CS480/680: Introduction to Machine Learning

Homework 2

Due: 11:59 pm, October 16, 2018, submit on LEARN.

Include your name, student number and session!

Submit your writeup in pdf and all source code in a zip file (with proper documentation).
[Text in square brackets are hints that can be ignored.]

Exercise 1: Classifying MNIST Dataset (60 pts)

The MNIST dataset (available on course website) consists of 50k training images and 10k test images of 10 digits (0-9). Train one of the following algorithms using the provided training set:

- perceptron
- linear regression
- kNN
- logistic regression
- decision tree
- bagging
- boosting
- random forest
- support vector machines.

Apply your algorithm to the provided MNIST test set. You may use any existing package (tensorflow, scikitlearn, etc) or implement your own. Feel free to apply any pre-processing, data augmentation, parameter tuning, etc. For evaluation, please submit your code to Learn, along with a *clear description of what you did.* If you use external packages, give explicit instructions on how to obtain them and how to run your code.

Name your team and your prediction csv file by your student id and submit to kaggle. You can monitor your relative ranking on the leaderboard.

[See here for a nice tutorial on how to submit to kaggle.]

Exercise 2: Adaboost (40 pts)

Recall the update rules of Adaboost:

$$p_i^t = \frac{w_i^t}{\sum_{j=1}^n w_j^t}, \quad i = 1, \ldots, n \tag{1}$$

$$\epsilon_t = \epsilon_t(h_t) = \sum_{i=1}^n p_i^t \cdot |h_t(\mathbf{x}_i) - y_i| \tag{2}$$

$$\beta_t = \frac{\epsilon_t}{1 - \epsilon_t} \tag{3}$$

$$w_i^{t+1} = w_i^t \beta_t^{1-|h_t(\mathbf{x}_i) - y_i|}, \quad i = 1, \ldots, n. \tag{4}$$

Here we use i and t to index the training examples and iterations, respectively. The only superscript that is understood as power is in $\beta_t^{1-|h_t(\mathbf{x}_i)-y_i|}$. As usual, $y_i \in \{0, 1\}$ and **we also assume** $h_t(\mathbf{x}_i) \in \{0, 1\}$. Thus, $|h_t(\mathbf{x}_i) - y_i|$ is either 0 or 1 depending on whether or not the prediction $h_t(\mathbf{x}_i)$ agrees with the true label y_i.

What is the training error

$$\epsilon_{t+1}(h_t) = \sum_{i=1}^{n} p_i^{t+1} \cdot |h_t(\mathbf{x}_i) - y_i| \tag{5}$$

of the weak classifier h_t on the next round $t+1$? Justify your answer. You may assume $0 < \epsilon_t < 1$ so that all quantities are well-defined.

[Hint: Try to split the n training examples into two groups, according to whether or not $h_t(\mathbf{x}_i)$ agrees with the true label y_i. Note also that w^t and p^t are proportional to each other, and the sum of numbers in p^t is exactly 1. All given equations need to be used at least once. Try to reduce everything to w^t.]