

Project 1 – Pattern Recognition (Spring 2021)

Name: Yildirim Kocoglu

- 1) Study the dataset (fisheriris) in terms of (a) Number of classes, (b) Number of features, and (c) What the data represents, i.e., gain some intuition about the problem domain. Based on your study, would you expect the features to perform well in this problem?
 - a. There are 3 classes in the “fisheriris” dataset:
 1. Setosa
 2. Versicolor
 3. Virginica
 - b. There are 4 features in the “fisheriris” dataset for all 3 classes:
 1. Sepal length
 2. Sepal width
 3. Petal length
 4. Petal width
 - c. The “fisheriris” dataset contains 3 classes of flowers (Setosa, Versicolor, Virginica) from the Iris genus with 4 features (Sepal length, Sepal width, Petal length, Petal width) for each example (150 examples). There are 50 examples in each class. The data set was collected by Ronald Fisher and was shared in his 1936 paper: “*The use of multiple measurements in taxonomic problems*” as an example for linear discriminant analysis. The definition of some of the terminology and pictures of 3 types of iris flowers used in the dataset are given below in Figure-1 and Figure-2 to understand the problem better.

One would normally expect these 4 given features to be enough to be able to distinguish the 3 different flower types from each other especially when all 4 features are considered together however, looking at the pictures of the 3 flowers in Figure-2, one can easily see that the sepals and petals of the Versicolor and Virginica classes look very similar to each other and therefore, it will most likely be very difficult to distinguish these 2 classes from one another. Setosa class on the other hand, looks distinguishable to the trained eye from the other classes due to the difference in shape of its petals and sepals. This can also be seen in Figure-3 and Figure-4 plots: “Sepal length vs Sepal width” and “Petal length vs Petal width” respectively. These plots show that Setosa can be linearly separated from the other 2 classes and the Versicolor and Virginica have a lot of points clustered together and are indistinguishable from each other in some cases.

Parts of a Flower

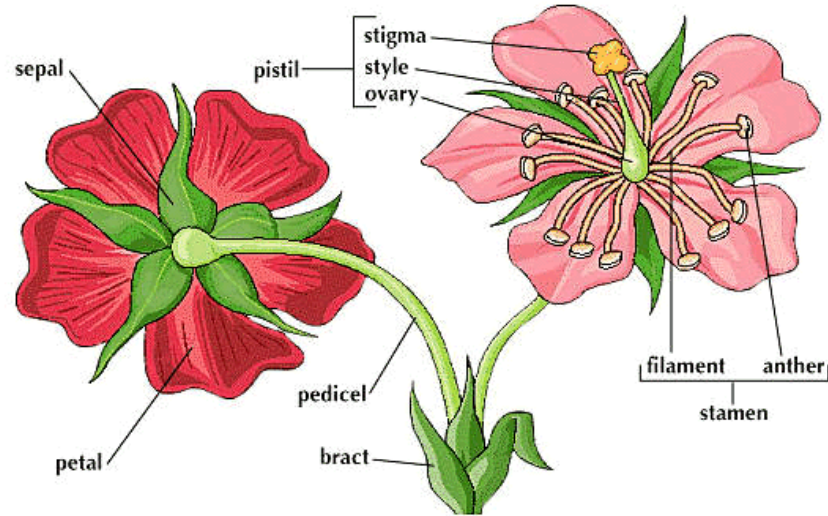


Figure – 1: Parts of a Flower (<https://www.britannica.com/science/sepal>)

Definitions of 3 terminologies used in the “fisheriris” dataset

Iris: Genus of 260-300 species of flowering plants with wide variety of colors.

Sepal: The outer parts of the flower (often green and leaf-like) that enclose a developing bud.

Petal: The parts of a flower that are often conspicuously colored.

