

CSE 417T: Homework 0

Due: 11am, January 21 (Tuesday), 2020

Notes:

- This is a special homework assignment for waitlisted students to complete. The instructor will check for correctness to make enrollment decisions. It will not be officially graded and will not factor in the final grades. However, the question will appear again at homework 1. The submissions to homework 1 will be graded by TA and will impact the final grades as specified in the syllabus.
- **Enrolled students do not need to submit this homework assignment.** The same question will appear in homework 1. Please submit your answers then.
- Please submit your homework via Gradescope. Please check the [submission instructions](#) for Gradescope provided on the course website. You must follow those instructions exactly.
- This special homework is due **by 11 AM on the due date. No late days are allowed.**
- The rule of academic integrity applies for this homework. If there is any suspicion of cheating (for example, answers are too similar to other students' submissions or to other resources), it will be reported to the university. The university maintains **permanent record** if students are found guilty.
- Please keep in mind the collaboration policy as specified in the course syllabus. If you discuss questions with others you **must** write their names on your submission, and if you use any outside resources you **must** reference them. **Do not look at each others' writeups, including code.**

Problems:

1. LFD Problem 1.3
 - The photocopy of the problem is in the next page in case you don't have access to the textbook yet.
2. Explain the reasons why you want/need to take this course in this semester. The enrollment priorities will be given to students who benefit the most by taking the course now. (This course is planned to be offered every semester).

Problem 1.3 Prove that the PLA eventually converges to a linear separator for separable data. The following steps will guide you through the proof. Let \mathbf{w}^* be an optimal set of weights (one which separates the data). The essential idea in this proof is to show that the PLA weights $\mathbf{w}(t)$ get “more aligned” with \mathbf{w}^* with every iteration. For simplicity, assume that $\mathbf{w}(0) = \mathbf{0}$.

- (a) Let $\rho = \min_{1 \leq n \leq N} y_n(\mathbf{w}^{*\top} \mathbf{x}_n)$. Show that $\rho > 0$.
- (b) Show that $\mathbf{w}^\top(t) \mathbf{w}^* \geq \mathbf{w}^\top(t-1) \mathbf{w}^* + \rho$, and conclude that $\mathbf{w}^\top(t) \mathbf{w}^* \geq t\rho$.
[Hint: Use induction.]
- (c) Show that $\|\mathbf{w}(t)\|^2 \leq \|\mathbf{w}(t-1)\|^2 + \|\mathbf{x}(t-1)\|^2$.
[Hint: $y(t-1) \cdot (\mathbf{w}^\top(t-1) \mathbf{x}(t-1)) \leq 0$ because $\mathbf{x}(t-1)$ was misclassified by $\mathbf{w}(t-1)$.]
- (d) Show by induction that $\|\mathbf{w}(t)\|^2 \leq tR^2$, where $R = \max_{1 \leq n \leq N} \|\mathbf{x}_n\|$.
- (e) Using (b) and (d), show that

$$\frac{\mathbf{w}^\top(t) \mathbf{w}^*}{\|\mathbf{w}(t)\|} \geq \sqrt{t} \cdot \frac{\rho}{R},$$

and hence prove that

$$t \leq \frac{R^2 \|\mathbf{w}^*\|^2}{\rho^2}.$$

$$\left[\text{Hint: } \frac{\mathbf{w}^\top(t) \mathbf{w}^*}{\|\mathbf{w}(t)\| \|\mathbf{w}^*\|} \leq 1. \text{ Why?} \right]$$

In practice, PLA converges more quickly than the bound $\frac{R^2 \|\mathbf{w}^*\|^2}{\rho^2}$ suggests. Nevertheless, because we do not know ρ in advance, we can't determine the number of iterations to convergence, which does pose a problem if the data is non-separable.