Homework 4

Due May 01, 2020 by 11:59pm

Instructions: Upload your answers to the questions below to Canvas. Submit the answers to the questions in a PDF file and your code in a (single) separate file, including for the data competition exercise. Be sure to comment your code to indicate which lines of your code correspond to which question part. There are 3 study assignments and 2 exercises in this homework.

Reading Assignments

- Review Lecture 4.
- Review Computer Lab. 4 in *canvas.uw.edu/courses/1371621/pages/course-materials* .
- Read and study Sec. 14.5-14.6 in *The Element of Statistical Learning* .

1 Exercise 1

In this exercise, you will implement in **Python** a first version of your own linear support vector machine with the huberized hinge loss.

The linear support vector machine with the huberized or smoothed hinge loss writes as

$$\min_{\beta \in \mathbb{R}^d} F(\beta) := \lambda \|\beta\|_2^2 + \frac{\rho}{n} \sum_{i=1;\, y_i=+1}^{n} \ell_{hh}(y_i, x_i^T \beta) + \frac{1-\rho}{n} \sum_{i=1;\, y_i=-1}^{n} \ell_{hh}(y_i, x_i^T \beta) \,,$$

where $h = 0.5$ and

$$\ell_{\mathrm{hh}}(y,t) := \begin{cases} 0 & \text{if} \quad yt > 1+h \\ \frac{(1+h-yt)^2}{4h} & \text{if} \quad |1-yt| \leq h \\ 1-yt & \text{if} \quad yt < 1-h \end{cases} \tag{1}$$

You know now the fast gradient algorithm, so no need to recall it here.

- Consider the Vowel dataset from *The Elements of Statistical Learning*. Standardize the data, if you have not done so already.
- Pick a positive class "+1" with examples from a single class and a negative class "-1" with examples from two other classes than the positive class.
- Define a training set, a validation set, and a testing set, with roughly equal proportions of examples from the positive class and the negative class resp. within each set.

- Derive the formula of the gradient of the objective function. You may proceed case by case and then define a general formula.

- Write a function *mylinearsvm* that implements the fast gradient algorithm to train the linear support vector machine with the huberized hinge loss. The function takes as input the training data, the regularization parameter λ, the imbalance parameter ρ, and the optimization accuracy ε.

- Train your linear support vector machine with the huberized hinge loss on the the Vowel dataset for $\lambda = 1$ and $\rho = 1$. Report the misclassification error for these hyperparameter values. Report the sensitivity and the specificity for these hyperparameter values.[1]

- Train your linear support vector machine with the huberized hinge loss on the the Vowel dataset for $\lambda = 1$ and $\rho = 0.1, 0.2, \ldots, 0.9, 1.0$. Plot the misclassification error vs. the imbalance parameter ρ. Plot on the same figure yet with different colors the sensitivity and the specificity vs. the imbalance parameter ρ.

- Select the best value of the pair λ and ρ on a grid of values of your choice in terms of area under the sensitivity/(1-specificity) curve obtained from the validation set. Compute the sensitivity and the specificity on the testing set for the best values found for the pair λ and ρ.

- Optional. Repeat all the above with obvious changes with a negative class "-1" with examples from all classes except positive class.

Exercise 3

In this exercise, you are facing several challenging situations and you are asked to suggest effective strategies to tackle them, based on your knowledge in statistical machine learning and data science in general. Please answer the questions for each situation in a few lines. Please feel free to write equations or draw figures if it helps.

Linear regression insomnia. Alex is Bobbie's manager. Alex is a workaholic and Bobbie suffers from insomnia. Bobbie wants to design a deep network to predict the covid-19 pandemic evolution. At 3am, Bobbie receives a text from Alex asking her to log in her computer for a zoom meeting. Alex wants Bobbie to design, implement, run, and validate a linear regression model on a dataset. Bobbie has 2hrs to deliver the model and the predictions to Alex. Tyler adamantly forbids Frankie to use any off-the-shelf package because of IP concerns. Bobbie calls **you** through zoom and asks for your advice regarding the best optimization algorithm to use in this case, since you took DATA 558 at UW. Which optimization algorithm do you advise your friend Bobbie to use? Explain why.

[1] You may set the optimization accuracy to a value of your choice based on your experience from previous homework assignments.

The next big thing. Joey is a colleague of yours. Joey has been working with Tyler on a binary classification task for a particular real-world problem. Joey and Tyler have collected a dataset to work on this classification task. Joey sends you a zoom meeting request. While his kid is jumping on his lap, he claims he has just revolutionized machine learning. You ask him to tell you more. Joey says his method manages to predict perfectly on the test set. A striking feature of Joey's revolutionary approach is to use the test set (examples and labels) at training time. Joey says he did some research and apparently no one has ever applied this strategy in research papers or books. What do you think about Joey's revolutionary approach? How much would you invest in Joey's new company?