Homework 3
Applied Machine Learning
Fall 2017
CSCI-P 556/INFO-I 526

Instructor: Hasan Kurban

October 17, 2017

Directions

Please follow the syllabus guidelines in turning in your homework. I am providing the LaTeX of this document too. This homework is due Friday Oct 27, 2017 11:59p.m. **OBSERVE THE TIME**. Absolutely no homework will be accepted after that time. Bring a hard-copy to Tuesday's class on the 1st. If you do not bring a hard-copy with the statement of your own work, the homework will not be accepted. All the work should be your own. Within a week, AIs can contact students to examine code; students must meet within three days. The session will last no longer than 5 minutes. If the code does not work, the grade for the program may be reduced. Lastly, source code cannot be modified post due date.

Linear and Logistic Regression

This part is provided to help you implement linear and logistic regression.

Notations

- Δ: data set
- m: number of training examples, n: number of features, x's: input variables, y's: output variable.
- $(x^{(i)}, y^{(i)}) : i^{th}$ training example
- $x_j^{(i)}$: value of feature j in i^{th} training example
- $x_0 = 1$ (the first feature (x_0) is a vector of 1's) – you should add $x_0 = 1$ to data before answering the questions.
- α: learning rate

Linear Regression

Parameters: $\theta = (\theta_0, \ldots, \theta_n)$

Hypothesis/Model: $h_\theta(x) = \theta(x) = \theta^T x = \theta_0 x_0 + \ldots + \theta_n x_n$

Cost Function: $J(\theta_0, \ldots, \theta_n) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$

$$\Rightarrow \frac{\partial}{\partial \theta_j} J(\theta_0, \ldots, \theta_n) = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})\ x_j^{(i)}$$

Logistic Regression

Parameters: $\theta = (\theta_0, \dots, \theta_n)$

Hypothesis/Model: $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$

Cost Function: $J(\theta_0, \dots, \theta_n) = \frac{1}{m}\big[-y^{(i)}\ log(h_\theta(x^{(i)})) - (1-y^{(i)})\ log(1-h_\theta(x^{(i)}))\big]$

$$\Rightarrow \frac{\partial}{\partial\theta_j} J(\theta_0, \dots, \theta_n) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})\ x_j^{(i)}$$

Linear Regression and Logistic Regression via Gradient Descent

Repeat until convergence$\{$

$$\theta_j \leftarrow \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m} h_\theta(x^{(i)}) - y^{(i)})\ x_j^{(i)}$$

$\}$ (simultaneously update θ_j for all j)

Problem 1 [100 points]

Implement gradient descent algorithm for linear regression and answer the following questions. In this question, you are asked to use your gradient descent implementation to fit linear models to Auto data set which can be found in the "ISLR" package.

```
> require("ISLR")
> Auto
```

Initialization of the parameters

- Add $x_0 = 1$ to data (all ones feature)
- Set the learning rate α to 0.01 and iteration number to 1000. You may need to use different α and iteration number values if you observe they are not sufficient.
- Initialize θ's as 0's – $(\theta_0, \dots, \theta_n) = (0, \dots, 0)$

Simple Linear Regression [45 points]

1.1 Perform a simple linear regression with "mpg" as the response and "horsepower" as the predictor. What are the parameters $\theta = (\theta_0, \theta_1)$? Is the relationship between horsepower and mpg positive or negative? [20 pt]

1.2 Plot the output variable and the input variable. Display the least squares regression line. [5 pt]

1.3 Use the model obtained in Q.1.1 to make predictions. What is the "mpg" value for "horsepower = 220"? [5 pt]

1.4 In a contour plot, show how $J(\theta)$ varies with changes in θ_0 and θ_1. Does $J(\theta)$ have a global minimum? [10 pt]

1.5 The closed-form solution to linear regression is $\theta = (\Delta^T\Delta)^{-1}\Delta^T y$. Report the coefficients using this formula. [5 pt]

