CS 541-B Artificial Intelligence: Mid-Term Exam

Instructor: Jie Shen

10/26/2022, 18:30 – 21:00 EST

Instructions:

- Open book exam, feel free to use any resource but no electronics;
- Discussion is not permitted;
- Always give your answer and explain it;
- 20 points per problem, totally 110 points (20 * 5 + 10).

0. Your name. (10 pts)

1. Discuss your understanding on algorithmic fairness and sketch your own approach to fortify an AI algorithm with fairness.

2. Give one example and one counterexample to each of the following statements, where X_1 and X_2 are random variables.

- $\mathrm{E}[X_1 X_2] = \mathrm{E}[X_1] \cdot \mathrm{E}[X_2]$
- $\mathrm{Var}[X_1 X_2] = \mathrm{Var}[X_1] \cdot \mathrm{Var}[X_2]$
- $\mathrm{Var}[X_1 + X_2] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2]$

3. Given an image, suppose that an expert will present you the correct label with probability at least 0.99, and he charges 100 dollars. On the other side, it is possible to distribute the image to a pool of non-experts, each of which charges you 1 dollar but the probability that he returns the correct label is as low as p ($p > 0.5$). Therefore, when going with the second option we have to hire many workers and take majority vote. Give a sufficient condition on p such that the following two are fulfilled simultaneously:

- the quality of the label from majority voting is as good as the one from the expert;
- it costs less to hire these non-experts.

Note: Phrase the condition as $f(p) \geq 0$ for some function f. You need to show what f is, but you do not need to calculate the value of p.

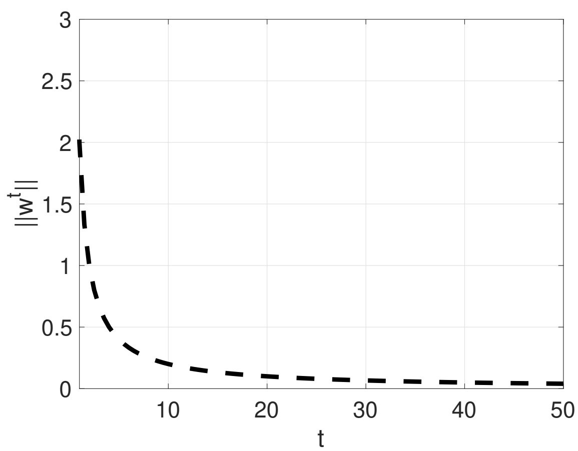
4. The classic PAC learning model of Valiant'84 made two fundamental assumptions: 1) the distributions of the training data and testing data are the same; and 2) all instances are labeled correctly. The goal of PAC learning is to find a hypothesis whose error rate, i.e. the probability that it misclassifies a new sample, is upper bounded by $\epsilon \in (0, 1)$. Give two respective examples to illustrate that if either assumption is violated, PAC learning becomes impossible.

5. Many machine learning problems boil down to solving the following optimization program:

$$\min_{\boldsymbol{w}} \; F(\boldsymbol{w}), \quad \mathrm{s.t.} \; \boldsymbol{w} \in \mathbb{R}^d. \tag{1}$$

Suppose that $d = 2$ and $F(\boldsymbol{w}) = \frac{1}{2}(w_1^2 + (w_1 + w_2)^2)$ where w_1 and w_2 are the first and second coordinates of $\boldsymbol{w}$ respectively.

- Calculate the gradient and the Hessian matrix of $F(\boldsymbol{w})$;
- Show that $F(\boldsymbol{w})$ is a strongly convex and smooth function, and calculate the tightest strong convexity parameter α and smoothness parameter L;
- Consider that we run gradient descent (GD) to find the global optimum of $F(\boldsymbol{w})$, starting from the initial iterate $\boldsymbol{w}^0 = (1, 1)$ and proceed with learning rate $\eta = 1/4$. Calculate the iterates $\boldsymbol{w}^1, \boldsymbol{w}^2, \boldsymbol{w}^3$.
- Suppose we are able to calculate more iterates $\boldsymbol{w}^4, \boldsymbol{w}^5, \ldots, \boldsymbol{w}^t, \ldots$ with $\eta = 1/4$, and we plot the curve "$\|\boldsymbol{w}^t\|_2$ v.s. t" as below. If we run GD with $\eta = 2/3$, what will the curve likely be? What about $\eta = 2$? Please plot them in the same figure and explain how you obtain these curves.



- Now consider minimizing the same function with stochastic GD, where the learning rate $\eta_t = 1/t$ at the t-iteration. Draw "$\|\boldsymbol{w}^t\|_2$ v.s. t" in the figure above.