max 100 pts.
starts with 10 pts.

[redacted] 1/12

Project 1

Dataset

The dataset includes 150 observations of Iris flowers. These observations have been labeled in 3 classes as Setosa, Vericolor, and Virginica (50 observations in each class). For each observation, 4 features have been measured as sepal length, sepal width, petal length, and petal width (Figure 1). The objective is to use these observations for training a classifier.

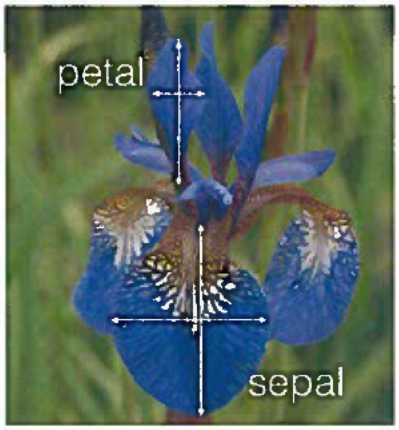


Figure 1. Iris flower and its features[1]

Since each observation has 4 features, their visualization is not possible (4D feature space). However, an intuition can be gained by pairwise illustration of features. Sepal length versus sepal width, and petal length versus and petal width of the observations are depicted in Figures 2 and 3, respectively.

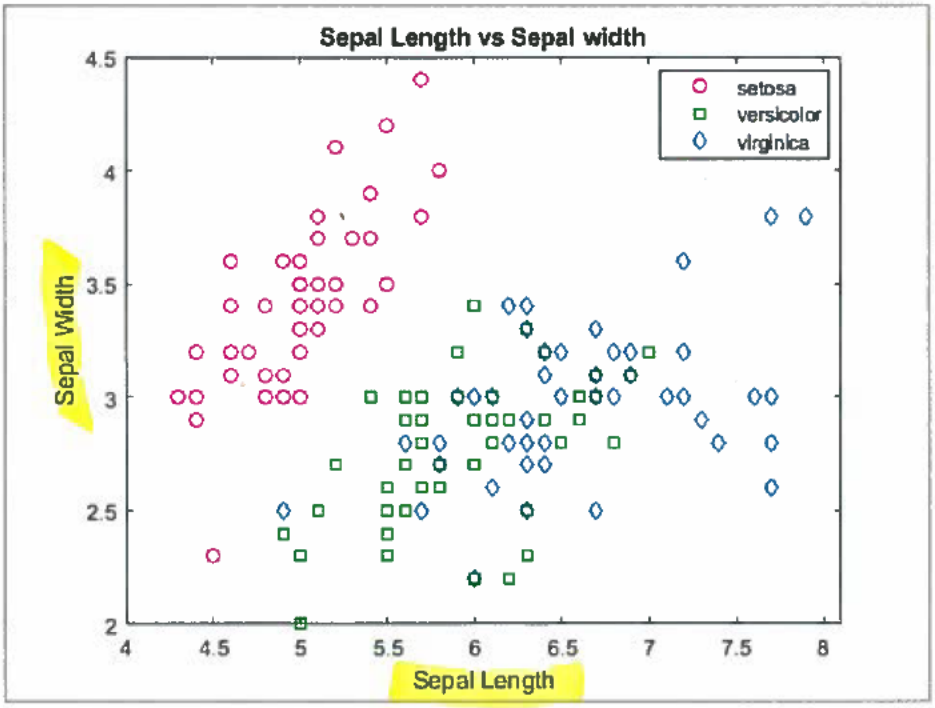


Figure 2. Sepal length versus sepal width

[1] http://5047-presscdn.pagely.netdna-cdn.com/wp-content/uploads/2015/04/iris_petal_sepal.png

