Homework #3

RELEASE DATE: 12/11/2018

DUE DATE: 1/15/2019, BEFORE 14:00, **in R217**

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. (80 points) Go register for the Coursera version of the second part of the class (`https://www.coursera.org/learn/ntumlone-algorithmicfoundations/`) and solve its homework 3. 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. (40 points) When using SGD on the following error function and 'ignoring' some singular points that are not differentiable, prove or disprove that $err(\mathbf{w}) = \max(0, -y\mathbf{w}^T\mathbf{x})$ results in PLA.

3. (40 points) Write down the derivation steps of Question 17 of Homework 3 on Coursera.

4. (20 points, *) For Questions 19 and 20 of Homework 3 on Coursera, plot a figure that shows $E_{in}(\mathbf{w}_t)$ as a function of t for both the gradient descent version and the stochastic gradient descent version on the same figure. Describe your findings. **Please print out the figure for grading.**

5. (20 points, *) For Questions 19 and 20 of Homework 3 on Coursera, plot a figure that shows $E_{out}(\mathbf{w}_t)$ as a function of t for both the gradient descent version and the stochastic gradient descent version on the same figure. Describe your findings. **Please print out the figure for grading.**

Bonus: Smart 'Cheating'

6. (Bonus 20 points) For a regression problem, the root-mean-square-error (RMSE) of a hypothesis h on a test set $\{(\mathbf{x}_n, y_n)\}_{n=1}^N$ is defined as

$$\text{RMSE}(h) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - h(\mathbf{x}_n))^2}.$$

Please consider a case of knowing all the (test) $\mathbf{x}_n$, none of the y_n, but allowed to query RMSE(h) for some h.

For any given set of hypotheses $\{h_1, h_2, \cdots, h_K\}$. Assume that every RMSE(h_k) $= e_k$ has been queried and thus known. Also, assume that RMSE(h_0) $= e_0$ is known where $h_0(\mathbf{x}) = 0$ for any $\mathbf{x}$. Let $H(\mathbf{x}) = \sum_{k=1}^{K} w_k h_k(\mathbf{x})$. Derive a closed-form solution from

$$\min_{w_1, w_2, \cdots, w_K} \text{RMSE}\left(H\right)$$

from $e_0, e_1, \ldots, e_K$, $h_0, h_1, \ldots, h_K$, and $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$. (*The solution can be used to optimize the "test" performance that aggregates all h_k's together, which is a common trick used in data mining competitions to fight on the leaderboard.*)