CSE474/574 Introduction to Machine Learning Programming Assignment 1

Linear Models for Supervised Learning

Group 48

Dhruv Sureshbhai Patel(#50321707)
Saumya Jay Dholakiya(#50320175)

Report 1

Calculate and report the RSME 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?

No.	Cases	Result
1	RMSE without intercept on test data	138.20
2	RMSE with intercept on test data	46.77
3	RMSE without intercept on test data	326.76
4	RMSE without intercept on test data	60.89

→ From the obtained results of RSME with intercept and without intercept, we can conclude that least RMSE is obtained with intercept.

→ With intercept, the decision boundary is more feasible to distinguish the data. Therefore RSME with intercept (test data) is more better.

Report 2

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

No.	Cases	Result
1	Gradient Decent Linear Regression RMSE on train data	57.05
2	Gradient Decent Linear Regression RMSE on train data	57.24

→ From the obtained result, it can be concluded that there is not much variation in error value of both training and test data.

Report 3

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

No.	Cases	Result
1	Perceptron Accuracy on train data	85.00
2	Perceptron Accuracy on test data	87.00

→ From the obtained result, accuracy of training data as well as testing data are quire nearer to each other.

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.

No.	Cases	Result
1	Logistic Regression Accuracy on train data	81.00
2	Logistic Regression Accuracy on test data	77.00

Report 5

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

No.	Cases	Result
1	SVM Accuracy on train data	81.00
2	SVM Accuracy on test data	88.00

→ From the obtained data, it is concluded that accuracy of SVM on testing data in more than training data.
→ But every time accuracy changes for each iteration. Training data accuracy ranges from 74 to 87 and testing data accuracy ranges from 77 to 88.
→ In most of the iteration, testing data accuracy is more than training data.

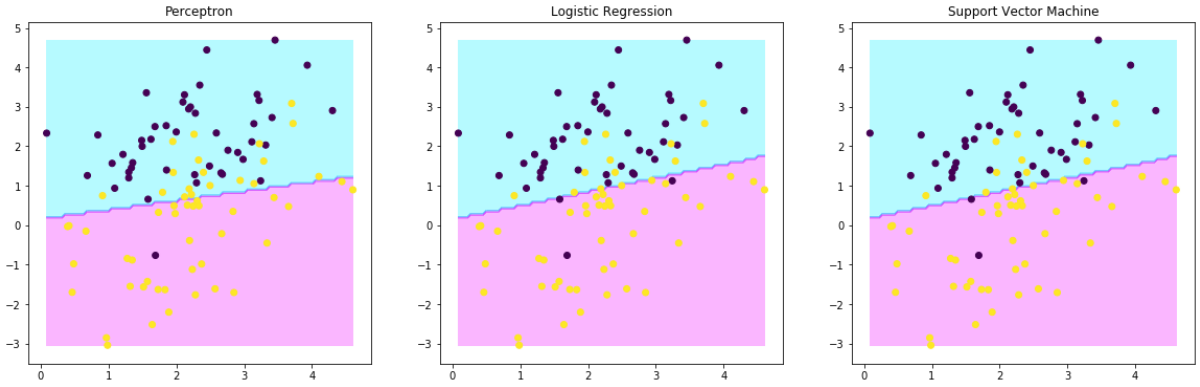
Report 6

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

	Perceptron	Logistic Regression	Support Vector Machine
Train	85.00	81.00	82.00
Test	87.00	77.00	88.00

→ From the above table, it is shown that support vector machine gives more accurate decision boundary. Though, SVM does not give more accurate value in each iteration. Therefore, it is also important to look at the accuracy of each model and then choose according to that.
→ Still, SVM gives optimal decision boundary compare to other models, as SVM try to maximize margin between closest support vector.

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



→ All three plots look similar as all model's accuracy are nearer to each other.

→ SVM has highest test accuracy among all other model, which indicate that it has more 'True Positive' and 'True Negative' points compare to others. Which means SVM has least number of data points which are identify as wrong targeted value.