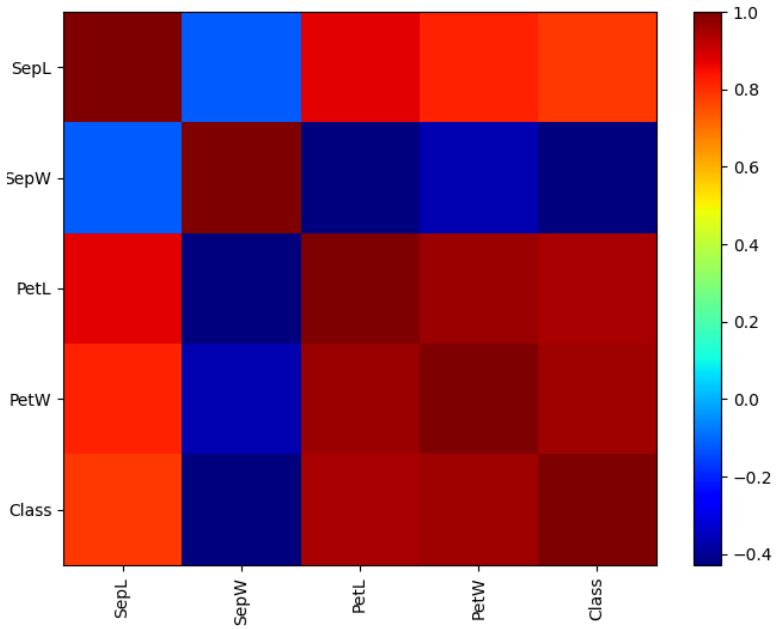

Figure 3: Correlation Coefficients Matrix

The correlation matrix, shown in figure 3, reveals some interesting insights about the relationships between features and iris classes:

- Petal length and width are highly positively correlated meaning larger pedaled flowers also tend to have wider petals. This would make sense biologically.
- Petal length and width are also strongly correlated with the iris class meaning these features clearly differentiate between different iris classes.

- Sepal width has only weak correlation to other features and classes. Sepal size seems less useful for discriminating iris classes.
- Sepal length and width are moderately negatively correlated, implying longer sepals accompany narrower ones. But this relationship is minor.
- So, in summary petal measurements are more predictive of iris type than sepal features. The strong class correlations indicate petal features will enable better linear separation between species.

