CS 229 Machine Learning, spring 2019

Homework 3:

Logistic Regression

Due Tuesday Feb 26, 11:59pm

Submit by the **blackboard system**

The goal of this homework is to become familiar with logistic regression.

**Question 1 (20 points)**

**Show the derivation of negative log-likelihood of logistic regression problem.**



**Question 2 (20 points)**

**Show how to maximize NLL(w) and find w\* by gradient descent.**

**Or other difference solutions (you can get a bonus of 20 points if you find a different solution)**

**Implementation Task:**

**Code:**

Write your code by any programming languages and submit your results together with the programs.

**Data:**

Please download data logreg\_data\_binary.txt. It includes four columns.

The **first column** coded the **target variable** of “apply to graduate school”, unlikely (0), or likely (1).

The **other three columns** are three variables as follows:

1. **parent**, which is a 0/1 variable indicating whether at least one parent has a graduate degree;
2. **public**, which is a 0/1 variable where 1 indicates that the undergraduate institution is a public university and 0 indicates that it is a private university,
3. **gpa**, which is the student's grade point average.

In other words, each undergraduate student is described by ***x***, which is a 3-dim vector. Can we make a prediction of his/her target ***t*** =?

**Learning method:** You can use gradient decent.

**Task 1 –** (40 points) **Logistic Regression with Binary target**

Implement the logistic regression algorithm for this binary classification problem.

1. Show the decreasing of NLL (negative log-likelihood) function with the increasing of iteration numbers.
2. Give the results of obtained coefficient, **w**, and discuss the meaning of the values in **w**
3. Download test data at test\_data\_binary.txt. How many target labels of test data are correctly predicted by the learned w?

**Task 2 –** (20 points) Use only the **gpa** as student’s feature, redo the training and testing.

1. What is the testing accuracy (the percentage of accurately classified testing samples among all testing data)?
2. Plot a figure with **gpa** as **x-axis,** and **t** as **y-axis**, show the **training** data (**gpa,t**) in blue stars, and the (**gpa, predicted-t**) in red circles.
3. If **0.5** is used as the threshold for determining the classification label (t=0 if prediction <0.5, t=1 if prediction >=0.5), what is the corresponding cutting value of GPA? (i.e., GPA > what, t will be 1)? NOTE: you may have normalized the feature values of GPA. Here, you are asked to give the GPA value in the original scale.

NOTE:

1. Data should be standardized, e.g., for one variable x using x’=(x-mean(x))/std(x) so that x’ has mean(x’)=0 and std(x’)=1.

Standardization should be down for all three variables.

The testing data should be standardized by the mean and std obtained from the variable values in training data.

1. One more dimension with value 1 should be added to each example.