Exercise report 3 for ML

季林成 2017012775 经71

20191009

1   linear

The experiment is carried out with polynomial kernal and linear kernel. The SVM is trained and tested with different size of datasets.

When testset-1 is applied, the result appears below:

linear_trainset_1_testset_1.log
train error: 0.7844
cross-validation error: 0.6125
test error: 0.5430
linear_trainset_1_testset_2.log
train error: 0.8063
cross-validation error: 0.5250
test error: 0.5125

When testset-2 is applied, the result appears below:

linear_trainset_2_testset_1.log
train error: 0.5851
cross-validation error: 0.5328
test error: 0.5140
linear_trainset_2_testset_2.log
train error: 0.5959
cross-validation error: 0.5328
test error: 0.5346

2   poly

When testset-1 is applied, the result appears below:

<div align="center">

poly_trainset_1_testset_1.log

train error: 0.6438

cross-validation error: 0.7000

test error: 0.5520

poly_trainset_1_testset_2.log

train error: 0.6625

cross-validation error: 0.6500

test error: 0.5513

</div>

When testset-2 is applied, the result appears below:

<div align="center">

poly_trainset_2_testset_1.log

train error: 0.5702

cross-validation error: 0.5656

test error: 0.5460

poly_trainset_2_testset_2.log

train error: 0.5694

cross-validation error: 0.5508

test error: 0.5506

</div>

3   conclusion

The conclusions comes from comparisons between error rate of same trainset and different testsets(such as trainset-1&testset-1 v.s. trainset-1&testset-2), or different trainsets and same testset(such as trainset-1&testset-2 v.s. trainset-2&testset-2). The mean error rate of 10 training and testing shows the SVM performs better when the testset is larger, which is intuitive as the more the samples are, the more difficult to classify them. However, the effect of changing trainsets is not obvious from the results.