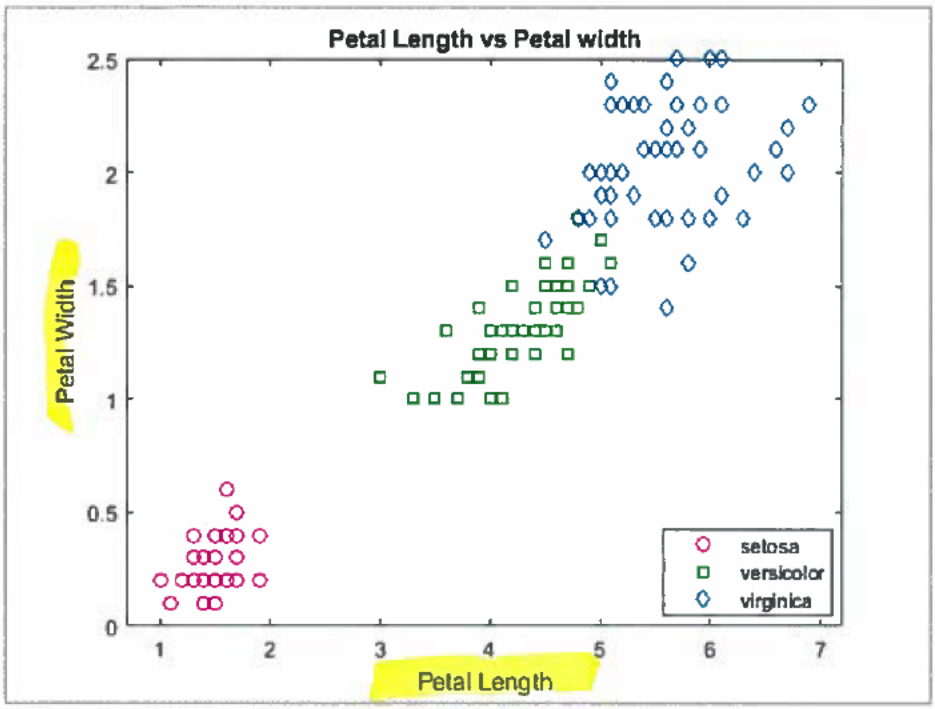


Figure 3. Petal length versus petal width

Figure 2 shows that if only sepal length and sepal width are employed, Setosa is expected to be linearly separated from the other two classes. However, these two features alone are not able to linearly separate Versicolor from Virginical.

Figure 3 shows that application of petal length and petal width will have a better result than the previous set. This case is expected to linearly separate Setosa from the other two classes and *almost* separate Versicolor and Virginical linearly.

The potential of simultaneous application of all the four features for separating the classes cannot be studied *visually*.

10 pts

- Basic info about dataset
- 2D plots using pairs of features, especially those shown in Fig. 2 & 3
- Can be concluded that features measured from petals may be more effective than those measured from sepals.

Statistics

Statistic	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

Petal length and *petal width* have interesting statistics; they both have relatively low within-class variance and high between-class variance. If it is intended that 2 of 4 features to be applied for classification, this characteristic makes these two features favorable for separating the classes. This is resulted from the property of the mentioned statistics; a small within-class variance of a feature means that there is a small difference between the values of this feature for different members of the same class, while a large between-class variance states that there is a meaningful difference in the feature values for members of different classes.

