Homework #3

RELEASE DATE: 12/20/2019

DUE DATE: 01/14/2020, BEFORE 14:00

QUESTIONS ABOUT HOMEWORK MATERIALS ARE WELCOMED ON THE FACEBOOK FORUM.

Please upload your solutions (without the source code) to Gradescope as instructed. 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/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. (20 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. (20 points) Write down the derivation steps of Question 9 of Homework 3 on Coursera.

4. (20 points) Write down the derivation steps of Question 16 of Homework 3 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 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.**

8. (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: More about Pseudo Inverse

9. (Bonus 20 points) In class, we derived that the linear regression solution weights must satisfy $\mathrm{X}^T\mathrm{X}\mathbf{w} = \mathrm{X}^T\mathbf{y}$. If $\mathrm{X}^T\mathrm{X}$ is not invertible, the solution $\mathbf{w}_{\text{lin}} = (\mathrm{X}^T\mathrm{X})^{-1}\mathrm{X}^T\mathbf{y}$ won't work. In this situation, there will be many solutions for $\mathbf{w}$ that minimize E_{in}. Here, you will derive one such

solution. Let ρ be the rank of X. Assume that the singular value decomposition (SVD) of X is $\mathrm{X} = \mathrm{U}\Gamma\mathrm{V}^T$, where $\mathrm{U} \in \mathbb{R}^{N\times\rho}$ satisfies $\mathrm{U}^T\mathrm{U} = \mathrm{I}_\rho$, $\mathrm{V} \in \mathbb{R}^{(d+1)\times\rho}$ satisfies $\mathrm{V}^T\mathrm{V} = \mathrm{I}_\rho$, and $\Gamma \in \mathbb{R}^{\rho\times\rho}$ is a positive diagonal matrix.

(a) Prove that $\mathbf{w}_{\text{lin}} = \mathrm{V}\Gamma^{-1}\mathrm{U}^T\mathbf{y}$ satisfies $\mathrm{X}^T\mathrm{X}\mathbf{w}_{\text{lin}} = \mathrm{X}^T\mathbf{y}$, and hence is a solution. (Note: $\mathrm{V}\Gamma^{-1}\mathrm{U}^T$ is actually one way to define the pseudo inverse $\mathrm{X}^\dagger$)

(b) Prove that for any solution that satisfies $\mathrm{X}^T\mathrm{X}\mathbf{w} = \mathrm{X}^T\mathbf{y}$, $\|\mathbf{w}_{\text{lin}}\| \le \|\mathbf{w}\|$. That is, the solution we have constructed is the "shortest" weight vector that minimizes E_{in}.