# README

2023-04-10

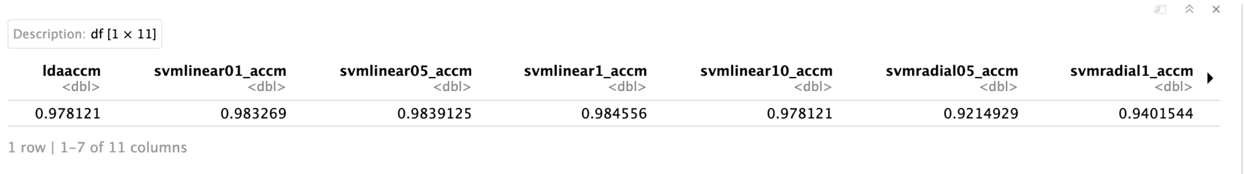
**DATA FILE** Two tab-delimited text files training_data.txt and test_data.txt are provided.The training data (labeled activity information included) is used to construct and test the ML algorithms. After that, applying algorithm to the test data (containing only feature information) and predict the activity corresponding to each time window.

**Task** *1. Build a binary classifier to classify the activity of each time window into static (0) and dynamic (1). For this task, consider postural transitions as static (0).*

a.Algorithm used in binary classifier: logistic regression, linear discriminant analysis, SVM with linear kernel, SVM with radial kernel

b.The baseline algorithms used in the binary classifier is the logistic regression. Without selecting the variables, the accuracy is extremely high and attain 0.998713. The problem might be the overfitting and when it throw into the test file, the accuracy might decrease.

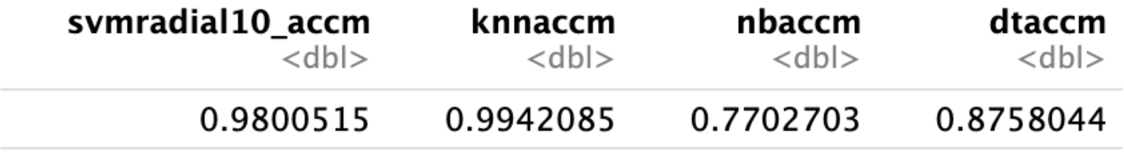
c.The final algorithms used in the binary classifier is the SVM with radial kernel with cost 5. When use SVM with different kernel to test the train data, the accuracy is very close to 1. For the SVM with linear kernel, the accuracy is always 1; For the SVM with radial kernel, it only have one error. Thus, I first train SVM with linear kernel with different cost in test file and throw it into github to see the result. It didn't show 100% accuracy. So I adjust the kernel to be radial and test with different cost. It turns out the SVM with radial kernel with cost 5 will accurately predict the test file with 100% accuracy.



Description: df [1 × 11]

| ldaaccm<br><dbl> | svmlinear01_accm<br><dbl> | svmlinear05_accm<br><dbl> | svmlinear1_accm<br><dbl> | svmlinear10_accm<br><dbl> | svmradial05_accm<br><dbl> | svmradial1_accm<br><dbl> |
|---|---|---|---|---|---|---|
| 0.978121 | 0.983269 | 0.9839125 | 0.984556 | 0.978121 | 0.9214929 | 0.9401544 |

1 row | 1–7 of 11 columns



| svmradial10_accm<br><dbl> | knnaccm<br><dbl> | nbaccm<br><dbl> | dtaccm<br><dbl> |
|---|---|---|---|
| 0.9800515 | 0.9942085 | 0.7702703 | 0.8758044 |

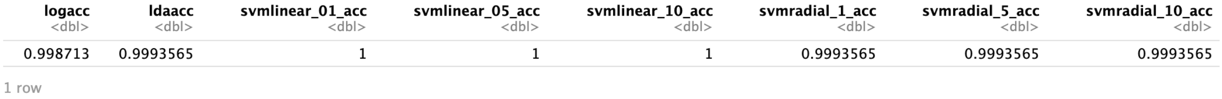
---

*2.Build a refined multi-class classifier to classify walking (1), walking_upstairs (2), walking_downstairs (3), sitting (4), standing (5), lying (6), and static postural transition (7)*

a.Algorithm used in multi-class classifier: knn method, naivebayes, decision tree, SVM with linear kernel, SVM with radial kernel, and linear discriminant analysis.

b.The baseline algorithms used in the multi class classifier is the knn method. By setting different k in to model, I found that when k = 1, the accuracy will attain the maximum. However, when compare to the result I submit to the leaderboard and got 0.962, it shows large difference. Thus, this high accuracy might occur due to overfitting.

c.The final algorithms used in the multi class classifier is the svm with linear kernel. By setting different cost, I found that the when cost = 1, the accuracy will attain the maximum. Also, I change different seed and validate this result. That might explain the tradeoff between variance and bias.



| logacc<br>&lt;dbl&gt; | ldaacc<br>&lt;dbl&gt; | svmlinear_01_acc<br>&lt;dbl&gt; | svmlinear_05_acc<br>&lt;dbl&gt; | svmlinear_10_acc<br>&lt;dbl&gt; | svmradial_1_acc<br>&lt;dbl&gt; | svmradial_5_acc<br>&lt;dbl&gt; | svmradial_10_acc<br>&lt;dbl&gt; |
|---:|---:|---:|---:|---:|---:|---:|---:|
| 0.998713 | 0.9993565 | 1 | 1 | 1 | 0.9993565 | 0.9993565 | 0.9993565 |

1 row

The data showing in the leaderboard

a.Binary classifier: 0.999(SVM linear with cost 10) -> 1(SVM radial with cost 5);

b.multiclass classifier: 0.962(SVM linear with cost 10) -> 0.965(SVM linear with cost 1);

During my training, I didn't penalize overfitting, that might explain the reason why knn method show pretty good result in my train_set but behave poorly when fit into test data. To further improve the accuracy, I need think about use lasso, or ridge regression to further reduce the influence of overfitting. Also, I haven't tried some algorithms that mentioned in the class, such as neural network, that might be further improve the classifier performance.