Machine Learning Homework1

RE6111024 葉嘉浤

Keywords—linear, classification, svm

I. INTRODUCTION (*HEADING 1*)

This homework aims to implement several different types of linear classifier and use existing SVM package,then compare the performance of each method on classification task.

II. METHODOLOGY OF USE

In this section,we introduce the methods used in this homework.

A. Linear classifier with update rule

The original linear classifier compute the dot product of weights and feature value,then add a constant as the prediction.The sign of prediction value is to verify if our prediction is correct.If the sign is positive,it represent our prediction is correct;otherwise,if the sign is negative,our prediction is wrong,then we will add the product of learning rate,feature value,and label value to original weights,the constant is added by label value.Then,we use the new weights and constant to predict this sample again,if the prediction is wrong,keep adjusting the weights and constant,until the prediction is correct or reach the maximum iteration times.

B. Linear classifier with least-squared manner

Least-squared method is a solution to help us find the weights that minimize the residual sum of square.With given data,we can get the closed form for weights to predict.

C. Voted perceptron and Average voted perceptron

Voted perceptron is an extension of linear classifier.While training,like original linear classifier,voted perceptron also calculate the dot product of weights and feature value,then add a constant as prediction.If the prediction is wrong,we will save the current weights and the number of examples that this weights got correct,then adjust the weights and constant.While testing,we calculate the prediction from all saved weights,multiply each prediction by the number it got correct,and take the sum over all predictions.

D. Large margin classifier

Large margin classifier aims to find the large margin to classify samples.During training,we collect the sampled observations,then use dot product of weights and features as prediction.When the prediciton is wrong,we adjust the weights,then we calculate the distance of all sampled observations to the line with original weights and new weights.if the total distance obtained from new weights is less than that obtatined from original weights,we will adopt new weights.

E. Soft large margin classifiers

Based on large margin classifier,soft large margin classifiers allow misclassifications,so we need to minimize the

misclassification error.We add a constraint,called slack variable,while calculating distance.To minimizing the error,we set a constant C to control the trade-off between margin maximization and loss minimization.

III. EXPERIMENT

In this part,we will do some experiment to evaluate performance of each model on classification task.

A. Experiment settings

In crx dataset,we remove the na values first,convert the categorical features to the numeric value with label encoder,then we normalize the data;in data dataset,since the feature values are all numeric,we normalize the data.
We split data into 80% training data and 20% testing data,learning rate is 0.1,maximum iteration is 10.Each experiment is conducted 5 times expect linear classifier with update rule and average voted perceptron.
In the first experiment,we compare all performance of implementations and existing SVM,the constant in soft large margin classifier is set as 1;in the second experiment,we try to find which constant value in soft large margin classifier will lead to the best performance.We set the constant as 1,3,5,7,9.

B. Performance on classification task

Dataset	Method	Accuracy
crx	Linear classifier (update rule)	<u>0.604±0.0</u>
	Linear classifier (least-squared manner)	0.489
	Voted perceptron	0.60±0.0
	Average Voted perceptron	0.41
	Large margin classifier	0.413±0.02
	Soft large margin classifier	0.454±0.08
	Existing SVM	0.88
data	Linear classifier (update rule)	0.404±0.0
	Linear classifier (least-squared manner)	0.377
	Voted perceptron	0.544±0.0
	Average Voted perceptron	0.684
	Large margin classifier	<u>0.907±0.01</u>
	Soft large margin classifier	0.905±0.01
	Existing SVM	0.96

TABLE I. . ACCURACY(MEAN±STD) BETWEEN DIFFERENT METHOD ON CALSSIFICATION.

Table 1. shows the accuracy of classification between each methods.Red color means the best model in a data,and the underline represents the best model which we implement.Existing SVM reach the highest accuracy on two datasets.Compare within our implementations,Linear classifier with update rule gets the highest accuracy on crx dataset,The second place is voted perceptron,which the accuracy is 0.6.The worst model is average voted perceptron.In data dataset,two types of large margin classifier

are the most appropriate models for this data,which their accuracy both greater than 0.9.Linear classifier with least-squared manner is the least suitable model for this data,which accuracy score is only 0.37.

Dataset	Slack variable	Accuracy
crx	1	0.425
	3	0.425
	5	0.418
	7	0.455
	9	0.373
data	1	0.912
	3	0.920
	5	0.920
	7	0.895
	9	0.895

TABLE II. . ACCURACY(MEAN±STD) BETWEEN DIFFERENT CONSTANT IN SOFT LARGE MARGIN CLASSIFIER

Table 2. shows the performance of soft large margin classifier with different constant.In crx dataset,the highest accuracy reachs when constant equals to 7,while in data dataset,model gets the best accuracy with constant equals to 3 or 5.

IV. CONCLUSION

In this work,we compare the performance between methods we implemented and use the existing SVM package.We try to find a constant in soft large margin classifier that will lead to the best performance.

There are some factors will influence the performance of linear classifiers,such as initialized weights ,the order of observations getting into classifiers,the setting of parameters.Since we do not set a seed to fix the initialized weights,the weights will lead to the accuracy.Take the original linear classifier for example,the performance will vary from 0.39 to 0.6 due to different initialized weights.The order of observations also influence the model,because each time we update weights,we consider the feature values of a observation,and it decides the direction and scale that weights are updated.Each time we train the model,we shuffle the data,and train the model again.We can see that the performance is different.The different value of parameters also influence the performance of a model.Take soft large margin classifier for example,the constant means how large we can tolerate from misclassification.We can see the performance is different if the constant is changed.There are still more factors that impact the model performance,we just discuss a little.The whole experiment procedure are available on https://github.com/chyeh1126/MachineLearning-2022/blob/main/hw1/ML_HW1.ipynb