MSBD6000B Deep Learning Project 1

QIU Feng (fqiuaa, 20398324)

1. Data pre-processing

By comparing Scale() function, MinMaxscale() function and Normalize() function respect to SVM model, the Scale() function has the highest accuracy.

Therefore, for data pre-processing, I use the Scale() function of sklearn for standardization.

1. Experiment deployment

In this project, I am using SVM, Logistic Regression and KNN to do training and cross validation. For SVM model, I use the SVC package of sklearn in python. For each of the classifier, I did 10-fold cross validation. The performance metric is accuracy.

1. Classifiers and experiment results

For all parameters, like SVM C and gamma, Logistic Regression C and KNN k-neighbors, optimizers are selected by cross validation and grid search. I only give the best results for each parameter selection.

* 1. SVM

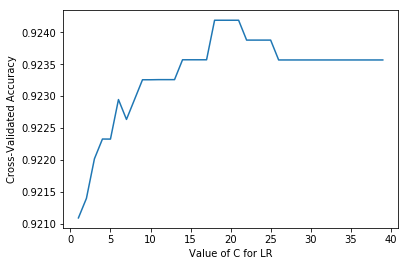
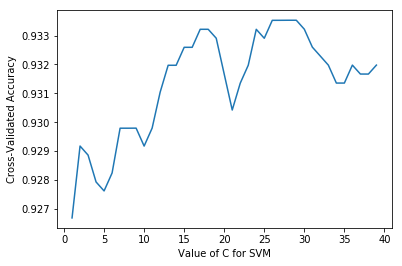
To find the best parameters in SVM, I use grid search and cross validation to find the best values of gamma and C. For SVC package in SVM, I am using rbf kernel, 28 for C value respect to the picture below and 0.01 for gamma value.

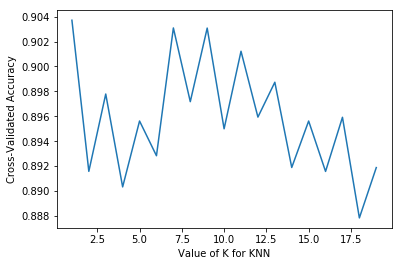
* 1. Logistic Regression

For LR, I use cross validation. The best C value from the picture below is 18.

* 1. KNN

For KNN, I use cross validation. The best k value from the picture below is 7.





* 1. Performance

|  |  |  |  |
| --- | --- | --- | --- |
| 10-fold accuracy | SVM | Logistic Regression | Knn |
| mean | 0.9335 | 0.9242 | 0.903 |
| std | 0.0187 | 0.0148 | 0.0139 |

1. Conclusion

In terms of accuracy and stability, SVM model is the best. Therefore, Ichoose SVM for prediction.