CSE 474-INTRO TO MACHINE LEARNING

Linear Models for Supervised Learning

Project Assignment 1

Christian Caballero

Clayton Dombrowsky

Ankit Sigroha

Group 34

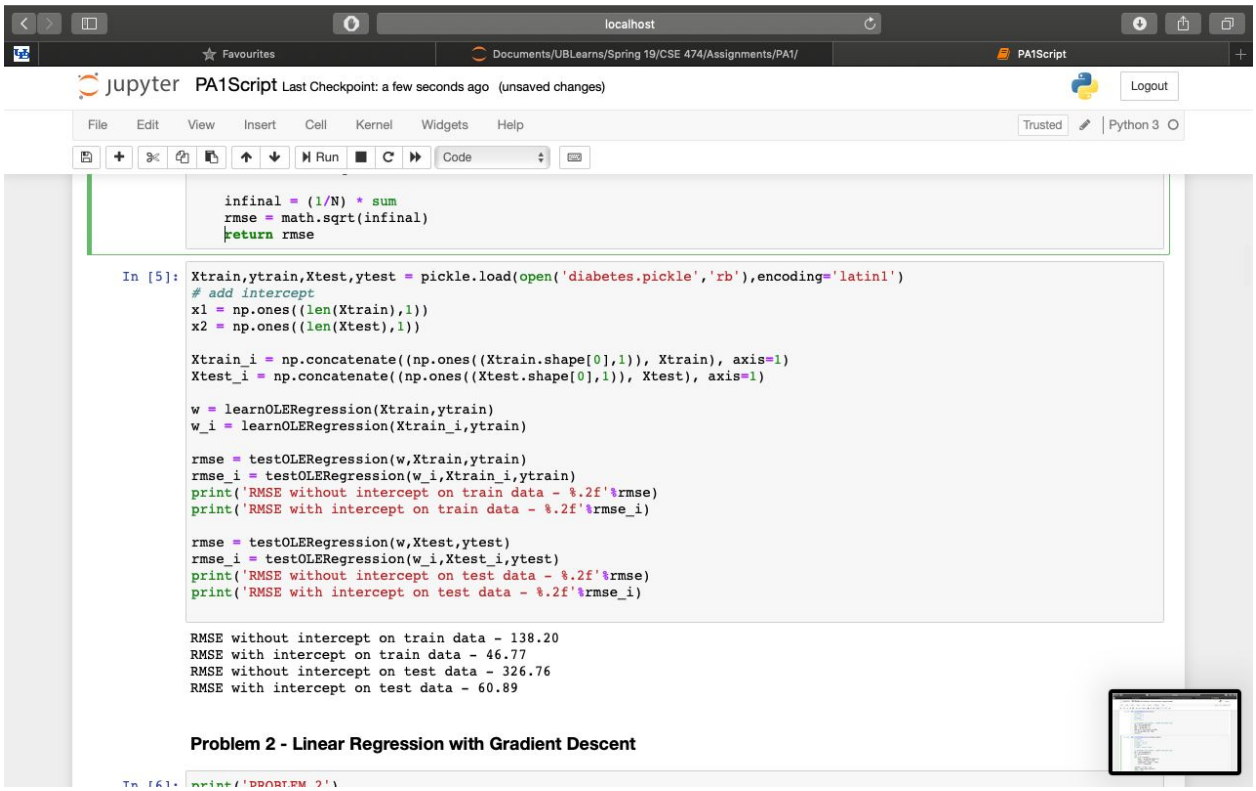
03.06.2019

Report 1:

Calculate and report the RMSE for training and test data for two cases: first, without using an intercept (or bias) term, and second with using an intercept. Which one is better?

Ans: The RMSE without intercept on train data is: 138.20
The RMSE with intercept on train data is: 46.77
The RMSE without intercept on test data is: 326.76
The RMSE with intercept on test data is: 60.89

The RMSE with intercept is better since it is less and we eventually want to minimize the error in the learning on the real data or the test data which is low in this case. Adding an intercept lowers the RMSE error.

