Part B Report

Name: Xinyi Yang (replace with your own name and delete this instruction!)

Assignment 6: Perceptron Classification and Training

CSE 415 Introduction to Artificial Intelligence, Winter 2021, University of Washington

Please answer each question using text in Blue, so your answers stand out from the questions.

QB1. How many epochs were required to train your perceptron on the 3-class Iris data having 4 features (the given training file, with 30 examples) ?

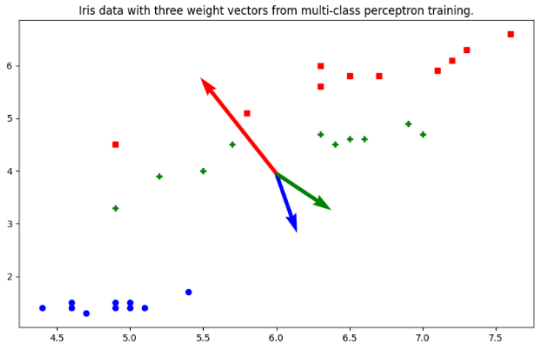
85

QB2. How many of the test data examples (out of 120) were mis-classified? Determine the percentage error rate and write that here.

14

Error rate is approximately 11.67%

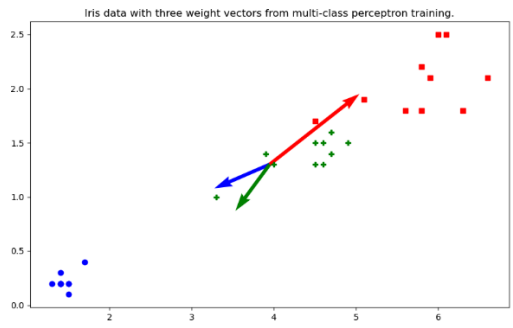
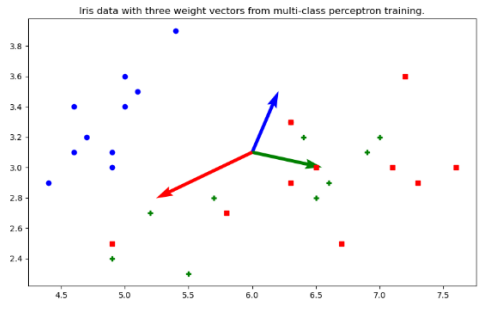
QB3. Capture the plot that is produced by the program showing the training data and the weight vectors when projected onto the 2-D subspace spanned by sepal length and petal length (which is the starter-code default in run_3_class_4_feature_iris_data.py. Paste it here, reduced to fit in the remaining space on this page.



Sepal length, petal length.

QB4. In the file run_3_class_4_feature_iris_data.py, adjust the commenting near lines 23-25 so you can see the data in the plot projects to features 2 and 3 (petal length and petal width). Describe the how

the data seems to be distributed in this view. Describe how the weight vectors seem to be pointing. Finally, describe the relationship between the weight vectors and the distribution of the data.

	
Petal length, petal width.	Sepal length, sepal width.

For the plot on the left:

The data can be divided into three groups of different colors and shapes. The data seems like lying on a diagonal line on the area. The three groups form almost clearly separated three clusters. The starting point of the vectors lies on the center of the plot. Each of them points to the cluster with the same color. The green and blue arrows are short while the red one is longer.

For the plot on the right:

The data in blue lies in the left top corner on the plot, separating from the other two kinds. While the data in green and red are splattered over the other half of the plot, which is difficult to be separated using a straight line. The arrow in green points to the direction where most of the green dots are. The other two arrows do not point to where most of the data with the same colors. The blue arrow points to the top and slightly right. The red arrow points to the bottom left corner where there is only one red dot. I don't really understand the reasons for their lengths and directions.