Homework Assignment 2

Zhaoyang Xu
Study Group: Yi Liu, JinZheng Li, Zhaoyang Xu

September 15, 2020

Contents

1 Introduction

Please document for example:

1. For the PAC-Bayesian aggregation method, what the bounds and the equations do you use?

   For PAC-Bayesian aggregation method, we trained each subset by SVM with RBF kernel and validated on the validation sets which were not used on training. These weak SVM classifiers were then applied for ρ-weithted majority vote. The computation of ρ were uesd by Theorem 3.32 on page 44 in lecture note.

$$\mathbb{E}_\rho[L(h)] \leq \frac{\mathbb{E}_\rho[\hat{L}^{val}(h,S)]}{1-\frac{\lambda}{2}} + \frac{\mathrm{KL}(\rho\|\pi) + \ln\frac{(n-r)+1}{\delta}}{\lambda(1-\frac{\lambda}{2})(n-r)}$$

   By following update rules, we can minimized bound.

$$\rho(h) = \frac{\pi(h)e^{-\lambda(n-r)(\hat{L}^{val}(h,S)-\hat{L}^{val}_{min})}}{\sum_{h'} e^{-\lambda(n-r)(\hat{L}^{val}(h',S)-\hat{L}^{val}_{min})}}$$

$$\lambda = \frac{2}{\sqrt{\frac{2n\mathbb{E}_\rho[\hat{L}(h,S)]}{(\mathrm{KL}(\rho\|\pi)+\ln\frac{n+1}{\delta})}+1}+1}$$

2. For the cross validation method, how do you set the parameters?

   The value of $\log_{10} C \in \{-3,-2,...,3\}$. For paremeter γ, $G(X_i) = \min_{(X_i,Y_i)\in S\bigcap Y_i\ Y_j} |X_i - X_j|$, $\gamma_J = 1/(2medianG^2)$, $\gamma \in \{10^{-4}\gamma_J, 10^{-2}\gamma_J, \gamma_J, 10^2\gamma_J, 10^4\gamma_J\}$.
   By grid search to select the γ and C, we used $C = 10$, $\gamma = 100\gamma_J \approx 0.71$

2 Implementation details

2.1 Dataset

Ex: What dataset do you use? How do you separate the training and testing data?

$ionosphere.data$ was loaded to implement and for convenience, the 'label' line 'g', 'b' were change into '0', '1'.
We random selected 200 samples as original training set, and other 151 samples as testing set.
For PAC-Bayesian Aggregation, the original training set were separated into training set with 35 samples and others as validation set. Also, we chose m subsets with 35 samples as training sets to construct the weak classifiers.

2.2 Cross Validation SVM

Ex: What tool do you use? How do you implement? (Please do not directly copy and paste the code in the report.)

By sklearn in python, we used function $GridSearchCV$ to select the parameters of model and also validated by $cv = 5$. Then the function SVC was used to construct model. After that, we can use $model.predict()$ and $accuracy\_score$ to get the loss directly.

2.3 PAC-Bayesian Aggregation

Ex: What tool do you use? How do you implement? (Please do not directly copy and paste the code in the report.)

For weak SVM classifiers, we still used function SVC, but this time we used $cross\_val\_score$ to get the loss of validation. Then the ρ and λ were iterated to get the optimal values. After that, we can calculate the loss. (But I'm not sure it is correct, my result and figure looks really odd)

2.4 Theoretical Bound

Ex: Which bound do you use?

I used the PAC-Bayes-kl Inequality,

$$\mathrm{kl}(\mathbb{E}_\rho[\hat{L}(h,S)]\|\mathbb{E}_\rho[L(h)]) \leq \frac{\mathrm{KL}(\rho\|\pi) + \ln\frac{2\sqrt{n}}{\delta}}{n}$$

We can get the bound on the expected loss of the classifier defined by ρ. It used the invertion of the kl divergence to get the bound on the expected loss.

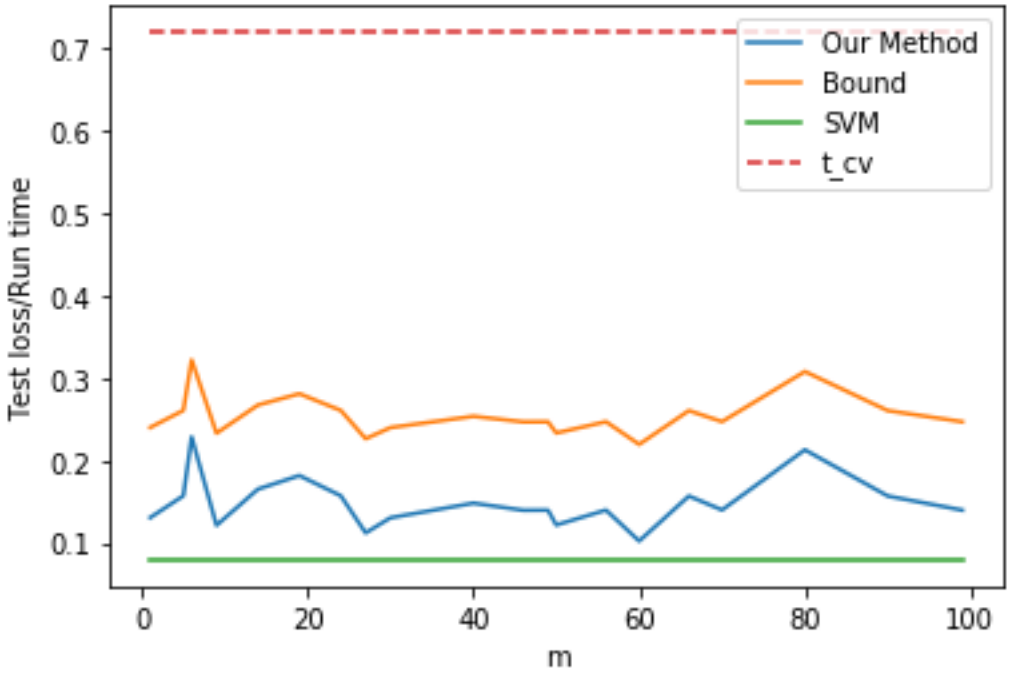
2.5 Runtime

Ex: How do you calculate the runtime?

The runtime of the CV-SVM includes cross validation and training on whole training set. So we can just $import\ \ time$ and put $start = time.clock()$ before the parameter selections and put $end = time.clock()$ in the end.
For PAC-Bayesian Aggregation, it should include the process of weak SVM's training, validation and the computation of ρ. So, we put a $time.clock$ after the iteration.

3 Figure

Please carefully annotate the figure and make some illustration.



SVM always shows the best. The bound of expect loss is the worst.