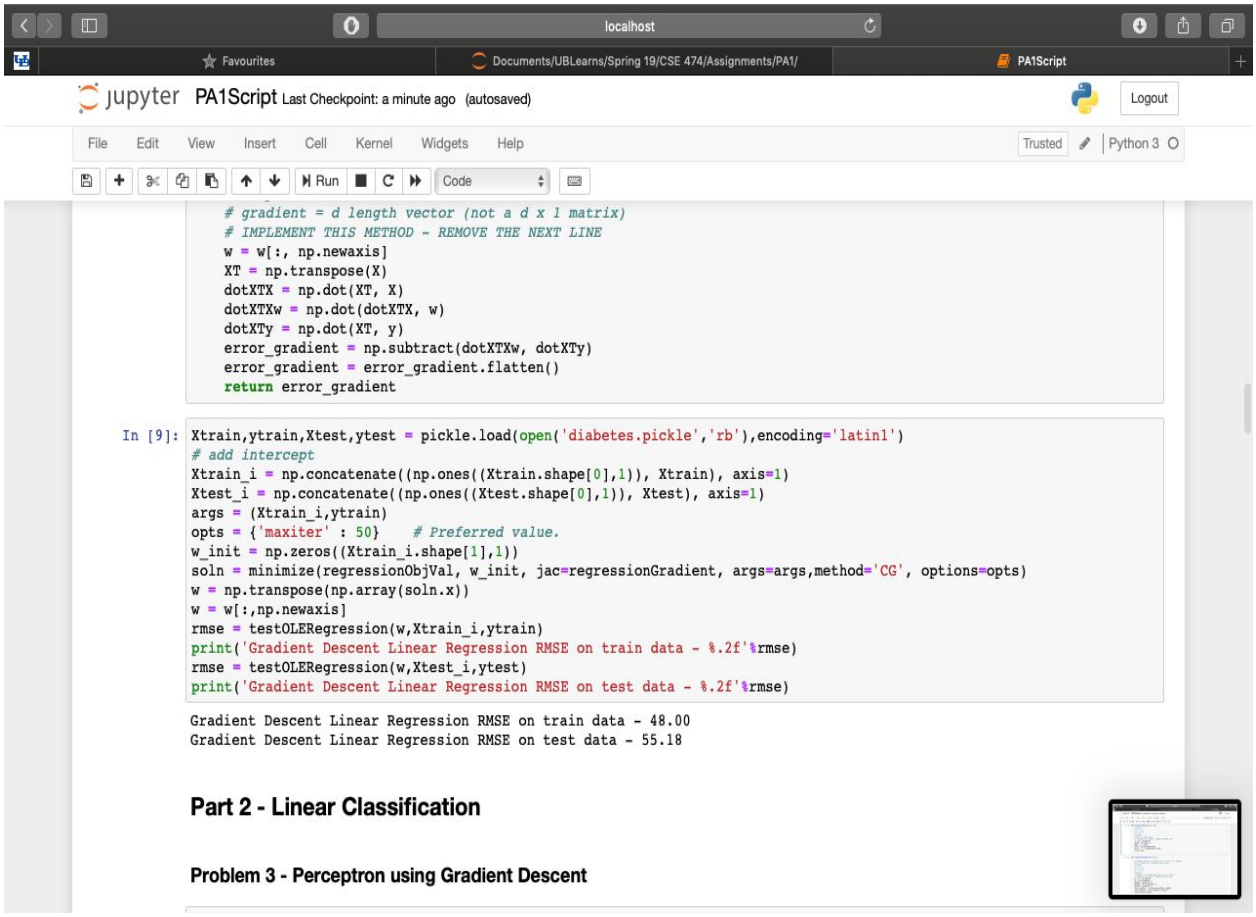


Report 2:

Using testOLERegression, calculate and report the RMSE for training and test data after gradient descent based learning. Compare with the RMSE after direct minimization. Which one is better?