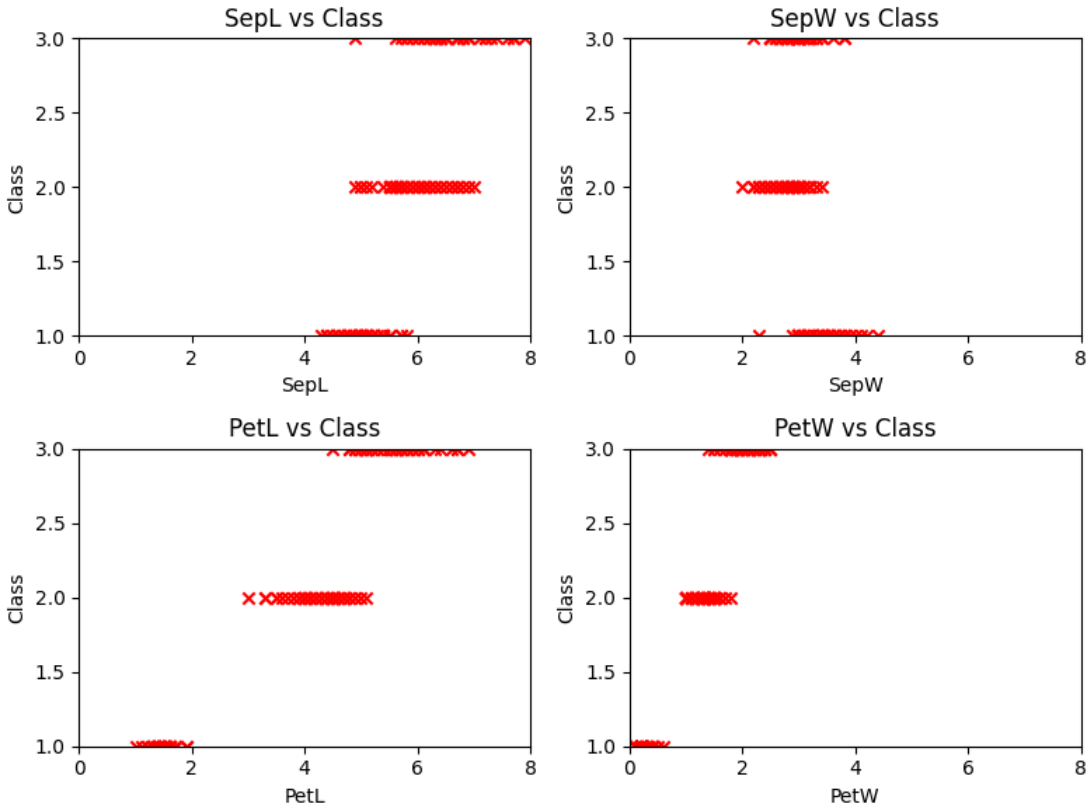

Figure 4: The Four Features vs the Class Label

Figure 4 shows the plots of each feature versus iris class which gives us insight on how useful the features are for classification:

- Sepal length alone cannot reliably separate the classes - data points of different species substantially overlap. (top left)
- Sepal width shows a similar overlap between classes. Using it alone would also make a poor classifier. (top right)
- Petal length on the other hand shows one class (Setosa) clearly distinct from the other two (Versicolor + Virginica). This feature could successfully classify at least one iris species. (bottom left)
- Petal width also shows clear separation of a single class (Setosa) just as the petal length.

So, the petal features show clearer separation between classes. Petal measurements are more suitable for classification than sepal features. Petal length and width show visible grouping of classes, but the sepal plots show a lot of overlapping of data points. A classifier using petal dimensions is likely to be more accurate since petal features better capture differences between iris species.

Classification Task:

Setosa Vs. Versi+Virigi	All Features	Batch Perceptron and LS

The Batch Perceptron converged after 23 epochs.

Batch Perceptron - Weights: [-0.3785703]
[0.38106101]
[0.06662289]
[-0.12696364]
[0.78181556]

Batch Perceptron - Misclassifications: 0

Least Squares - Weights): [0.13205954]
[0.48569574]
[-0.44931423]
[-0.11494546]
[-0.76355422]

Least Squares - Misclassifications: 0

Too many dimensions to Plot.

Classification Task:

Setosa Vs. Versi+Virigi	Features 3 and 4 only	Batch Perceptron and LS

The Batch Perceptron converged after 139 epochs.

Batch Perceptron - Weights: [-0.1785687]
[0.21608379]
[0.2971174]

Batch Perceptron - Misclassifications: 0

Least Squares - Weights: [-0.5026581]

[0.01966852]

[1.53206671]

