Machine Learning Online | Assignment - III

Logistic Regression | Classification Problem

Consider the log-likelihood function for logistic regression:

```
L(θ) = y(i) loghθ(x(i)) + (1−y(i))log(1−hθ(x(i)))
```

1. The files “logisticX.csv”and “logisticY.csv” in the datasets folder contain the inputs (x(i) ∈ R2) and outputs (y(i) ∈ {0,1}) respectively for a binary classification problem, with one training example per row. Implement Mini-Batch Gradient Descent method for optimizing L(θ), and apply it to fit a logistic regression model to the data. Initialize params θ = ∼0 (the vector of all zeros). What are the coefficients θ resulting from your fit? (Remember to include the intercept term.)

2. Plot the training data (your axes should be x1 and x2 , corresponding to the two coordinates of the inputs, and you should use a different symbol for each point plotted to indicate whether that example had label 1 or 0). Also plot on the same figure the decision boundary fit by logistic regression. (i.e., this should be a straight line showing the boundary separating the region where $h(x) > 0.5$ from where $h(x) \leq 0.5$.)

3. **Bonus Problem** Read and Implement Newton's Method (given in Andrew NG CS229 Notes) for Part-A. Compare your results - Final Loss and Total Iterations taken before convergence.

4. Repeat Part-A and Part-B using Sci-kit Learn Logistic Regression Class. Compute 5-fold Cross Validation Score on Training Set. There is no test set for the current problem.