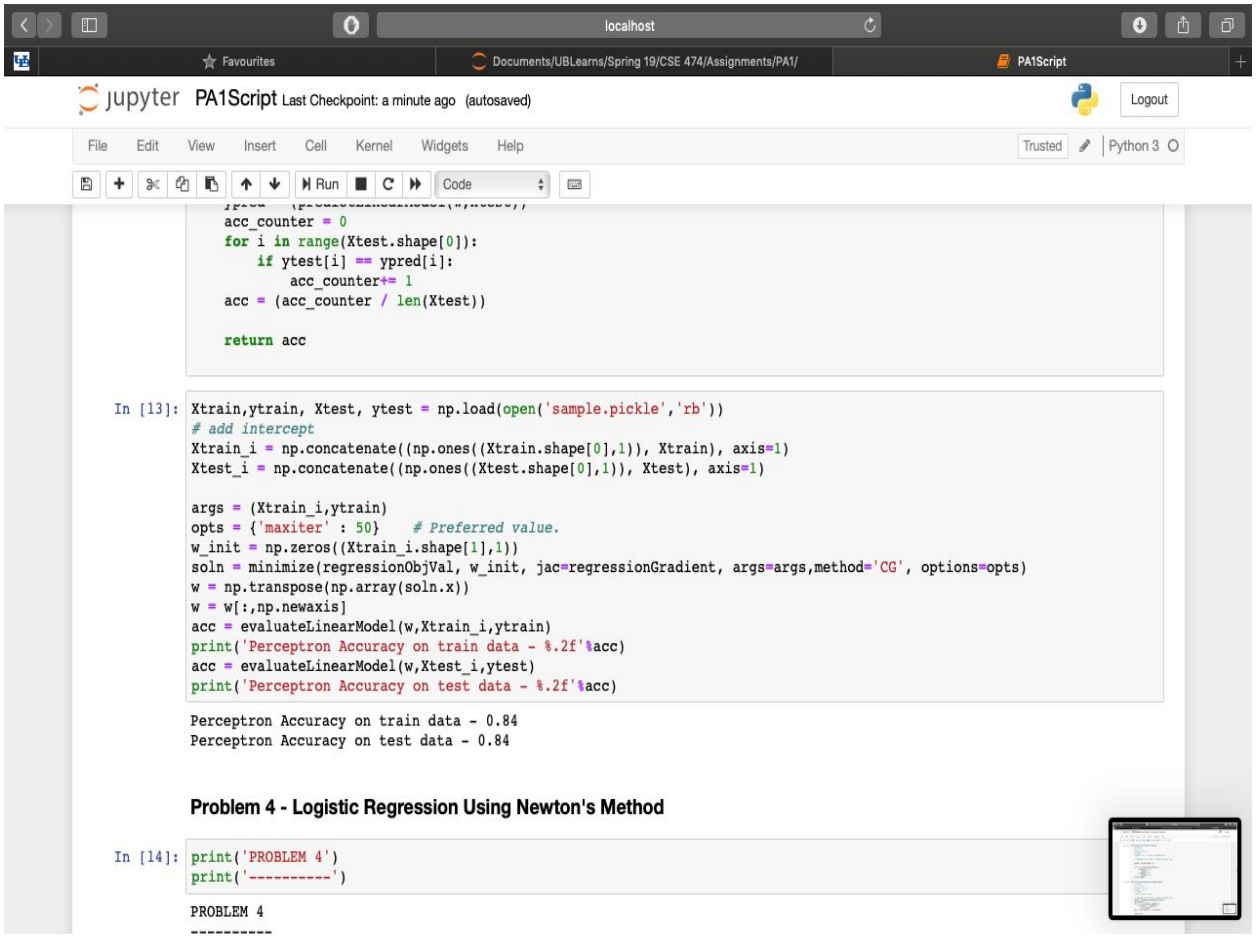
Ans: Gradient Descent Linear Regression RMSE on train data is: 48.00
Gradient Descent Linear Regression RMSE on test data is: 55.18

The RMSE after the minimization using gradient descent is better as it lowers the error than RMSE with an intercept.