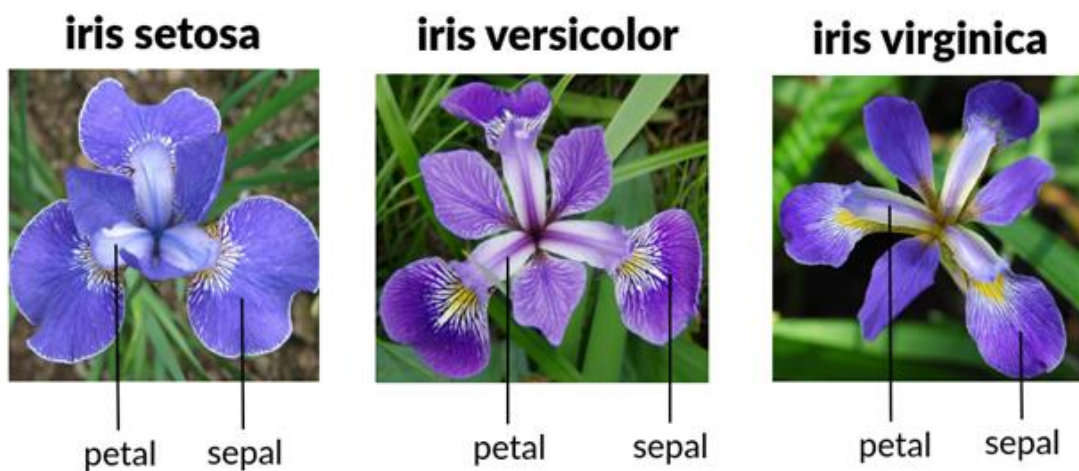


Figure – 2: The sepal and petal parts of the 3 types of Iris flower: Setosa, Versicolor, Virginica that was used in the “fisheriris” dataset (<https://morioh.com/p/eafb28ccf4e3>)