Multivariate Linear Regression [55 points]

1.6 First, perform feature scaling (mean normalization) over the Auto data set to make gradient descent converge faster. Then, train a multivariate linear regression with "mpg" as the response and all other variables except name as the predictors. Report the parameters (θ's). What does the coefficient for the "year" variable suggest? [30 pt]

1.7 Use the model obtained in Q.1.6 to make predictions. What is the "mpg" value for $(x_1, \ldots, x_7) = (4, 300, 200, 3500, 11, 70, 2)$? [5 pt]

1.8 In this question, you are asked to test different learning rates. Run your gradient descent for 100 iterations at the chosen learning rates ($\alpha_1 = 3, \alpha_2 = 0.3, \alpha_3 = 0.03, \alpha_4 = 0.00003$). For each learning rate, make a plot that shows how $J(\theta)$ changes at each iteration. Discuss the plots? i.e., which one looks better? does it converge? [15pt]

1.9 Calculate the coefficients using the normal equations. [5pt]

Problem 2 [35 points]

From textbook, Chapter 3 exercises 13, 14 and 15 (Pages 124-126).

Problem 3 [65 points]

Implement gradient descent algorithm for logistic regression and answer the following questions. In this question, you are asked to use your gradient descent implementation to train logistic regression models over Auto data set.

Initialization of the parameters

- Add $x_0 = 1$ to data (all ones variable)
- Set the learning rate α to 0.01 and iteration number to 1000. You may need to use different α and iteration number values if you observe they are not sufficient.
- Initialize θ's as 0's – $(\theta_0, \ldots, \theta_n) = (0, \ldots, 0)$

Data Preprocessing

- Create a binary variable, "mpg01", that contains a 1 if mpg contains a value above its median, and a 0 if mpg contains a value below its median
- Create a new data set called new.Auto by extracting the features mpg01, cylinders, displacement, horsepower, weight from Auto data set. Keep the order of the features as shown here in the new.Auto data set
- Do feature scaling (mean normalization) over the input variables of the new.Auto data set to make gradient descent converge faster.
- Use the new.Auto data set to answer the following questions

Logistic Regression

3.1 Implement the sigmoid function and make a plot of it by testing different inputs. $\left(g(z) = \frac{1}{1+e^{-z}}\right)$ [5 pt]

3.2 Perform logistic regression on the new.Auto data set in order to predict "mpg01" using the input variables: cylinders, displacement, horsepower, weight. Report the parameters (θ's). [30 pt]

3.3 What is the error of the model over new.Auto data set? [10 pt]

3.4 Use the model obtained in Q.3.1 to make predictions. What is the "mpg01" value for $(x_1, \ldots, x_4) = (8, 340, 200, 3500)$ (first, scale the data point with the parameters obtained earlier while normalizing the features) [5 pt]

3.5 In this question, you are asked to test different learning rates. Run your gradient descent for 100 iterations at the chosen learning rates ($\alpha_1 = 3, \alpha_2 = 0.3, \alpha_3 = 0.03, \alpha_4 = 0.00003$). For each learning rate, make a plot that shows how $J(\theta)$ changes at each iteration. Discuss the plots? i.e., which one looks better (faster)? does it converge? [15pt]

What to Turn-in – Submission Instructions

- Zip the files requested below for your submission. The zipped folder should be named as "usename-section number", i.e, hakurban-P556
  - The *tex and *pdf of the written answers to this document
  - *`R`files for:
    - Gradient descent implementation for problems 1.1 and 1.6 – file name: "*linearRegression.R*".
    - Normal equation implementation for problems 1.5 and 1.9 – file name: "*normalSolution.R*".
    - Gradient descent implementation for problem 3.2 – file name: "*logisticRegression.R*".
    - new.Auto data set – file name: "*newAuto.R*".
  - A README file that explains how to run your code and other files in the folder