Homework #4

RELEASE DATE: 1/2/2018

DUE DATE: 1/16/2018, BEFORE 14:00

QUESTIONS ABOUT HOMEWORK MATERIALS ARE WELCOMED ON THE FACEBOOK FORUM.

Unless granted by the instructor in advance, you must turn in a printed/written copy of your solutions (without the source code) for all problems.

For problems marked with (*), please follow the guidelines on the course website and upload your source code to designated places. You are encouraged to (but not required to) include a README to help the TAs check your source code. Any programming language/platform is allowed.

Any form of cheating, lying, or plagiarism will not be tolerated. Students can get zero scores and/or fail the class and/or be kicked out of school and/or receive other punishments for those kinds of misconducts.

Discussions on course materials and homework solutions are encouraged. But you should write the final solutions alone and understand them fully. Books, notes, and Internet resources can be consulted, but not copied from.

Since everyone needs to write the final solutions alone, there is absolutely no need to lend your homework solutions and/or source codes to your classmates at any time. In order to maximize the level of fairness in this class, lending and borrowing homework solutions are both regarded as dishonest behaviors and will be punished according to the honesty policy.

You should write your solutions in English or Chinese with the common math notations introduced in class or in the problems. We do not accept solutions written in any other languages.

> This homework set comes with 200 points and 20 bonus points. In general, every homework set would come with a full credit of 200 points, with some possible bonus points.

1. (60 points) Go register for the Coursera version of the second part of the class (`https://www.coursera.org/teach/ntumlone-algorithmicfoundations/`) and solve its homework 4. The registration should be totally free. Then, record the highest score that you get within up to 3 trials. Please print out a snapshot of your score as an evidence. (*Hint: The problems below are simple extensions of the Coursera problems.*)

2. (20 points) Write down the proof of your answer for Question 3 of Homework 4 on Coursera. If you choose "none of the other choices", please write down your derived update rule and prove it.

3. (20 points) Write down the proof of your answer for Question 4 of Homework 4 on Coursera. If you choose "none of the other choices", please write down a counter-example for each of the other choices.

4. (20 points) Write down the derivation steps of Question 5 of Homework 4 on Coursera.

5. (20 points) Write down the derivation steps of Question 11 of Homework 4 on Coursera.

6. (20 points) Write down the derivation steps of Question 12 of Homework 4 on Coursera.

7. (20 points, *) For Questions 14 and 15 of Homework 4 on Coursera, plot a figure that countains two curves, one being $E_{in}(g_\lambda)$ with respect to $\log_{10}\lambda$; the other being $E_{out}(g_\lambda)$ with respect to $\log_{10}\lambda$. Describe your findings. **Please print out the figure for grading.**

8. (20 points, *) For Questions 16 and 17 of Homework 4 on Coursera, plot a figure that countains two curves, one being $E_{train}(g_\lambda^-)$ with respect to $\log_{10}\lambda$; the other being $E_{val}(g_\lambda^-)$ with respect to $\log_{10}\lambda$. Describe your findings. **Please print out the figure for grading.**

Bonus: More about Leave One Out

9. (Bonus 20 points) In this problem, you will derive some interesting E_{loocv} results.

(a) Consider a binary classification algorithm $\mathcal{A}_{\text{majority}}$ that returns a constant classifier that always predicts the majority class (i.e., the class with more instances in the data set that it sees), and another classification algorithm $\mathcal{A}_{\text{minority}}$ that returns a constant classifier that always predicts the minority class (i.e., the class with less instances in the data set that it sees). For a data set with 1126 positive instances and 1126 negative instances. What is $E_{loocv}(\mathcal{A}_{\text{majority}})$ and $E_{loocv}(\mathcal{A}_{\text{minority}})$? Which algorithm would be chosen if we use E_{loocv} for algorithm selection? Prove your answer.

(b) Consider a regression algorithm $\mathcal{A}_{\text{average}}$ that returns a constant hypothesis that always predicts the average value within the data set that it sees. Prove that $E_{loocv}(\mathcal{A}_{\text{average}})$ is a scaled version of the variance of $\{y_n\}_{n=1}^{N}$.