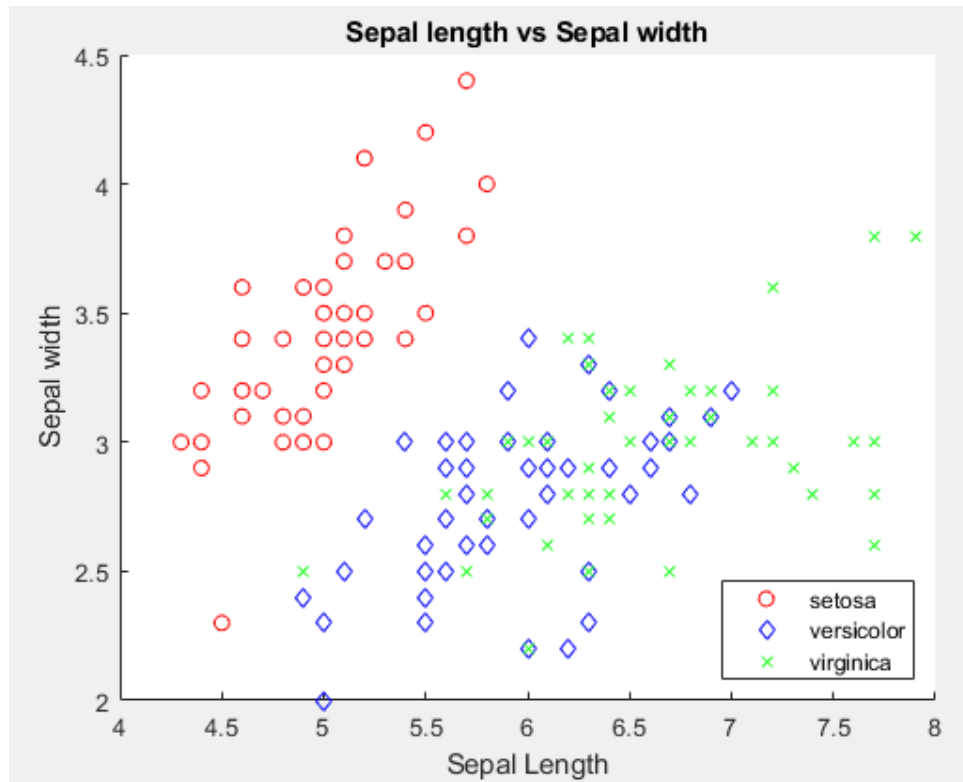


Figure – 3: Sepal length vs Sepal width

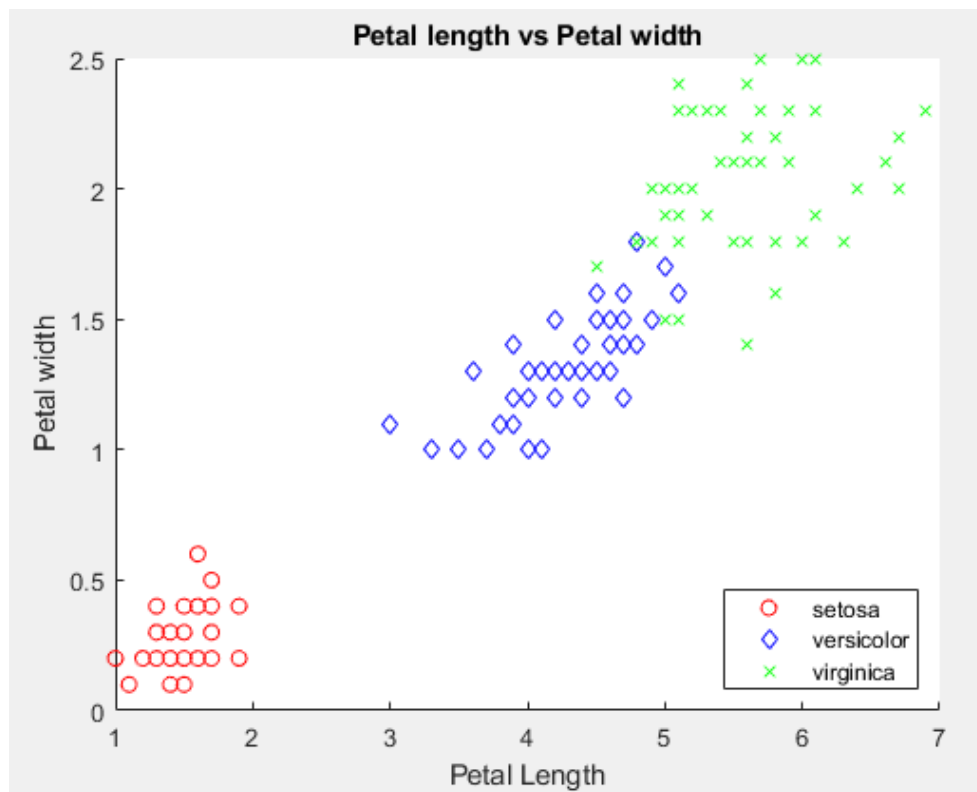


Figure – 4: Petal length vs Petal width

- 2) Compute the following quantities for each feature. Do you observe anything of interest from these statistics?

	Sepal length	Sepal width	Petal length	Petal width
Minimum	4.3000	2.0000	1.0000	0.1000
Maximum	7.9000	4.4000	6.9000	2.5000
Mean	5.8433	3.0573	3.7580	1.1993
Variance	0.6857	0.1900	3.1163	0.5810
Within-Class variance	0.2650	0.1154	0.1852	0.0419
Between-Class variance	0.4214	0.0756	2.9140	0.5361

Minimum and **Maximum**, together may indicate how the examples collected for each feature may differ within all the classes or whether one feature may have similar values to another feature which is not desirable if, both features will be used in the analysis to separate classes. It is not very useful in indicating separability between classes in this case but, it may be useful to indicate the feature values being different from one another. When considered together with **Mean** of the values, it is a stronger measure to indicate similarity of values between features. It can be seen from the table that none of the features are very similar to each other which is desirable if, all the features were to be used together for the analysis of the 3 classes.