Report 3:

Train the perceptron model by calling the scipy.optimize.minimize method and use the evaluateLinearModel to calculate and report the accuracy for the training and test data.
Ans: Perceptron accuracy on train data is: 0.84
Perceptron accuracy on test data is: 0.84

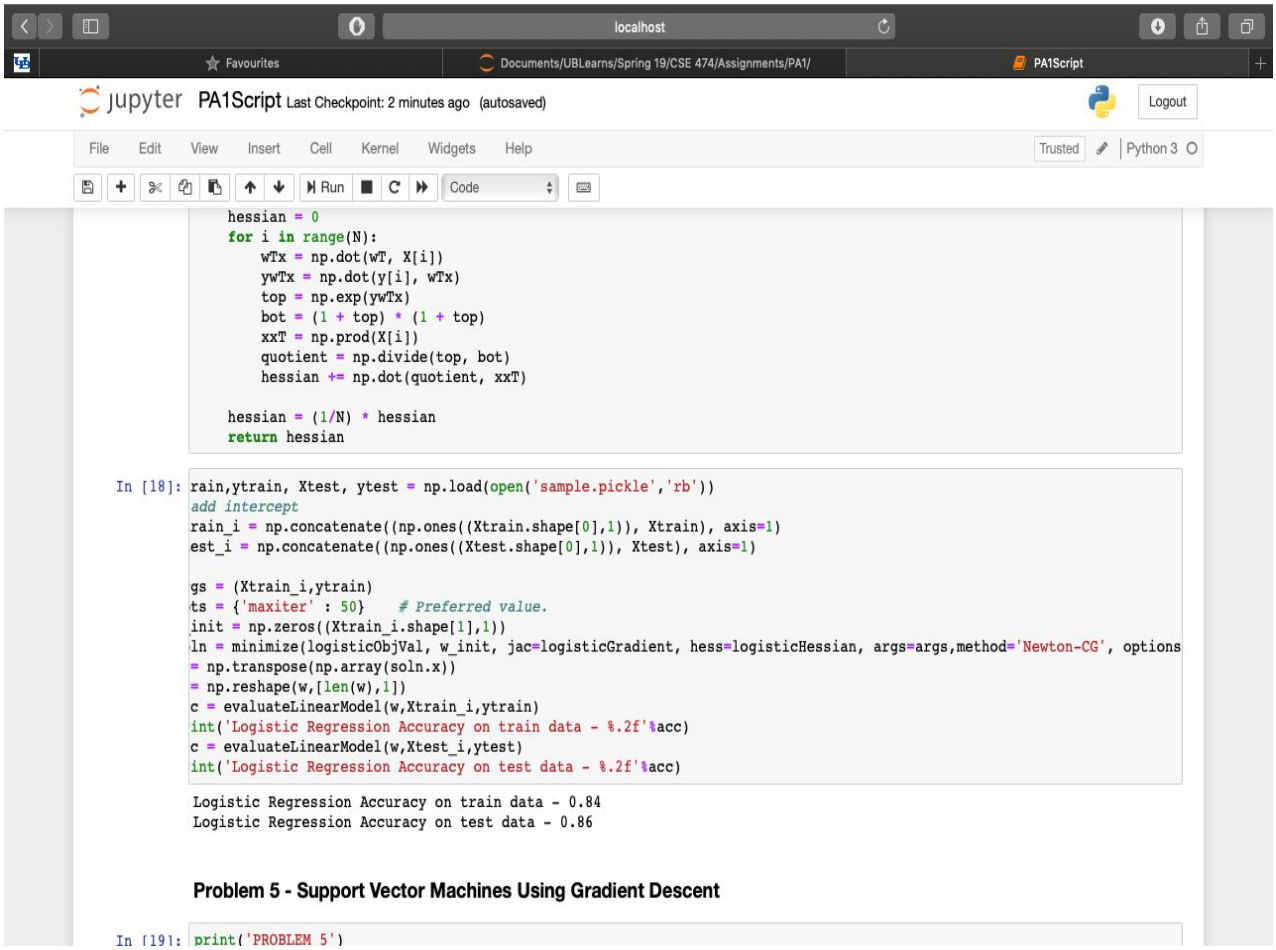


Report 4:

Train the logistic regression model by calling the scipy.optimize.minimize method, and use the evaluateLinearModel to calculate and report the accuracy for the training and test data.

Ans: Logistic Regression accuracy on train data is: 0.84

Logistic Regression accuracy on test data is: 0.86

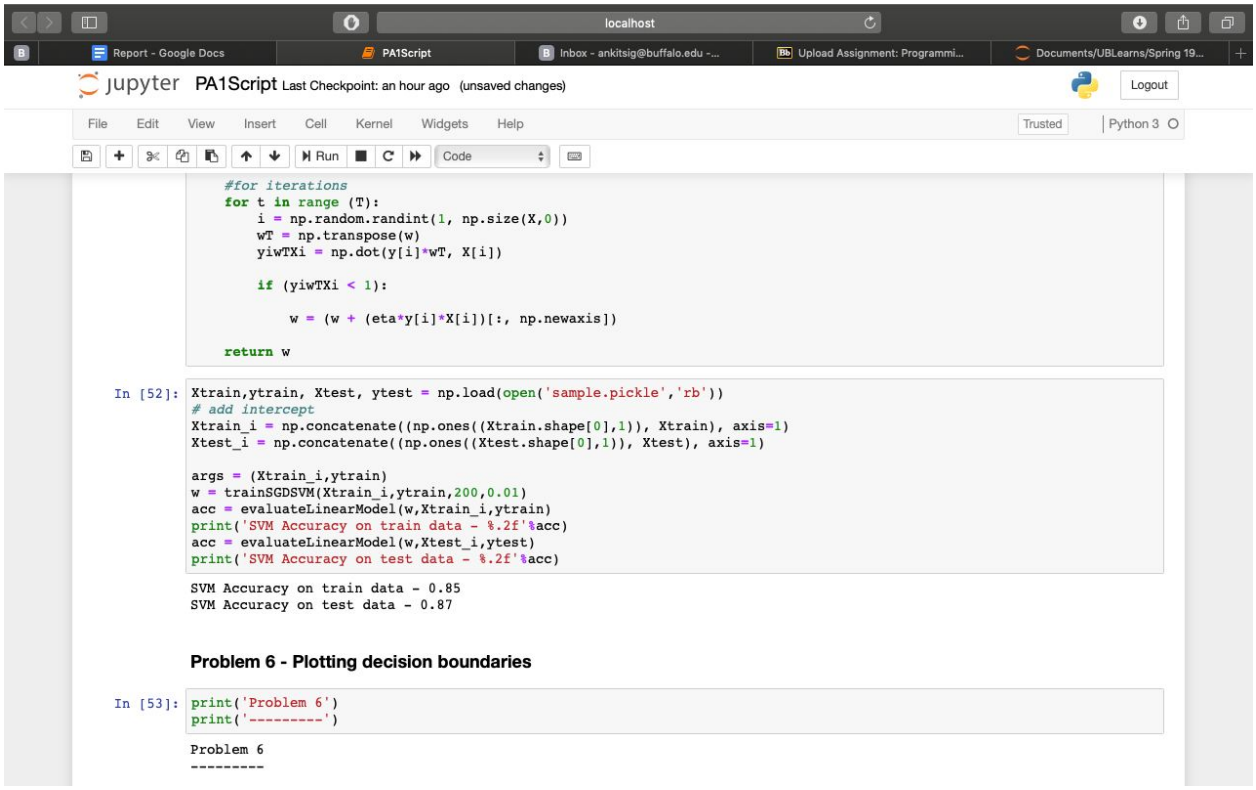


Report 5:

