CSC 588: Homework 3

Kwang-Sung Jun

Due: Mar 11, 2023, 11:59pm MST

Instructions for reporting your answers – not following these will result in deducted grading points:

- Show all work along with answers.
- Place your final answer into an 'answer box' that can be easily identified (unless the answer is a proof). Your final answer should not be based on symbols that you defined (okay to define it in the answer box). Your answer should be in the most explicit and simplified form.
- The grading will also be based on the clarity. There should be no room for interpretation about what you are writing. Otherwise, I will assume that they are wrong. The same goes with undefined symbols.
- Each subproblem's answer must be separated clearly from that of the other subproblems; e.g., do not mix up answers for subproblem 1.1 with that of subproblem 1.2. (This is due to the way gradescope works). Please write down the subproblem number like 4.(a), followed by the answer.
- If you get stuck, you can post your questions on Piazza. Try posing your questions to be generic, so you maintain the academic integrity while promoting discussion among your classmates.
- You are encouraged to discuss the homework questions with your classmates, but the discussions should only be at a high level, and you should write your solutions in your own words. For every question you have had discussions on, please mention explicitly whom you have discussed with; otherwise it may be counted as academic integrity violation.
- There will be no late days. Late homework result in zero credit. Not even one minute. It is a good idea to set yourself up your own deadline like one day before it is due.
- For detailed homework policies, please read the course syllabus carefully, available on the course website.
- Each subproblem (i.e., Problem X.Y) is worth 10 points unless noted otherwise.

Submission instruction:

- Submit homework via gradescope. You can hand-write your answers and scan them to make it a PDF, or type up your answers as pdf using LaTeX. If you use your phone camera, I recommend using TurboScan (smartphone app) or similar ones to avoid uploading a slanted image or showing the background. Make sure you rotate it correctly.
- Watch the video and follow the instruction for the submission: `https://youtu.be/KMPoby5g_nE`
- Report the code as part of the answer as texts. You should also submit the code to a separate submission entry in gradescope 'HW# - code' as well.

Problem 1

In this exercise, we run experiments on AdaBoost using a simple benchmark dataset `diabetes` in `openml.org`. You may use any programming languages you like. Some preparations:

1. Go to `https://www.openml.org/d/37` and download the dataset.
2. The last column of the dataset gives the classes of the examples - use $+1$ to denote class 'tested_positive' and -1 to denote class 'tested_negative'.
3. Choose a random subset of size 100 as the training set, and use the remaining 668 examples as the test set.

Answer the following questions:

(a) Define base hypothesis class $\mathcal{B} = \left\{\sigma \cdot (2I(x_i \leq t) - 1) : \sigma \in \{\pm 1\}, i \in \{1, \ldots, d\}, t \in \mathbb{R}\right\}$ as the set of bi-directional decision stumps. Let the weak learner $\mathcal{A}$ be: given a weighted dataset, return the classifier $h \in \mathcal{B}$ that has the smallest weighted error. Implement AdaBoost with $\mathcal{A}$, and run it for 3000 iterations. At time t, suppose the following cumulative voting classifier

$$H_t(x) = \operatorname{sign}(f_t(x)), \quad f_t(x) = \sum_{s=1}^{t} \alpha_s h_s(x)$$

is produced. Plot AdaBoost's learning curves: the training error of H_t, the test error of H_t, and the training exponential loss of f_t, as functions of iteration t. What do you see?

(b) Given voting classifier f_t, define its normalization as

$$\bar{f}_t(x) = \frac{f_t(x)}{\sum_{s=1}^{t} \alpha_s} = \frac{\sum_{s=1}^{t} \alpha_s h_s(x)}{\sum_{s=1}^{t} \alpha_s}$$

Now, given an example (x, y), define its normalized margin at time step t as $y\bar{f}_t(x)$. At iterations 3, 10, 30, 100, 300, 1000, 3000, plot histograms of normalized margins of training examples. Do you see any trend as t increases?

Problem 2

Show that for AdaBoost, at iteration t, the updated distribution D_{t+1} satisfies that

$$\sum_{i=1}^{m} D_{t+1}(i) I(h_t(x_i) \neq y_i) = \frac{1}{2}.$$

Intuitively, why is this formula reasonable?

Problem 3 (20pts)

Most of the problems we have seen in class so far are about classification. Consider instead a regression problem, where we have a distribution D over $\mathcal{X} \times \mathcal{Y}$, where the feature space $\mathcal{X} = \left\{x \in \mathbb{R}^d : \|x\|_\infty \leq R\right\}$ and the label space $\mathcal{Y} = [-Y, Y]$. Consider the hypothesis class $\mathcal{H} = \left\{h_w(x) := \langle w, x \rangle : \|w\|_1 \leq B\right\}$, and define the loss function to be the square loss $\ell_{\text{sq}}(\hat{y}, y) = (\hat{y} - y)^2$. For any predictor $h : \mathbb{R}^d \to \mathbb{R}$, define $L_D(h) = \mathbb{E}_{(x,y)\sim D} \ell_{\text{sq}}(h(x), y)$ its generalization loss. Now, given a set of examples $S = ((x_1, y_1), \ldots, (x_m, y_m))$ drawn iid from D, define the ERM $\hat{h} = \arg\min_{h \in \mathcal{F}} \mathbb{E}_S \ell_{\mathrm{sq}}(h(x), y)$. For any $\delta > 0$, can you show a tight upper bound on

$$L_D(\hat{h}) - \min_{h' \in \mathcal{H}} L_D(h')$$

that holds with probability $1 - \delta$? (You might want to use the contraction inequality of Rademacher complexity to solve this problem.)