Variance can indicate how closely clustered all classes are or in other words, how far away the points are from the **Mean**. In this case, it is preferred not to have a lower value due to the fact that all classes are considered at the same time and if all the classes are very similar in the given feature, that feature may not be a useful feature to separate the classes from each other. **Petal length** has the largest value in this case which indicates it to be a better feature to distinguish between classes.

Within-Class variance ($sw(i) = \sum_{j=1}^M P_j * \sigma_{ji}^2$) can indicate how closely clustered or widespread each class is and if one class has more outliers in a given feature than others. In this case, **Sepal length** has the largest value in between all the features but, it's still believed to be small enough to not cause any significant issues. The lower Within-Class variance is also an indication of less noise in the data or tight clustering of each class within a given feature. It is preferred to have a Within-Class variance value that is as low as possible to recognize classes easier for a given feature.

Looking at the **Between-Class variance** ($sb(i) = \sum_{j=1}^M P_j * (\mu_{ji} - \mu_i)^2$), one can notice the **Petal length** has a much larger value compared to other features, which means that at least one class or more classes have means that are further away from the mean of all the examples (regardless of class) for Petal length. This also may indicate that Petal length is a good candidate as a feature to separate the classes from each other. It is preferred to have a Between-Class variance value as high as possible to separate classes easier for a given feature.

- 3) Compute and display the correlation coefficients exactly as shown below (left figure). Do you observe anything interesting from this display?

Correlation coefficient is closely related to the covariance which is a measure that indicates similarity (linear relationship) between two features. The formula that relates correlation coefficient to covariance is given below which can be thought as the normalized version of covariance:

$$COR(X, Y) = \frac{COV(X, Y)}{\sqrt{VAR(X) * VAR(Y)}}$$

Covariance itself is calculated using the formula below:

$$COV(X, Y) = \frac{1}{n-1} * \sum_{i=1}^n (xi - E(X)) * (yi - E(Y))$$

First thing to notice is that the diagonal values of the correlation coefficient heat map is of the same value = 1. The reason for this is because any feature with itself will have a perfect correlation (a perfect linear line can be fitted through the points without any outlier present). Second thing to notice is that the heat map given in Figure-5 is constructed from using a symmetrical matrix meaning that feature (X,Y) will have the same correlation value with feature (Y,X).

Correlation coefficient is a strong indication of linear relationship between two features X and Y. In this case, PetW, PetL, and Class features have the highest positive correlation between each other (> 0.9). There are also negatively correlated features present but, the value of correlation is not high (around -0.4). Correlation coefficient is also an indication of how closely clustered the points are when two features X and Y are plotted in the same plot. From this heatmap, it can be said that PetL, PetW, and Class features are more tightly clustered compared to other features however, it can't be said that this is a good indication of linear separability but, rather linear dependence between two features.

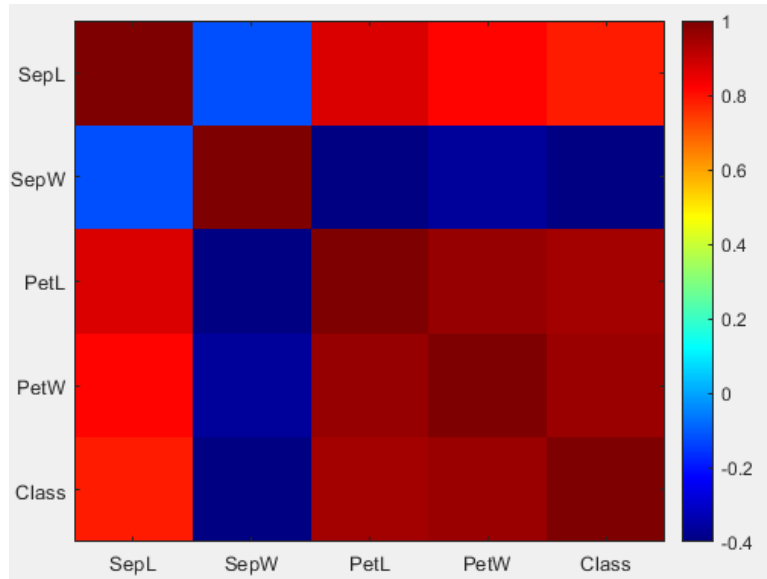


Figure – 5: Correlation coefficients heat map

- 4) Display each of the four features versus the class label, exactly as shown below (right figure). What can you state about how well the features may perform in classification?