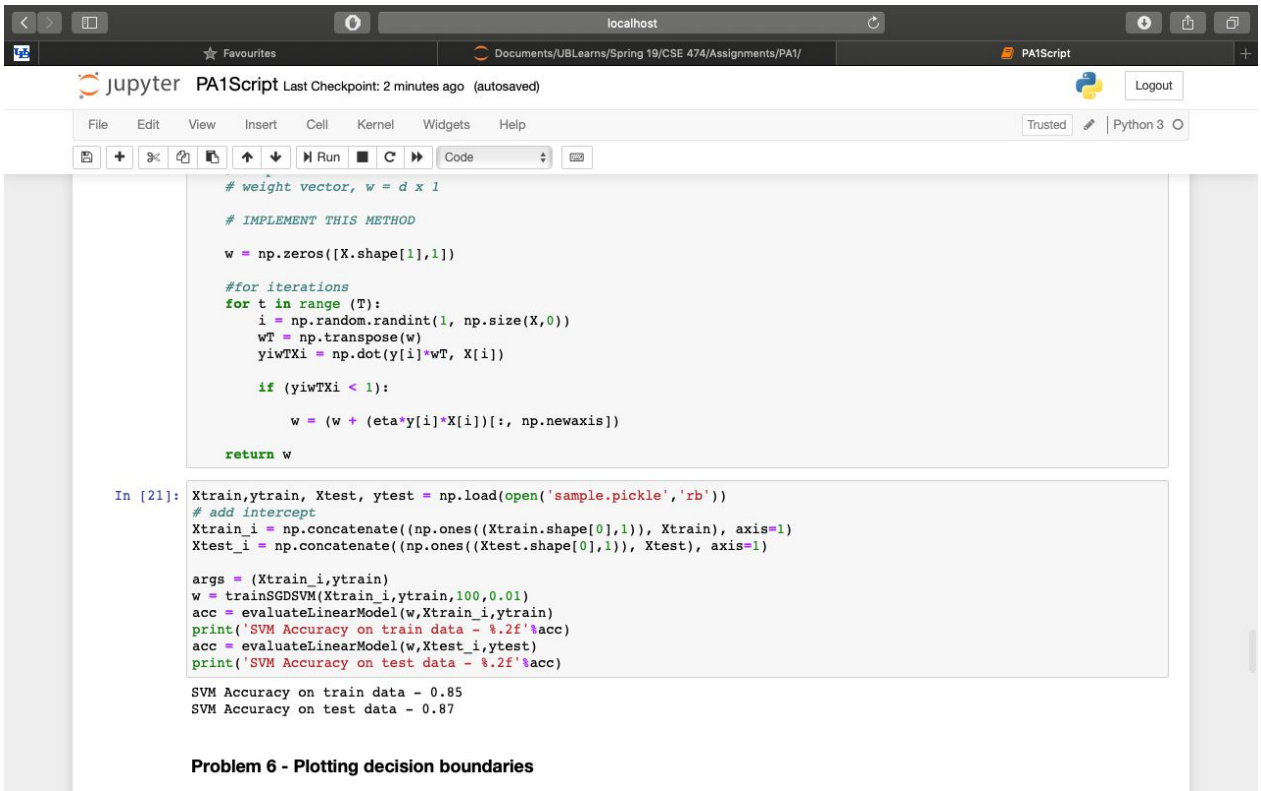
Train the SVM model by calling the trainSGDSVM method for 200 iterations (set learning rate parameter η to 0.01). Use the evaluateLinearModel to calculate and report the accuracy for the training and test data.

