Lab Assignment 5 (Group Assignment)
CS 331 - CS 530: Machine Learning
Course Instructor : Dr. Satyanath Bhat
Teaching Assistants : Vipin Gautam, Divya Mishra, Shreyansh Aswal
Faculty Fellow: Pooja Dalvi

INSTRUCTIONS: Please submit your solutions for assignment in google classroom as ipynb files. Only one member needs to make the submission in the classroom.

I Perceptron

Answer all of the following questions

1. (Perceptron) Using the make_blobs function in sklearn generate a dataset of 100 points with two classes. Write the perceptron algorithm by scratch and show the intermediate hyperplanes generated.

2. (Perceptron) Repeat the above exercise with batch gradients instead. Step size might need to be adjusted for convergenence.

3. (Perceptron) Using the make_circles function in sklearn generate two classes so that they form different concentratic circles. Generate 100 points. Create second order features and train the perceptron. Using the contour function show the final decision boundary in the original two dimensional space.

4. (Perceptron) Repeat the two perceptron example above (make_blobs and make_circles) generating 1000 points. We are not interested to see the plots now. We instead want to split the dataset into test (50%) and train(50%). Report the test accuracy of perceptron in each case.

II Regression

Reference: Stanford lecture notes: `https://see.stanford.edu/materials/aimlcs229/cs229-notes1.pdf`

1. Implement LMS algorithm for linear regression from scratch. Visualize the learnt house prices on the scatter plot of the input training dataset

2. On the full batch gradient descent visualize the contours of J(W) for different values of the learning rate η.

3. On the stochastic gradient descent visualize the contours of J(W) for different values of the learning rate and batch sizes η.

4. Implement locally weighted linear regression as described in Stanford lecture notes.

III Logistic Regression

1. Using the make_blobs function in sklearn generate a dataset of 100 points with two classes. Implement Logistic regression with cross entropy loss
2. Using the make_blobs function in sklearn generate a dataset of 100 points with two classes. Implement Logistic regression with least mean square loss
3. Compare the trajectory of gradient descent (batch) with both cross entropy loss and least mean square loss

IV Regularization

1. Generate points with the model $y = ax+b+\epsilon$ where epsilon is standard gaussian. x is distributed as uniform rv between [0,10]. Train a linear regression model with following polynomials

   - 2
   - 5
   - 10

   Study the out of sample performance for each of the above. Compare this when training dataset size is changed.

2. On the above, fix a suitable training dataset size, train a 10 degree polynomial which exhibits overfitting. Implement the following regularization schemes

   - Lasso
   - Ridge
   - Elastic

   Observe how the coeffients change via a plot for different values of regularization constant. Using a validation approach fix a regularization constant. Implement a six-fold cross validation method for fixing regularization constant.