From Figure-6, it can be clearly seen that the Setosa class values for PetL and PetW do not overlap with the other two classes (Versicolor and Virginica) and therefore, these two features is most likely the best features to choose to separate Setosa from the rest of the classes. However, in all of the features, Versicolor and Virginica values will overlap which indicates that these two classes are not linearly separable from each other and there will be misclassifications in both classes if linear separation is attempted.

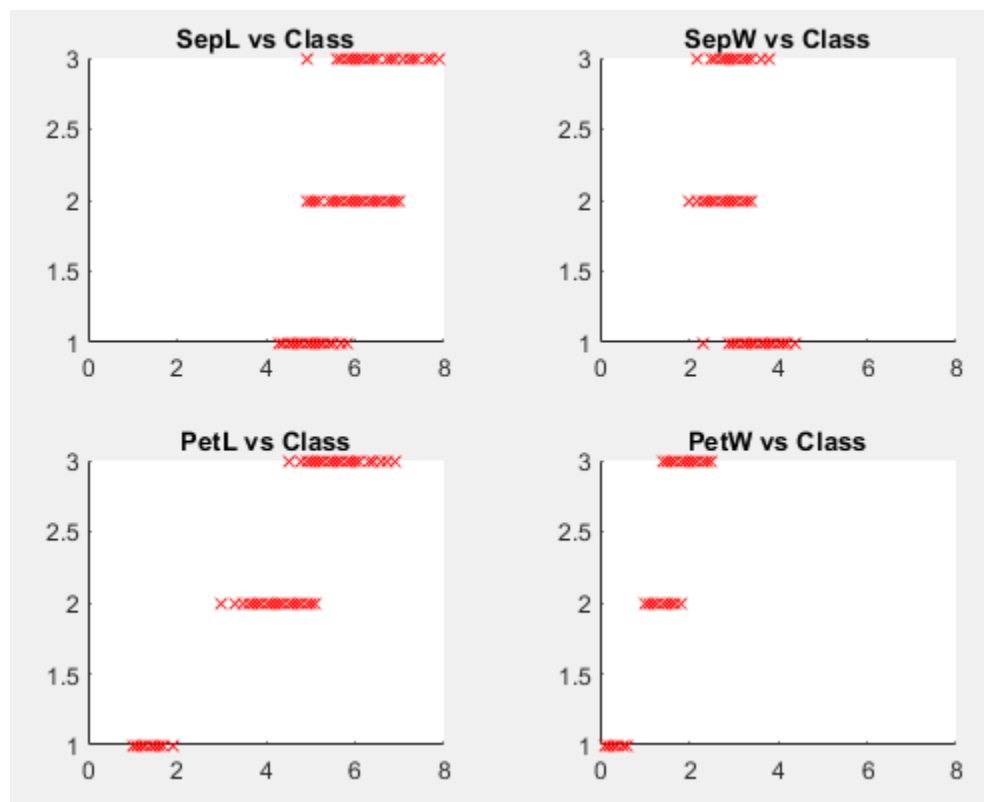


Figure – 6: All four features vs the class labels

- 5) Perform the following classification tasks.

Setosa Vs. Versi+Virigi	All Features	Batch_Perceptron and LS
Setosa Vs. Versi+Virigi	Features 3 and 4 Only	Batch_Perceptron and LS
Virgi Vs. Versi+Setosa	All Features	Batch_Perceptron and LS
Virgi Vs. Versi+Setosa	Features 3 and 4 Only	Batch_Perceptron and LS
Setosa Vs. Versi Vs. Virigi	Features 3 and 4 Only	Multiclass LS

For each case, report whether the method converged. If so, report (a) No. of epochs, (b) Computed weight vector, (c) No. of training misclassifications, and whenever appropriate, (d) plot of feature vectors, as well as the computed decision boundary.