- Check all values in table — 10 pts
- Similar Comments as those highlighted

Correlation Coefficients

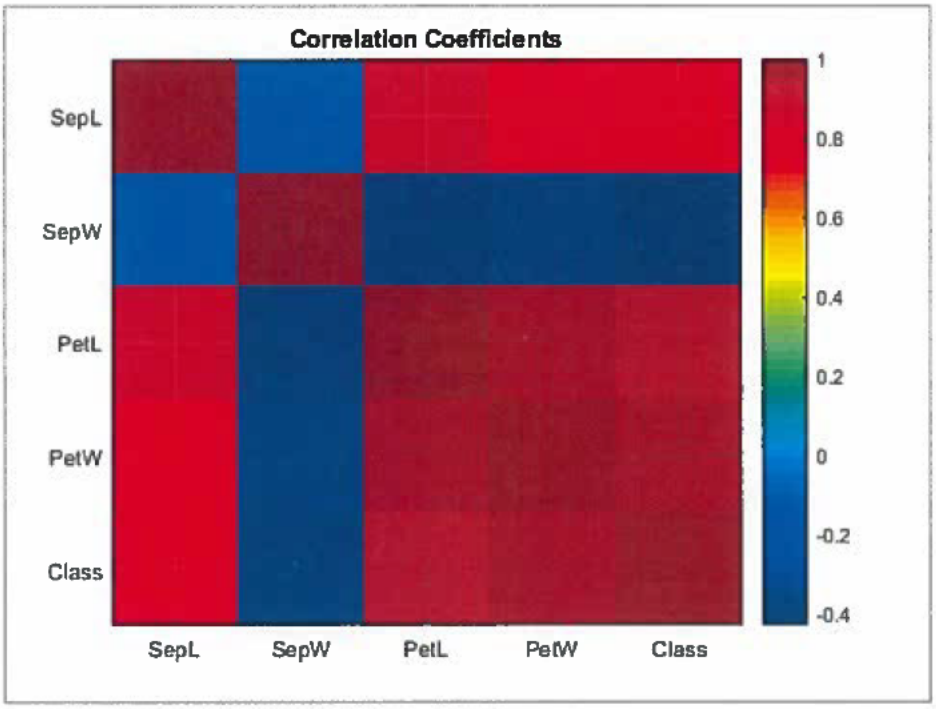


Figure 4. Correlation coefficients of features and classes

Figure 4 shows the correlation coefficients of features and class labels. Sepal width shows an interesting pattern: it has negative correlation with the other three features and the class label. However, the absolute values of the correlation coefficients are not big enough to draw a comprehensive conclusion.

Further, petal width has a high correlation with petal width and class label. It means that the flowers with bigger petal width have bigger length and higher odds of belonging to classes with bigger labels (e.g. Virginica rather than Setosa). There is a similar pattern in the correlation between petal length and class label but with lower odds (due to the smaller correlation coefficient).

Due to the high correlation of petal width and petal width, it may be decided to use only one of them for the classification.

- Check the plot for Correctness 5 pts
- high Correlation between petal width & Class label
- Low " " Sepal " " " "
  absolute

Features vs. Class Labels

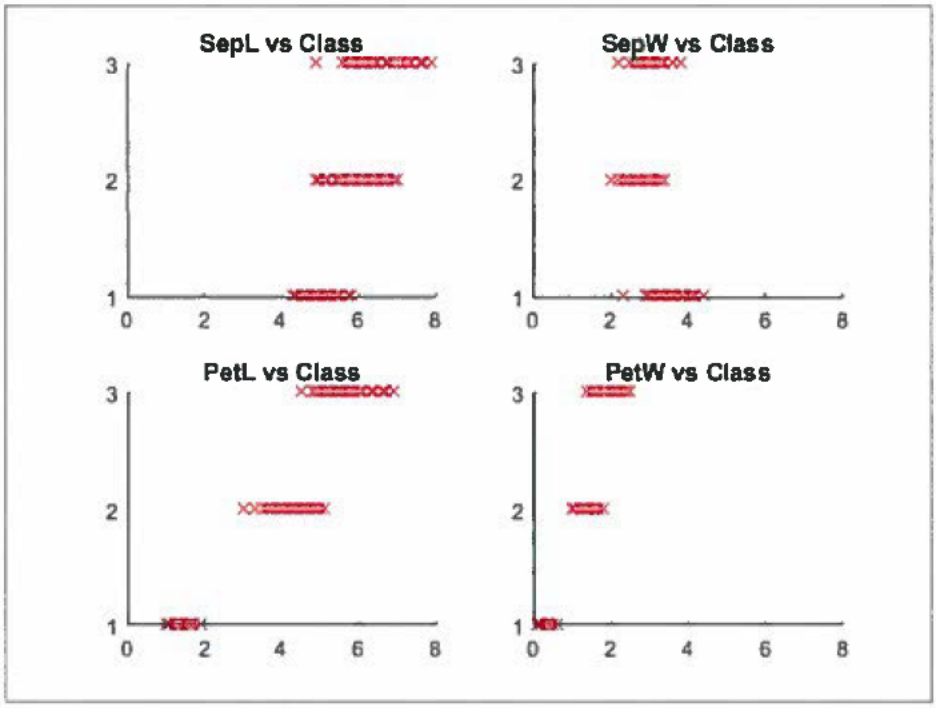


Figure 5. Features versus class labels

Figure 5 depicts the values of each feature versus the class label. The points showing the values of the petal width are more congested in each class; the property that was concluded from the within-class variance two. Also, assuming a certain value for a feature (a vertical line in each graph), petal length and petal width have less points in multiple classes in comparison to sepal length and sepal width. Hence, if it is aimed to select only two out of the four features (to measure only sepal or petal dimensions), the petal length and width will probably perform better than the sepal length and width.

