CS 340 Data Mining, Fall 2018

Homework 1: Regression

Due Nov 19, 11:59pm

Submit by the **blackboard system**

***Requirements:***

1. Discussion is encouraged, but independent write-up and submission in your own hand is required for credit.
2. The work should be written in a \***clear**\* way.
   1. The process of your solution must be unambiguously outlined or explicitly shown.
   2. The explanation and discussion must be well-founded by examples, tables, figures, and literatures and so on.
3. Work submitted after the due date may be graded for correctness, but not credited.

**Linear Regression**

(**5 points**) The goal of this homework is to become familiar with the gradient descent algorithm and linear regression algorithm.

**Data**: Please download the data regression\_x\_t.txt. It includes two columns, one is the variable ***x***, and the other is the target ***t***. Randomly choose 80% of the data for learning the regression function. Use the remaining 20% for evaluating the learned function.

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

**Regression basis function:** Choose any kind of basis function, e.g., polynomial function, or Gaussian basis function. Try an appropriate number of basis functions.

**Questions:**

1. (2.5 pts) implement the ***batch gradient descent*** algorithm.
2. Show the decreasing of error function with the increasing of iteration numbers.
3. Give the results of obtained coefficient, **w**
4. Show the predicted ***f*(*x*)** when applying the learned regression function on testing data ***x***, and compare it with the corresponding actual target ***t*** on the same figure;
5. Did you make good prediction?
6. (2.5 pts) implement the ***one-step*** solution for ***minimizing sum-of-square error***.
7. Give the results of obtained coefficient, **w**
8. Compare the value of **w** obtained at 1) by gradient descent algorithm, and 2) by one-step solution
9. Show the predicted ***f*(*x*)**, and compare it with the corresponding actual target ***t*** on the same figure;
10. Compare the performance of prediction in this question with that in 1)

**Neural Network for Regression** (**5 points**)

1. (2 pts) Applying Neural Network on the regression problem above. What is the performance of Neural Network? Is it better than linear regression models?
2. (3 pts) Apply Neural Network on a non-linear regression problem.

**Data**: Please download the training data Reg\_Train.txt and the testing data Reg\_Test.txt. Each example is described by 18 features in one line. The last column (the last value of each line) is the target value.

**Tasks**: learn a Neural Network from the training data, and use it to predict the target values in testing data. What is the **best** prediction result? (give the **average** and **standard deviation** of prediction **errors** on testing data)