- The 10 tables (Table-1 to Table-10) and the 3 figures (Figure-7 to Figure-9) below are used as a way of reporting the requested answers using both Batch perceptron and Least Squares as well as Multi-class Least Squares for the last classification task. Please note that “BP” stands for Batch perceptron and “LS” stands for Least Squares.
- Labels were given as [1 0] for Setosa vs Versi + Virigi and [1 0] for Virigi vs Versi + Setosa which will impact the decision boundaries for the Least Squares method.
- Batch Perceptron is not affected by the given labels since the labels are not used in the formula for updating the weight vectors.
- Classification was made using $d(x) \geq 0.5$ (class 1) and $d(x) < 0.5$ (class 2) for Least Squares method.
- In Batch Perceptron, if $d(x) < 0$ then that example is considered misclassified and correctly classified if $d(x) \geq 0$ due to the multiplication of class 2 with -1. During testing phase, since the unlabeled class is not multiplied with -1, if $d(x) \geq 0$ then it belongs to class 1 and if $d(x) < 0$, then it belongs to class 2.
- In Multiclass Least Squares, labels are given as [1 0 0] for class 1, [0 1 0] for class 2, and [0 0 1] for class 3. The classification is made based on the column value of maximum value. For example: if prediction equals [0.95 0.03 0.02] then the maximum value is 0.95 therefore, the column value is 1 and the class = 1.

Setosa Vs. Versi + Virigi (All Features)

	Setosa vs. Versi + Virigi (All features)
Converged	Yes
No. of Epochs	7
No. of misclassifications (LS)	0
No. of misclassifications (BP)	0

Table – 1: Setosa vs. Versi + Virigi (All features) results

	Setosa vs. Versi + Virigi (All features)				
w_LS	0.066	0.242848	-0.22466	-0.05747	0.118223
w_BP	7.6634	25.77837	-41.2355	-17.9552	4.904719

Table – 2: Setosa vs. Versi + Virigi (All features) computed weight vectors

Setosa Vs. Versi + Virigi (Features 3 & 4)

	Setosa vs. Versi + Virigi (Feature 3 & 4)
Converged	Yes
No. of Epochs	8
No. of misclassifications (LS)	1
No. of misclassifications (BP)	0

Table – 3: Setosa vs. Versi + Virigi (Features 3 & 4) results

	Setosa vs. Versi + Virigi (Feature 3 & 4)		
w_LS	-0.251	0.009834	1.266033
w_BP	-5.288	-8.73925	20.62585