Least Squares - Misclassifications: 1

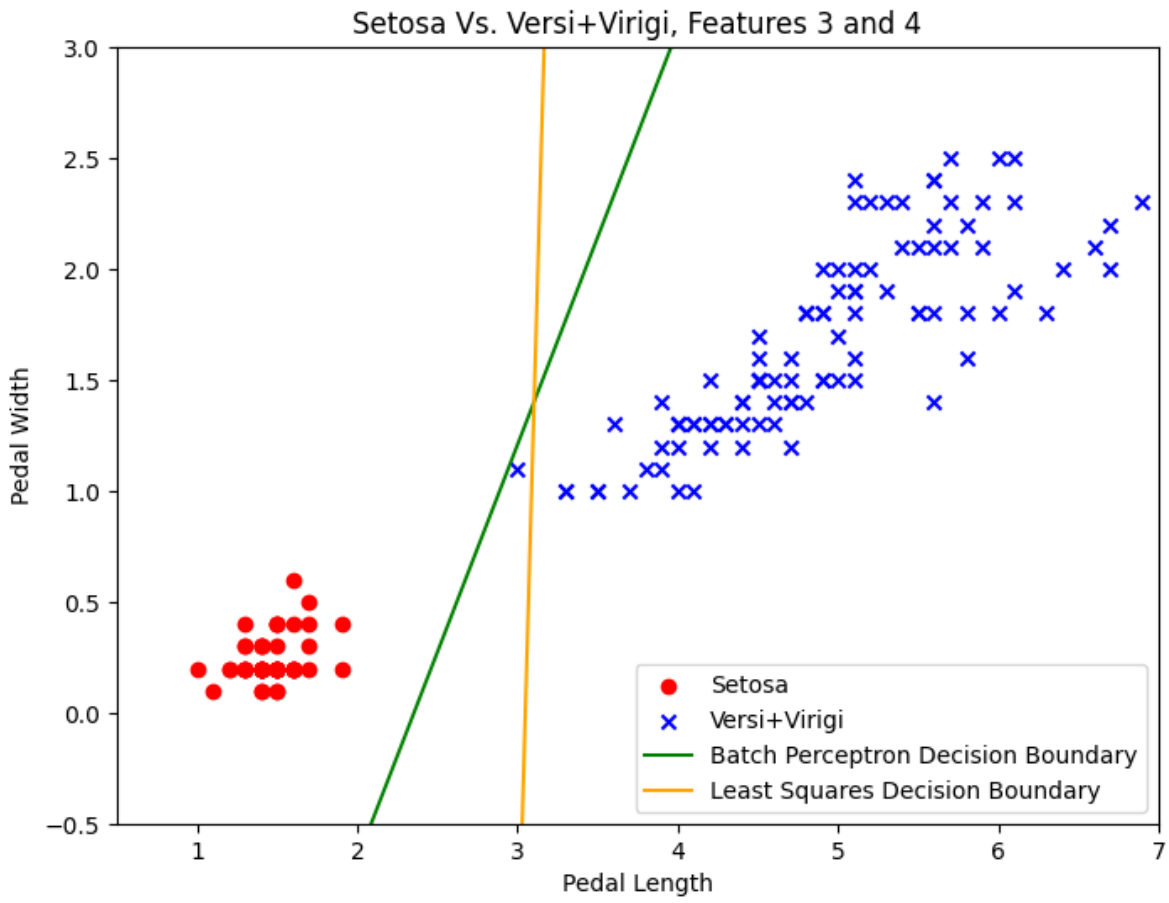

Figure 5. Plot of Feature Vectors and decision boundary

Figure 5 shows how the Perceptron and Least Squares classifiers performed at separating Setosa class from Versicolor + Virginica classes using the petal length and width features.

The Perceptron boundary clearly divides the Setosa points from the combined Versi+Virigi points. It perfectly classifies the Setosa species with no misclassification.

The Least Squares decision boundary mostly separates the classes except one where it misclassifies one Versi+Virigi sample as Setosa. This single Versicolor or Virginica sample falls on the wrong side of the Least Squares decision boundary.

Classification Task:

Virgi Vs. Versi+Setosa	All Features	Batch Perceptron and LS

The Batch Perceptron did not converge after 5000 epochs. (max_epoch = 5000)

Batch Perceptron - Weights: [-0.10114545]
[-0.16547327]
[0.24154305]
[0.41596261]
[-0.7153075]

Batch Perceptron - Misclassifications: 5

Least Squares - Weights: [-0.09175217]
[0.40553677]
[0.00797582]
[1.10355865]
[-2.39056373]

Least Squares - Misclassifications: 11

Too many dimensions to Plot.

