

# CSC 665: Homework 3

Chicheng Zhang

October 31, 2019

Please complete the following set of problems. You must do the exercises completely on your own (no collaboration allowed). The exam is due **on Nov 14, 12:30pm, on Gradescope**. You are free to cite existing theorems from the textbooks and course notes.

## Problem 1

Consider the homogeneous, soft-margin SVM optimization problem:

$$\underset{w, \xi}{\text{minimize}} \quad \frac{\lambda}{2} \|w\|^2 + \sum_{i=1}^n \xi_i \tag{1}$$

$$\begin{aligned} \text{s. t.} \quad & y_i(\langle w, x_i \rangle) \geq 1 - \xi_i, & \forall i \in \{1, \dots, n\}, \\ & \xi_i \geq 0, & \forall i \in \{1, \dots, n\}. \end{aligned} \tag{2}$$

1. Introducing dual variables  $\alpha_i \geq 0$  for each constraint  $i$ ,  $i \in \{1, \dots, n\}$  and  $\beta_i \geq 0$  for each constraint  $i$ ,  $i \in \{1, \dots, n\}$ , compute the Lagrangian function  $L(w, \xi, \alpha, \beta)$ .
2. Derive the dual optimization problem.
3. Use the KKT condition to interpret: which of the training examples are “support vectors” that contribute to the SVM solution?

## Problem 2

Suppose we have  $k$  finite hypothesis classes  $\mathcal{H}_1, \dots, \mathcal{H}_k$ , and  $m$  training examples drawn iid from  $D$ . In addition we are given the promise that there exists  $i_0 \in \{1, \dots, k\}$  such that  $\min_{h \in \mathcal{H}_{i_0}} \text{err}(h, D) = 0$  (but we don't know the identity of  $i_0$ ); Can we design an algorithm that produces classifiers with generalization error  $O(\frac{\ln |\mathcal{H}_{i_0}|}{m})$  with high probability? Why or why not?

## Problem 3

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

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

Why is this a reasonable update?

## Problem 4

In this exercise, we conduct experiment with AdaBoost with a simple benchmark dataset named *ringnorm*.

1. Generate 100 training and 100 test examples from the following distribution  $D$  supported on  $\mathbb{R}^{10} \times \{\pm 1\}$ :  $\mathbb{P}_D(Y = -1) = \mathbb{P}_D(Y = +1) = \frac{1}{2}$ ,  $X|_{Y=+1} \sim N((0, \dots, 0), 4I)$ ;  $X|_{Y=-1} \sim N((\frac{2}{\sqrt{20}}, \dots, \frac{2}{\sqrt{20}}), I)$ .
2. Define base hypothesis class  $\mathcal{H} = \{\sigma \cdot (2I(x_i \leq t) - 1), \sigma \in \{\pm 1\}, i \in \{1, \dots, d\}, t \in \mathbb{R}\}$  as the set of bi-directional decision stumps. Let the weak learner  $\mathcal{B}$  be: given a weighted dataset, return the classifier  $h \in \mathcal{H}$  that has the smallest weighted error. Implement AdaBoost with  $\mathcal{B}$ , and run it for 300 iterations. At time  $t$ , suppose the following cumulative voting classifier

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

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 a function of iteration  $t$ . What do you see?

3. Given a voting classifier  $f_t$ , define its normalization as

$$\bar{f}_t(x) = \frac{f_t(x)}{\sum_{s=1}^t \alpha_s} = \frac{\sum_{s=1}^t \alpha_s h_s(x)}{\sum_{s=1}^t \alpha_s}. \quad (3)$$

Now, given an example  $(x, y)$ , define its normalized margin at timestep  $t$  as  $y\bar{f}_t(x)$ . At iterations  $t = 10, 30, 50, 100, 300$ , show histograms of normalized margins of training examples. Do you see any tendency at  $t$  increases?

## Problem 5 (No need to submit)

Show that AdaBoost produces large-margin voting classifiers under the  $\gamma$ -weak learning assumption. If at every iteration  $t$ ,  $\epsilon_t \leq \frac{1}{2} - \gamma$ , then after  $T$  rounds, the *margin error* of the output classifier will also decrease exponentially in  $T$ . Specifically, show:

$$\frac{1}{m} \sum_{i=1}^m \mathbf{1}(y_i \bar{f}_T(x_i) \leq \gamma) \leq \exp\{-\Omega(T\gamma^2)\}.$$

where  $\bar{f}_T$  is the normalized voting classifier defined as per Equation (3).