Ans: The Accuracy for T = 200 and eta = 0.01 is computed below:-

Ans: SVM accuracy on train data is: 0.85



SVM accuracy on test data is: 0.87



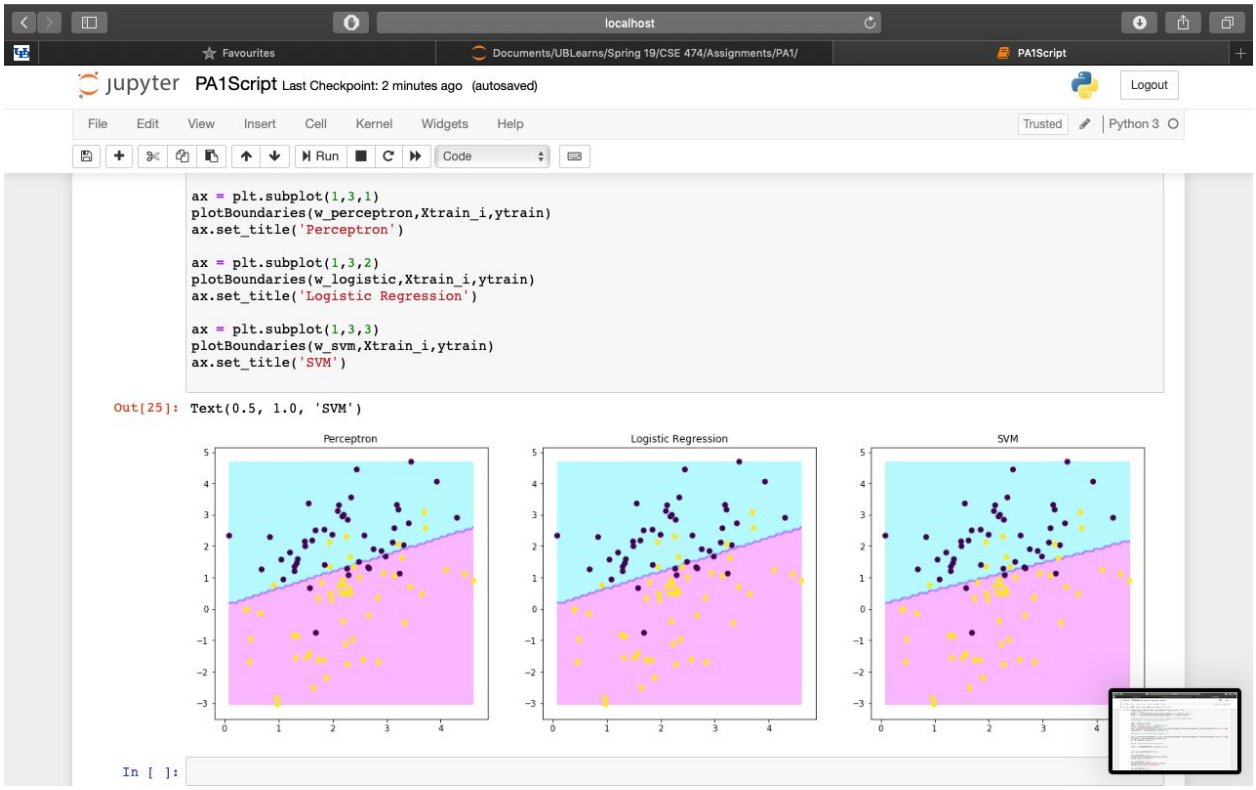
Report 6:

1. Use the results for test data to determine which classifier is the most accurate?

Ans: Out of the three classifiers the SVM is most accurate of all. The Linear regression is second most accurate followed by the perceptron.

2. Plot the decision boundaries learnt by each classifier using the provided plotDecisionBoundary function which takes the learnt weight vector, w as one of the parameters. Study the three boundaries and provide your insights.

Ans: The three decision boundaries are:



Although the accuracy differs by a little but the SVM decision boundary is efficient as it provides a less slack for the points which are quite not separable. The homogeneous classification with the less risk of error makes it most accurate whereas the Perceptron and Logistic Regression classification seems to differ by a little and the decision boundary is flexible to make some errors or to take risk on classifying the data.