5 pts

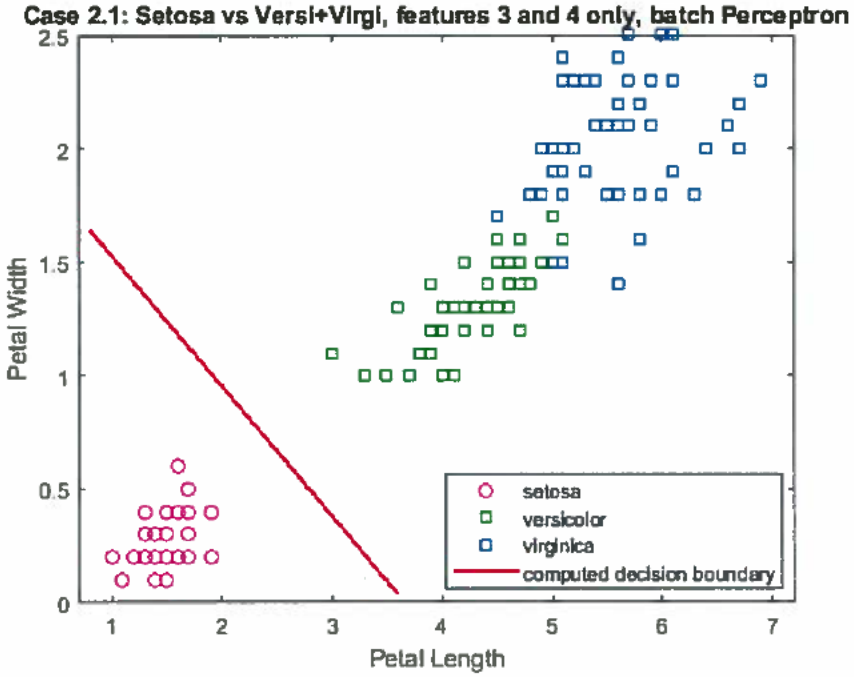
- check plots for correctness
- Again, petal length & petal width exhibit small within-class variances & large between-class variances

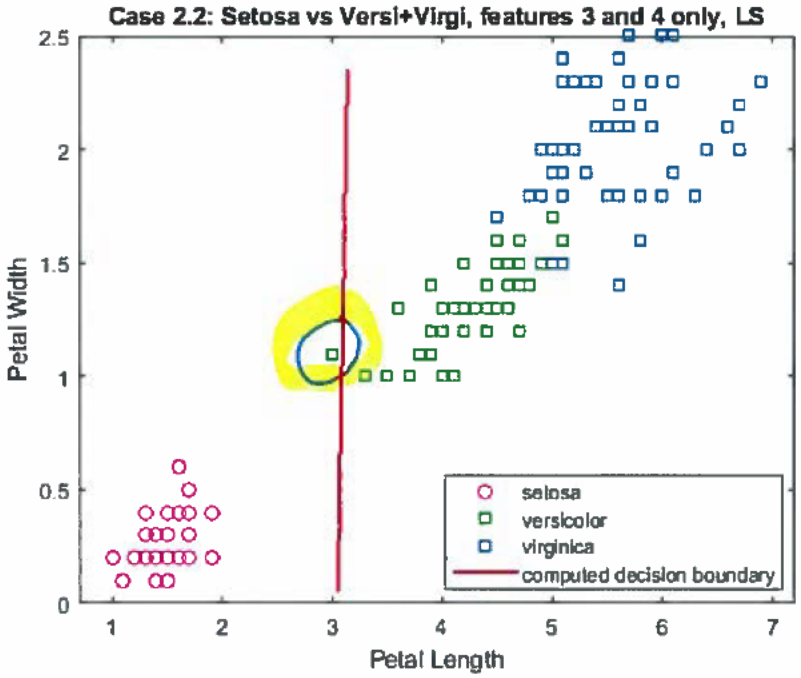
Classification

Case 1.1. Setosa vs. Versicolor+Viriginica; all features; batch Perceptron	
a. Did the method converge?	Yes
b. No. of epochs	7
c. Computed weight vector	$\begin{bmatrix} 5.5050 \\ 13.6700 \\ -19.1850 \\ -8.8000 \\ 2.7500 \end{bmatrix}$
d. No. of training misclassifications	0
e. Plot of feature vectors and computed decision boundary	N/A

10 pts.

