Assignment-2

Machine Learning

Wen-Hai Tseng
Nation Cheng Kung University
w13281328@gmail.com

***Abstract*—The objective of this paper is to implement Kernel methods for various machine learning models and explore variations with Random Forest without utilizing packages, aiming to enhance students' understanding of models and applications.**

I. Introduction (*Heading 1*)

The paper aims to address two practical issues. The first problem involves implementing Kernel methods for diverse machine learning models and integrating them using ensemble learning. The second problem pertains to the observation that Random Forest models may not achieve optimal predictive performance. To address this, we propose replacing the trees with Multilayer Perceptron to enhance predictive capabilities.

II. Method

A. *Q1: kernel methods in different types of machine learnin g models, and then combining these kernel-enhanced mo dels in an ensemble learning framework.*

In this paper, we incorporate two models, linear classification and KNN, into Kernel methods. Subsequently, we perform stacking on the results of these two models, followed by learning using a Decision Tree.

$$k(x, y) = e^{-\sigma \|x-y\|^2} \quad (1)$$

B. *Q2: design an ensemble learning framework where multi ple such MLPs are trained on different subsets of the dat a or features and then aggregated to form a more robust model.*

I adopt the approach of replacing the trees in the Random Forest with 2-layer Multilayer Perceptrons (MLPs), with each MLP trained on a different subset of features.

III. Experments

A. *Linear Classifier with RBF Kernel methods*

```
The Val Accuracy of our classifier is: 0.542463317029791

The Test Accuracy of our classifier is: 0.5313472654513117
```

Fig. 1. The accuracy of the Linear Classifier model's predicti ons on the Validation and Test datasets.

B. *K Nearest Neighbors with RBF Kernel methods*

```
The Val Accuracy of our Knn is: 0.5264562027567808

The Test Accuracy of our Knn is: 0.5224544241885283
```

Fig. 2. The accuracy of the K Nearest Neighbors model's predictions on the Validation and Test datasets

C. *Decision Tree with Stacking*

```
The Val Accuracy of our new feature is: 0.5429079590929302

The Val Accuracy of our classifier old feature is: 0.542463317029791

The Val Accuracy of our knn old feature is: 0.5264562027567808
```

Fig. 3. The accuracy of the Decision Tree vs Linear Classifier vs K Nearest Neighbors predictions on the Validation

We conducted experiments comparing the performance of a Linear Classifier model using RBF Kernel methods with that of K Nearest Neighbors using RBF Kernel methods. The results indicated that the Linear Classifier model outperformed K Nearest Neighbors. Further, stacking the results of both models with a Decision Tree yielded improvements beyond the performance of the Linear Classifier alone..

D. *Deep Random Forest*

```
The Val Accuracy of our Knn is: 0.5015562472209871

The Test Accuracy of our Knn is: 0.5024455313472654
```

Fig. 6. The accuracy of Deep Random Forest's predictions on the Validation and Test datasets.

We replaced the trees inside the Random Forest with 2-layer Multilayer Perceptrons; however, the results did not indicate an improvement in performance. This could be attributed to the fact that the dataset may not possess sufficient complexity to benefit from such a substitution.

References

[1] https://devpress.csdn.net/python/62f62634c6770329307fc080.html
[2] https://zhuanlan.zhihu.com/p/23966698

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.