

Linear Models for Supervised Learning

Programming Assignment 1

CSE 474: Introduction to Machine Learning

Group 38

Project Members :

Vineeth Nallabotu(vineethn@buffalo.edu)

Maniteja Manchikalapudi(maniteja@buffalo.edu)

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?

RMSE without intercept on train data - 138.20.

RMSE without intercept on test data - 326.76

RMSE with intercept on train data - 46.77

RMSE with intercept on test data - 60.89.

From the above two cases we can conclude that RMSE with intercept is better as it has lesser values which means it gives us better estimate.

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?

Gradient descent linear regression RMSE on train data - 48.01

Gradient descent linear regression RMSE on test data - 54.93

After comparing to the RMSE in report 1, RMSE after gradient descent based learning is worse for train data and is better on test data

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.

Perceptron accuracy on train data - 0.84

Perceptron accuracy on test data - 0.84

Both are same.

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.

Logistic regression accuracy on train data - 0.84

Logistic regression accuracy on test data - 0.86

Report 5:

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

SVM accuracy on train data – 0.83

SVM accuracy on test data – 0.87

Report 6:

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

From the above results , we can conclude that SVM is the most accurate .

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.