Case 1.2. Setosa vs. Versicolor+Viriginica; all features; LS	
a. Did the method converge?	Yes (LS always gives an answer)
b. No. of epochs	1 (LS is a one-step algorithm)
c. Computed weight vector	$\begin{bmatrix} 0.0660 \\ 0.2428 \\ -0.2247 \\ -0.0575 \\ 0.1182 \end{bmatrix}$
d. No. of training misclassifications	0
e. Plot of feature vectors and computed decision boundary	N/A

Case 2.1. Setosa vs. Versicolor+Viriginica; features 3 and 4 only; batch Perceptron	
a. Did the method converge?	Yes
b. No. of epochs	7
c. Computed weight vector	$\begin{bmatrix} -2.7050 \\ -4.7050 \\ 9.9000 \end{bmatrix}$
d. No. of training misclassifications	0
e. Plot of feature vectors and computed decision boundary exact location of boundary is not important	

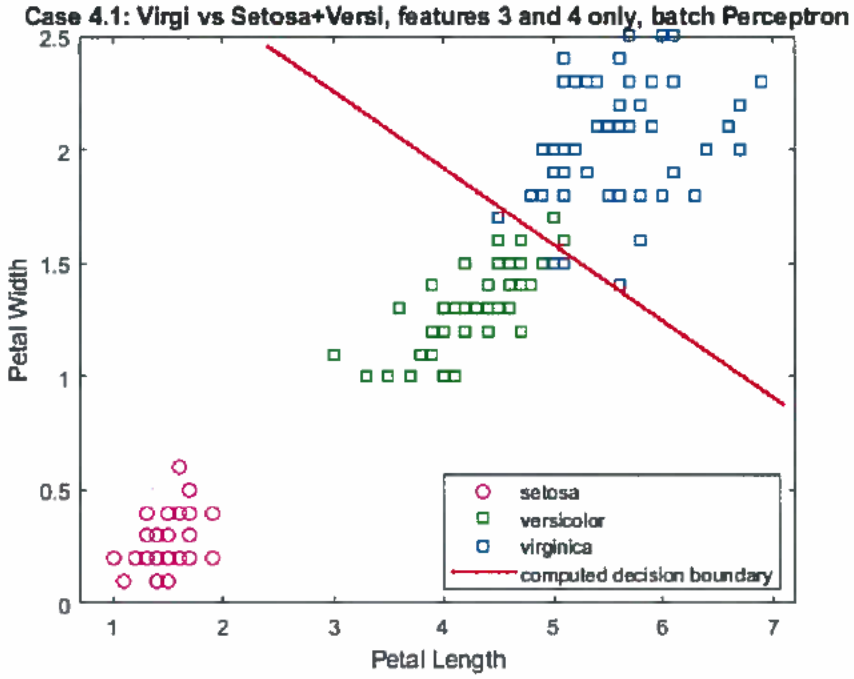
Case 2.2. Setosa vs. Versicolor+Viriginica; features 3 and 4 only; LS	
a. Did the method converge?	Yes (LS always gives an answer)
b. No. of epochs	1 (LS is a one-step algorithm)
c. Computed weight vector	$\begin{bmatrix} -0.2513 \\ 0.0098 \\ 1.2660 \end{bmatrix}$
d. No. of training misclassifications	1 ← Could be Zero but make sure plot Corresponds to this answer
e. Plot of feature vectors and computed decision boundary	

10 pts.

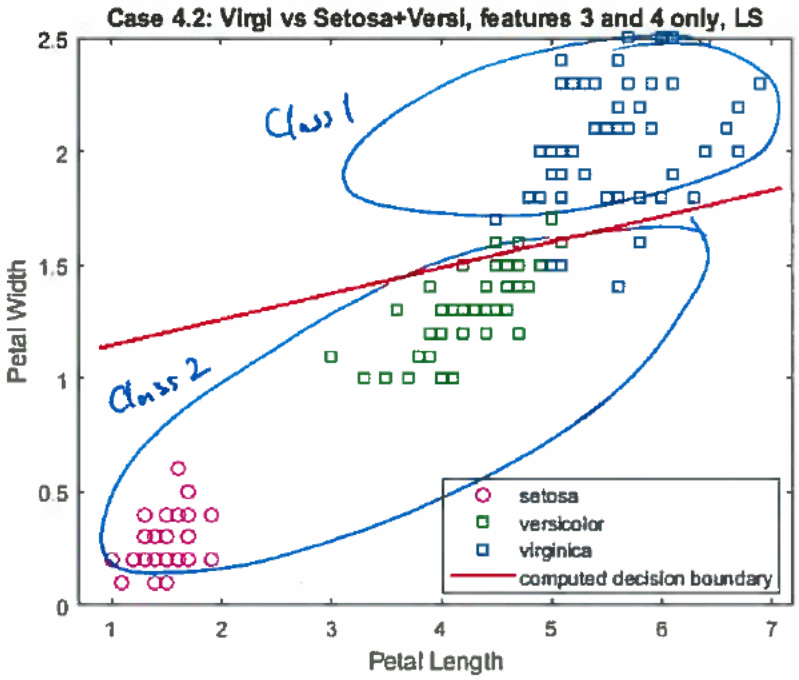
Case 3.1. Viriginica vs. Setosa+Versicolor; all features; batch Perceptron	
a. Did the method converge?	No
b. No. of epochs	1000
c. Computed weight vector	$\begin{bmatrix} -48.7850 \\ -42.4500 \\ 71.5550 \\ 60.8200 \\ -31.0000 \end{bmatrix}$
d. No. of training misclassifications	2
e. Plot of feature vectors and computed decision boundary	N/A