Classification Task:

Virgi Vs. Versi+Setosa	Features 3 and 4 Only	Batch Perceptron and LS

The Batch Perceptron did not converge after 5000 epochs. (max_epoch = 5000)

Batch Perceptron - Weights: [0.03559014]
[0.12612772]
[-0.36449704]

Batch Perceptron - Misclassifications: 4

Least Squares - Weights: [-0.14601222]

[1.28051055]

[-1.32037839]]

Least Squares - Misclassifications: 8

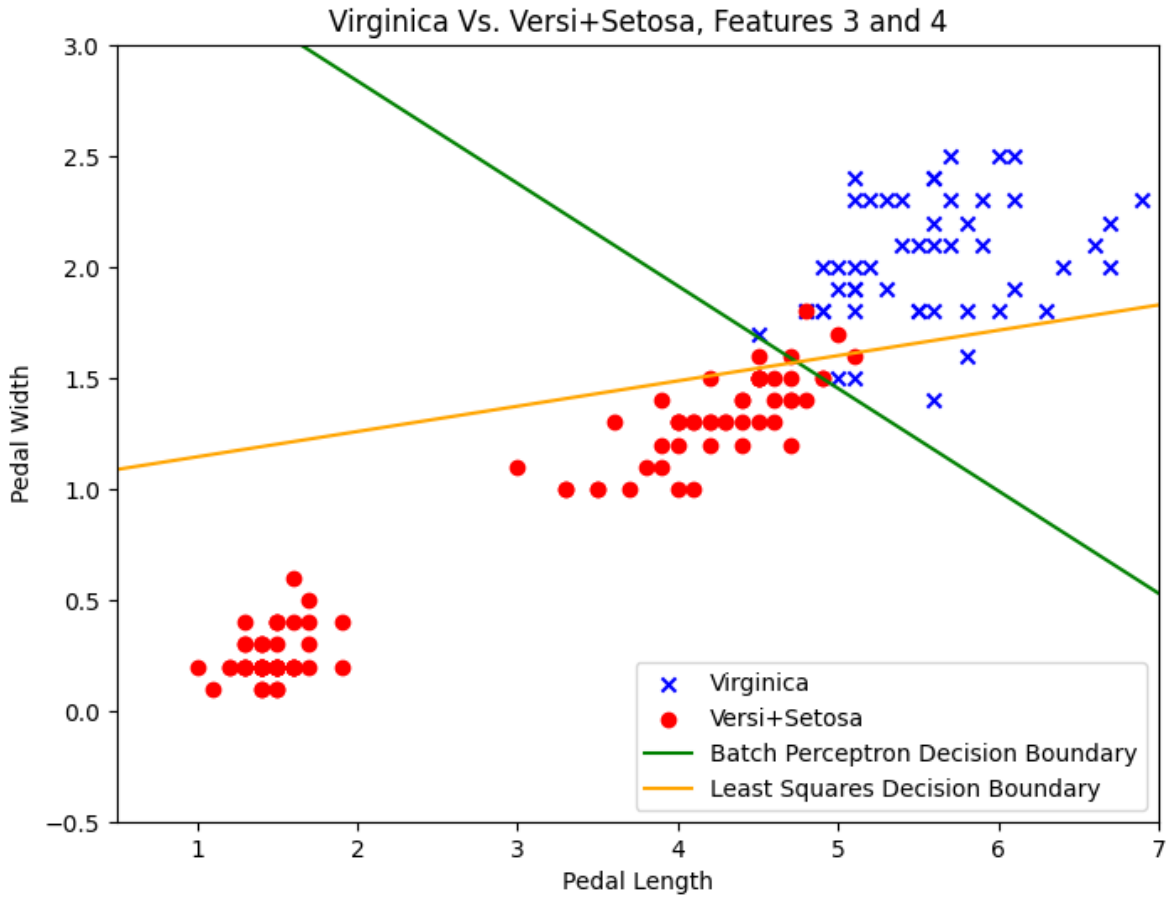

Figure 6. Plot of Feature Vectors and decision boundary

Figure 6 shows how the Perceptron and Least Squares classifiers performed at separating Virginica class from Versicolor + Setosa classes using the petal length and width features.

The Perceptron boundary (green) surprisingly is able to separates the Virginica class with no misclassifications. The Least Squares boundary (orange) also is able to classify Virginica from the Versi+Setosa class with pretty decent accuracy as well.

The obvious overlap between Versicolor and Virginica near the boundaries makes it impossible for an error free linear separation.

