Assignment 1

1. Implement the linear regression algorithm to estimate the weight parameters for the feature matrix (X) and the class label vector (y). (a) Plot the cost function vs the number of iterations. (b) Plot the cost function (J) vs w1 and w2 in a contour or 3D surf graph (w= [w0 w1 w2]). Please use the dataset "data.xlsx". (Use for or while loop for the implementation)

2. Implement stochastic gradient descent for the linear regression problem in question number 1. (a) Plot the cost function vs the number of iterations. (b) Plot the cost function vs w1 and w2. (Please use the dataset "data.xlsx"). (Use for or while loop for the implementation)

3. Implement the ridge regression problem by considering both batch gradient descent and stochastic gradient descent. (a) Plot the cost function vs the number of iterations for both the cases. (b) Plot the cost function (J) vs w1 and w2 in a contour or 3D surf graph for both the cases. (Please use the dataset "data.xlsx"). (Use for or while loop for the implementation)

4. Implement the Vectorized linear regression problem to evaluate the weight parameters for question number 1. Compare the weight parameters with the weights obtained using both gradient descent and stochastic gradient descent based algorithms. (Please use the dataset "data.xlsx").

5. Implement Least angle regression to estimate the weight parameters for the feature matrix (X) and the class label vector (y) by considering both gradient descent and stochastic gradient descent based algorithms. (Please use the dataset "data.xlsx"). (Use for or while loop for the implementation).

6. Implement K-means clustering based unsupervised learning algorithm for the dataset ("data2.xlsx"). Plot the estimated class labels vs features. Use the number of clusters as K=2.

7. Implement the logistic regression algorithm for the binary classification using the dataset ("data3.xlsx"). Divide the dataset into training and testing using hold-out cross-validation technique with 60 % of instances as training and the remaining 40% as testing. Evaluate the accuracy, sensitivity and specificity values for the binary classifier.

8. Implement the multiclass logistic regression algorithm using both "One VS All" and "One VS One" multiclass coding techniques. Evaluate the performance of the multiclass classifier using individual class accuracy and overall accuracy measures. Use the hold-out

cross-validation approach (60% training and 40% testing) for the selection of training and test instances. (Please use the dataset “data4.xlsx”)

9. Evaluate the performance of multiclass logistic regression classifier using 5-fold cross-validation approach. Evaluate the individual class accuracy and overall accuracy measures for the multiclass classifier along each fold. (Please use the dataset “data4.xlsx”)

10. Use the likelihood ratio test (LRT) for the binary classification using the dataset (“data3.xlsx”). Divide the dataset into training and testing using hold-out cross-validation technique with 60 % of instances as training and the remaining 40% as testing. Evaluate the accuracy, sensitivity and specificity values for the binary classifier.

11. Implement the Maximum a posteriori (MAP) decision rule for multiclass classification task. Use the hold-out cross-validation approach (70% training and 30% testing) for the selection of training and test instances. (Please use the dataset “data4.xlsx”).

12. Implement the Maximum likelihood (ML) decision rule for multiclass classification task. Use the hold-out cross-validation approach (70% training and 30% testing) for the selection of training and test instances. (Please use the dataset “data4.xlsx”).

13. Please write in your own words that what you have learned by solving the Assignment 1.

Dataset description

data4.xlsx: (This dataset contains 4 features for different instances as four columns and fifth column is for class label)
data3.xlsx: (This dataset contains 4 features for different instances as four columns and fifth column is for class label)
data2.xlsx: (This dataset contains 4 features for different instances as four columns)
data.xlsx (First 2 columns as features and last column as class labels (continuous values)