10 pts.

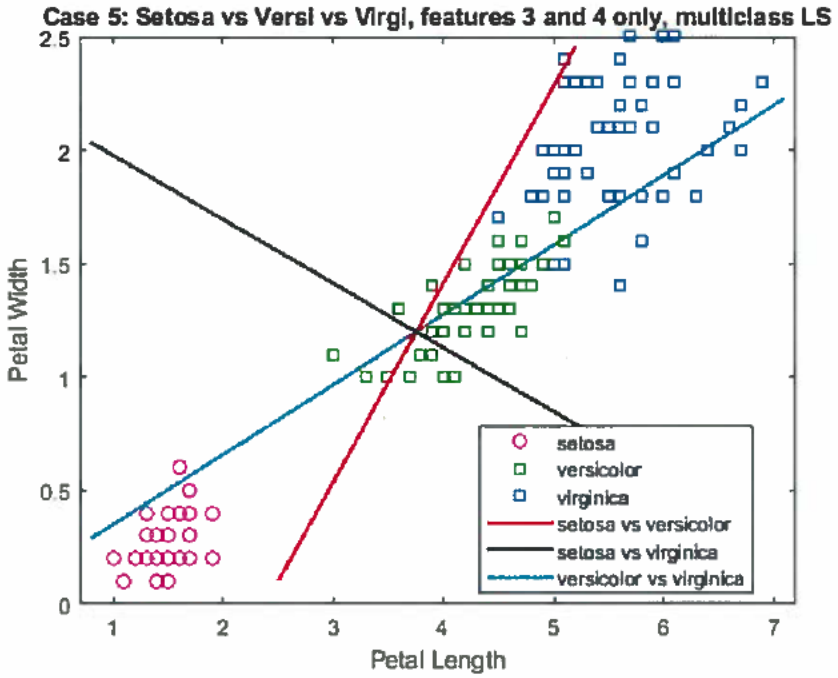
Case 3.2. Viriginica vs. Setosa+Versicolor; all features; LS	
a. Did the method converge?	Yes (LS always gives an answer)
b. No. of epochs	1 (LS is a one-step algorithm)
c. Computed weight vector	$\begin{bmatrix} -0.0459 \\ 0.2028 \\ 0.0040 \\ 0.5518 \\ -0.6953 \end{bmatrix}$
d. No. of training misclassifications	11 Actual value is not important
e. Plot of feature vectors and computed decision boundary	N/A

Case 4.1. Viriginica vs. Setosa+Versicolor; features 3 and 4 only; batch Perceptron	
a. Did the method converge?	No
b. No. of epochs	1000
c. Computed weight vector	$\begin{bmatrix} 9.4400 \\ 27.9700 \\ -91.4500 \end{bmatrix}$
d. No. of training misclassifications	6
e. Plot of feature vectors and computed decision boundary	

