CS760 HW4 Report

Yifan Mei

1. I made stratified 10-fold cross validation on the "chess-KingRookVKingPawn.arff" data set to compare naive Bayes and TAN. The way I do so is firstly build cross validation groups 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 with data number 320, 320, 320, 320, 320, 320, 319, 319, 319, 319 to make sure the ratio of "won" and "nowin" are nearly the same for each CV group. Then I make a loop to test both TAN and Naïve Bayes methods given the same CV groups.

   I list the CV group number, each group's containing data number, the number of the test-set examples that were correctly classified given TAN method, the number of the test-set examples that were correctly classified given Naïve Bayes method, the accuracy of TAN method, and the accuracy of Naïve Bayes.

CV group number	Test number	The number of the test-set examples that were correctly classified given TAN method	The number of the test-set examples that were correctly classified given Naïve Bayes method	Accuracy of TAN method	Accuracy of NB method
1	320	298	277	0.93125	0.865625
2	320	302	280	0.94375	0.875
3	320	293	285	0.915625	0.890625
4	320	302	288	0.94375	0.9
5	320	300	274	0.9375	0.85625
6	320	299	279	0.934375	0.871875
7	319	297	284	0.931034483	0.890282132
8	319	289	271	0.905956113	0.849529781
9	319	292	287	0.915360502	0.89968652
10	319	294	283	0.921630094	0.887147335

2. I made a paired t test for the accuracy of TAN method and the accuracy of Naïve Bayes method.

t = 7.47403437
p-value = $3.79486156*10^{-5}$, which is less than 0.0001.

Appendix

The result of paired t test given R code:

```
> t.test(dif)

        One Sample t-test

data:  dif
t = 7.47403437, df = 9, p-value = 3.79486156e-05
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.0344628338172 0.0643792509828
sample estimates:
   mean of x
0.0494210424
```