

# 10601 Assignment Report

## Milestone 3

Andrew ID: jiacongh, shenwu

Finished Date: 06/11/2013

Did you receive any help whatsoever from anyone in solving this assignment?

No

Did you give any help whatsoever to anyone in solving this assignment?

No

### 1. Tuning the classifier learner.

In this task, I tuned parameters in the SMO, LWL and logitBoost classifiers. I ran those classifiers on some training data and choose the best performance parameter set as the parameter I choose for the next phase.

Here are some parameter setting results in the credit-a dataset for the SMO classifier.

Parameter: C(epsilon=-12;tolerance parameter=0.001)-

0.1	0.2236
0.2	0.2232
0.4	0.2231
0.5	0.2231
0.6	0.2231
0.8	0.223
1.0	0.229
1.2	0.2228

Parameter :epsilon 1.0\*En (C=0.5;tolerance =0.001)

-1	0.229
-6	0.2238
-8	0.223
-10	0.2231
-12	0.2231
-14	0.2231
-16	0.2231

Parameter :tolerance parameter (C=0.5,epsilon=-12)

1	0.2675
0.8	0.2228

0.7	0.2229
0.6	0.2229
<b>0.5</b>	<b>0.2228</b>
0.4	0.2229
0.3	0.2231
0.1	0.2231
0.01	0.2231
0.001	0.2231
0.0001	0.2231

In the tuning phase, we found the SMO classifier is not very sensitive to the parameter, the result usually doesn't change for more than 1% when we vary the parameter. The LWL classifier is more sensitive. When changing the weightKernal parameter from 1 to 2, the error rate result in some datasets vary for over 5%. The logitBoost classifier is not sensitive to some parameter but quite sensitive to other parameter. When changing weightThreshold from 100 to 1000, the error rate change no more than 0.1%. However, the result varies obviously when changing the value of shrinkage.

The table up is the parameter tuning table for SMO classifier. The highlight parameters are parameters we choosed to be the final result for the next phase. Other two classifiers follow the same rule.

Finally, the parameter set of SMO is C=1.2 L=0.5 P=1.0E-8;

The parameter set of LWL is U=0 k=1;

The parameter set of logitBoost is P=100 F=0 R=1 H=1.5;

## 2. Evaluating Classifiers

In this section, we used the selected classifiers to run all test data. And compare the classifier which has the best performance with the untuned classifier. Then we will find it is more robust. When comparing with all three tuned classifiers, we found the logitBoost classifier is the one best robust. Then we compare the result of it with the untuned logitBoost classifier.

filename/classifier	LogitBoost(new)	LogitBoost
anneal	<b>0.02303</b>	0.02703
audiology	<b>0.173333</b>	0.18666
autos	<b>0.411764</b>	0.47058
balance-scale	<b>0.101942</b>	0.31553
breast-cancer	0.351063	<b>0.265957</b>

colic	0.223140	0.19834
credit-a	<b>0.218070</b>	0.219298
diabetes	0.308300	0.25296
glass	0.563380	<b>0.338028</b>
heart-c	0.67	<b>0.16</b>
hepatitis	<b>0.078431</b>	0.098039
hypothyroid	<b>0.0040161</b>	0.00723

We can find from the table that when tuning parameters in logitBoost, new LogitBoost classifier get lower error rate which means the tuned LogitBoost classifier is more robust.