location of boundary not important

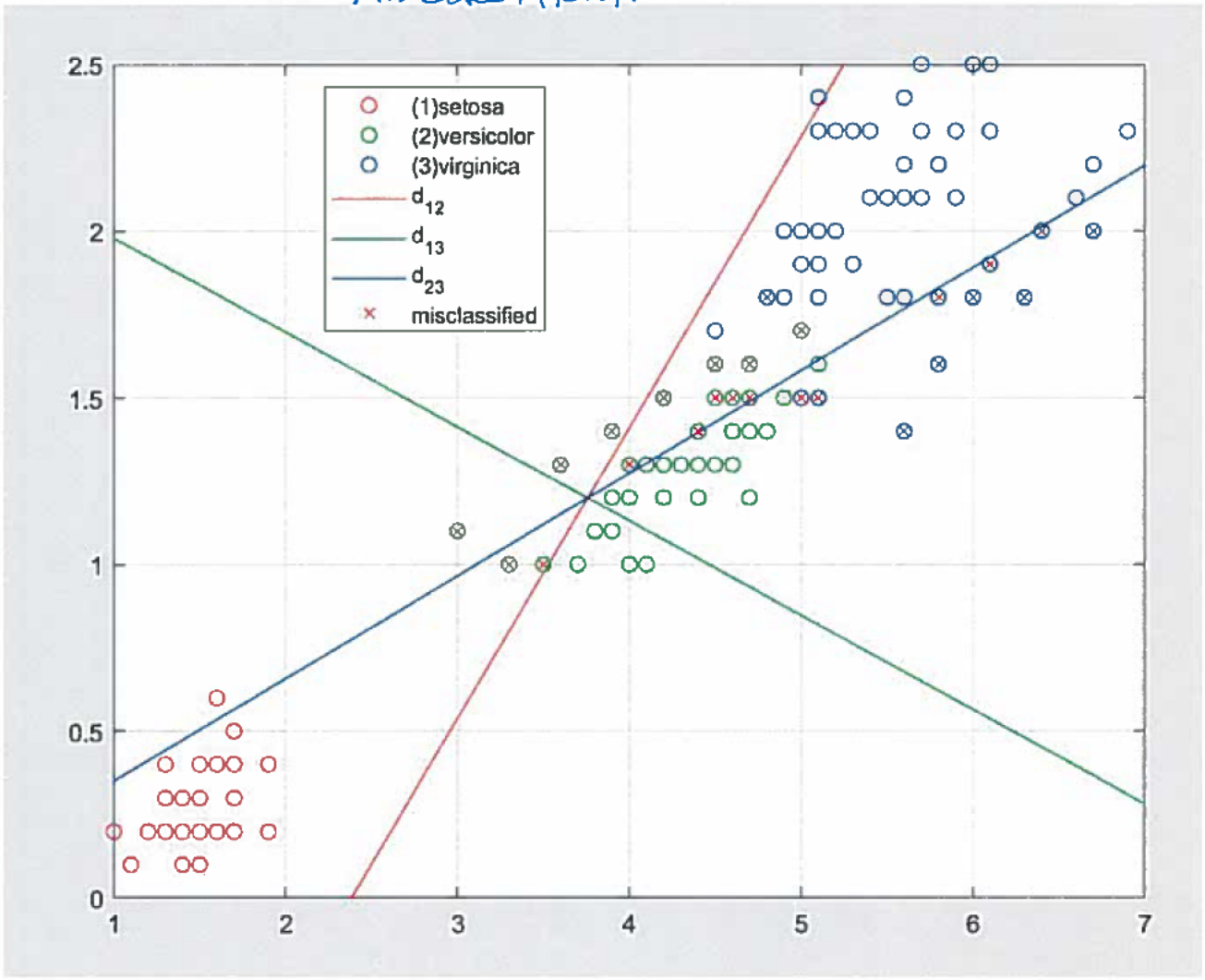
Case 4.2. Viriginica vs. Setosa+Versicolor; features 3 and 4 only; LS	
a. Did the method converge?	Yes (LS always gives an answer)
b. No. of epochs	1 (LS is a one-step algorithm)
c. Computed weight vector	$\begin{bmatrix} -0.0730 \\ 0.6403 \\ -0.1602 \end{bmatrix}$
d. No. of training misclassifications	8 exact no. not important
e. Plot of feature vectors and computed decision boundary Verify that boundary is trying to separate Class 1 from Class 2	

10 pts.

Case 5. Setosa vs. Versicolor vs. Viriginica; features 3 and 4 only; multiclass LS	
a. Did the method converge?	Yes (LS always gives an answer)
b. No. of epochs	1 (LS is a one-step algorithm)
c. Computed weight vector	$\begin{bmatrix} -0.2513 & 0.3243 & -0.0730 \\ 0.0098 & -0.6501 & 0.6403 \\ 1.2660 & -0.1058 & -0.1602 \end{bmatrix}$
d. No. of training misclassifications	34
e. Plot of feature vectors and computed decision boundary	

See next page

total of 34 misclassifications



make sure boundaries are correct

20 pts.