Table – 4: Setosa vs. Versi + Virigi (Features 3 & 4) computed weight vectors

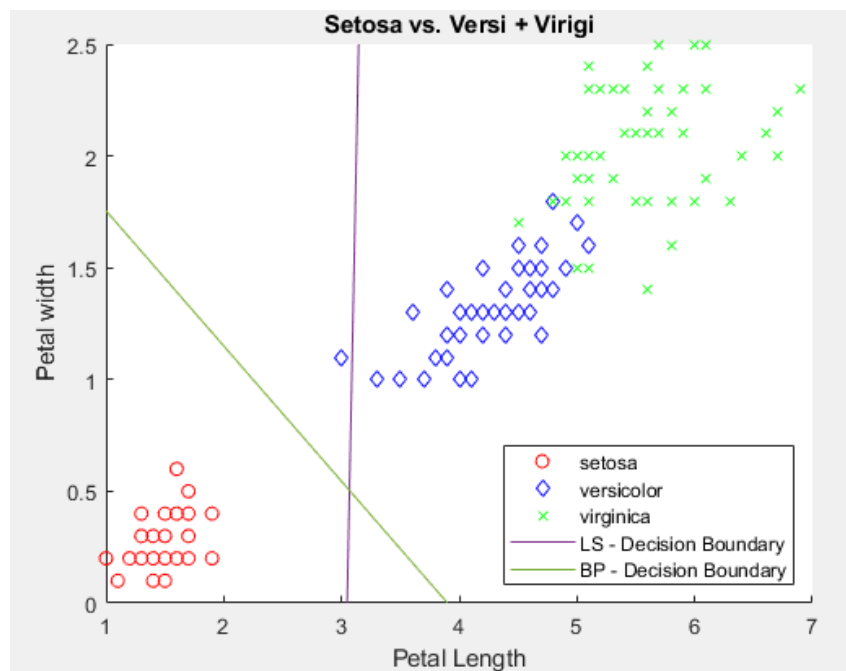


Figure – 7: Setosa vs Versi + Virigi with least squares (LS) and batch perceptron (BP) decision boundaries given (using only Petal width and Petal length features).

Virigi Vs. Versi + Setosa (All Features)

	Virigi vs. Versi + Setosa (All features)
Converged	No
No. of Epochs	Max Epochs = 1000
No. of misclassifications (LS)	11
No. of misclassifications (BP)	6

Table – 5: Virigi vs. Versi + Setosa (All Features) results

	Virigi vs. Versi + Setosa (All features)				
w_LS	0.0459	-0.20277	-0.00399	-0.55178	1.695282
w_BP	-82.04	-89.3716	141.4545	115.5848	-105.415

Table – 6: Virigi vs. Versi + Setosa (All Features) computed weight vectors

Virigi Vs. Versi + Setosa (Features 3 & 4)

	Virigi vs. Versi + Setosa (Feature 3 & 4)
Converged	No
No. of Epochs	Max Epochs = 1000
No. of misclassifications (LS)	8
No. of misclassifications (BP)	6

Table – 7: Virigi vs. Versi + Setosa (Features 3 & 4) results

	Virigi vs. Versi + Setosa (All features)		
w_LS	0.073	-0.64026	1.160189
w_BP	18.802	56.51075	-182.674

Table – 8: Virigi vs. Versi + Setosa (Features 3 & 4) computed weight vectors

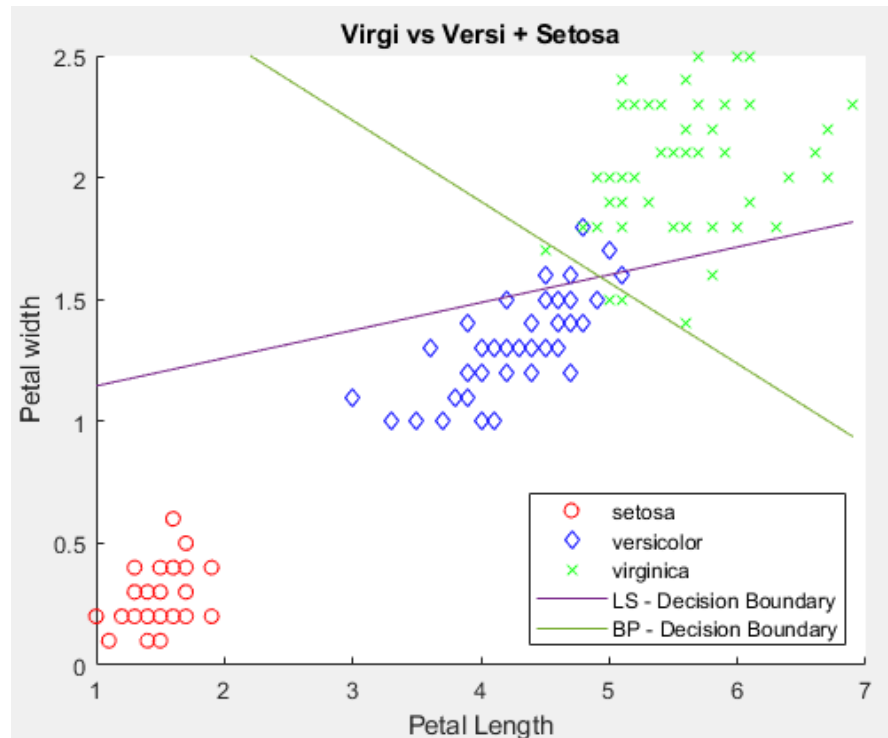


Figure – 8: Virigi vs Versi + Setosa with least squares (LS) and batch perceptron (BP) decision boundaries given (using only Petal width and Petal length features).