Classification Task:

Setosa Vs. Versi Vs. Virigi	Features 3 and 4 Only	Multiclass LS

Least Squares - Weights: [-0.25132905 0.32433516 -0.07300611]

[0.00983426 -0.65008953 0.64025527]

[1.26603335 -0.10584416 -0.16018919]

Misclassifications: 34

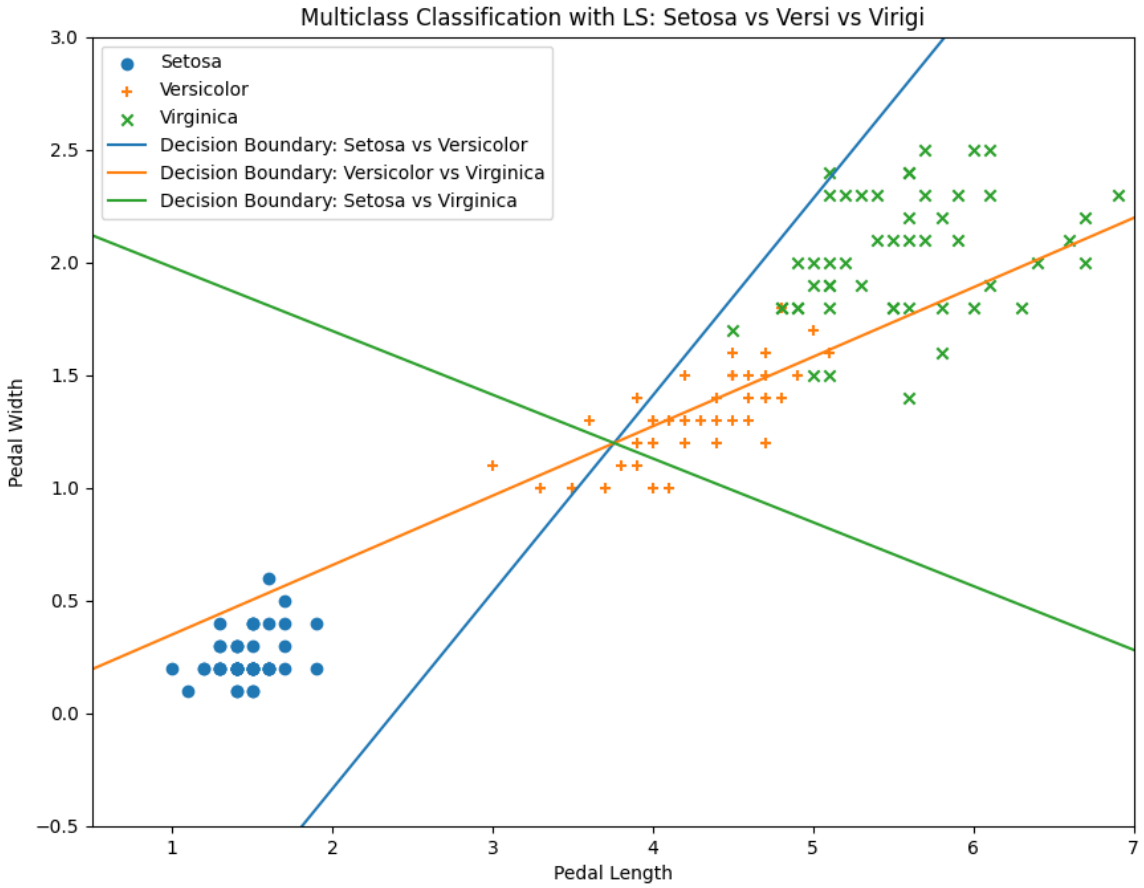

Figure 7. Plot of Feature Vectors and decision boundary

The plot in figure 7 shows the multiclass classification of all three iris classes using Least Squares on petal length and width data. Setosa(blue) is linearly separable from the other classes and the boundary decision line results in no misclassifications, but notable overlap exists between Versicolor (orange) and Virginica (green) which results in the least square boundary decision not being able to completely separate the two classes. The boundary between Versicolor and Virginica misclassifies 34 samples, indicating the Least Squares method cannot reliably distinguish these two species using only the petal features.