Setosa Vs. Versi Vs. Virigi (Features 3 & 4)

	Setosa vs. Versi vs Virigi (Feature 3 & 4)
No. of misclassifications (Multi-LS)	34

Table – 9: Setosa vs. Versi vs. Virigi (Features 3 & 4) results

	w_multi_LS (Feature 3 & 4)		
w1	-0.251	0.009834	1.266033
w2	0.3243	-0.65009	-0.10584
w3	-0.073	0.640255	-0.16019

Table – 10: Setosa vs. Versi vs. Virigi (Features 3 & 4) computed weight vectors

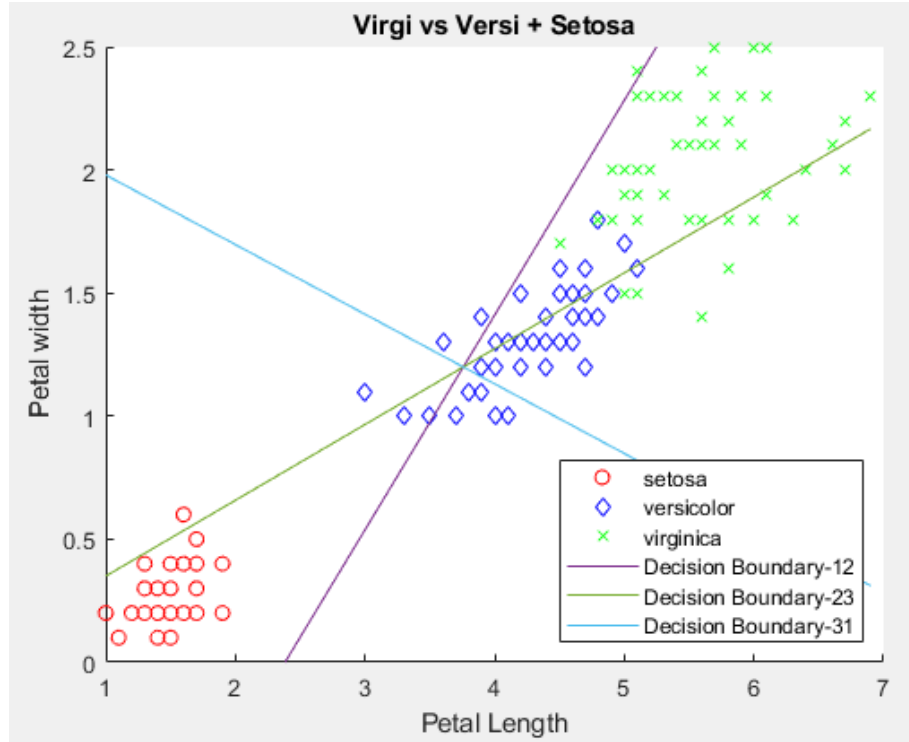


Figure – 9: Virgi vs Versi vs Setosa with multi-class least squares (multi-class LS) decision boundaries given with Decision Boundary -12 indicating the decision boundary between class 1 and 2 (using only Petal width and Petal length features).

The decision boundaries in Figure-9 are drawn using the formula below:

- $d1(x) - d2(x) > 0$
- $d2(x) - d3(x) > 0$
- $d3(x) - d1(x) > 0$

where, $d1(x) = w1 * x$, $d2(x) = w2 * x$, and $d3(x) = w3 * x$ and $w1, w2$, and $w3$ are all taken from the $M * (l + 1)$ W matrix calculated using the Multi-class Least Squares with $M = 1,2,3$ (3 classes) in this case.