Florida Atlantic University (FAU)

Assignment 3

Project objective: Modeling assignment: Using meta learning schemes with a strong and a weak learner for classification.

Written By: Kevin Tudor

Table of Contents

Requirements

This last section of the project will allow you to evaluate the benefits of using meta learners on this data set:

Using the same methods as above (10-fold cross-validation on the fit data), determine the preferred (optimal cost ratio) model, and evaluate the models using the test dataset for:

1. Cost sensitive classifier combined with bagging and J48
2. Cost sensitive classifier combined with bagging and Decision Stump
3. Cost sensitive classifier combined with boosting (AdaBoostM1) and J48
4. Cost sensitive classifier combined with boosting (AdaBoostM1) and Decision Stump

Use the default settings for the meta learners (bagging, boosting,) and the learner (J48, decision stump) but vary the cost ratio in the same way as in Part 4 of Assignment II.

Provide the command lines you used.

How do the results of each classifier compare to the cost-sensitive tree obtained in Part 4 of Assignment II? Comment.

Now, set the number of iterations of each meta learner to 25 and repeat the experiments. Don't forget to analyze the results.

Iterations (meta learner) = 10

Default Cost ratio (1 – 1)

10-fold -> Ratio (1-1)	Type I %	Type II %
bagging and J48	8.8%	27.3%
bagging and Decision Stump	11.7%	20.0%
boosting (AdaBoostM1) and J48	7.3%	21.8%
boosting (AdaBoostM1) and Decision Stump	12.4%	32.7%

User Test Set -> Ratio (1-1)	Type I %	Type II %
bagging and J48	5.9%	21.4%
bagging and Decision Stump	11.8%	14.3%
boosting (AdaBoostM1) and J48	7.4%	32.1%
boosting (AdaBoostM1) and Decision Stump	16.2%	10.7%

Custom Cost ratio (1 – 0.5)

10-fold -> Ratio (1-0.5)	Type I %	Type II %
bagging and J48	5.1%	36.4%
bagging and Decision Stump	6.6%	38.2%
boosting (AdaBoostM1) and J48	7.3%	21.8%
boosting (AdaBoostM1) and Decision Stump	10.2%	32.7%

User Test Set -> Ratio (1-0.5)	Type I %	Type II %
bagging and J48	5.9%	28.6%
bagging and Decision Stump	11.8%	14.3%
boosting (AdaBoostM1) and J48	5.9%	32.1%
boosting (AdaBoostM1) and Decision Stump	14.7%	14.3%

Custom Cost ratio (1 – 1.5)

10-fold -> Ratio (1-1.5)	Type I %	Type II %
bagging and J48	9.5%	23.6%
bagging and Decision Stump	16.1%	16.4%
boosting (AdaBoostM1) and J48	5.1%	23.6%
boosting (AdaBoostM1) and Decision Stump	13.9%	27.3%

User Test Set -> Ratio (1-1.5)	Type I %	Type II %
bagging and J48	10.3%	17.9%
bagging and Decision Stump	22.1%	7.1%
boosting (AdaBoostM1) and J48	5.9%	35.7%
boosting (AdaBoostM1) and Decision Stump	17.6%	17.9%

Summary

How do the results of each classifier compare to the cost-sensitive tree obtained in Part 4 of Assignment II?

The results of each classifier in this assignment do not linearly compare with the cost-sensitive tree obtained in Part 4 of Assignment II. In assignment II, when the cost of Type II is increased the Type I misclassifications go up while the Type II misclassifications go down. When the cost of Type II is decreased the Type I misclassifications go down while the Type II misclassifications go up. In the classifiers from this assignment, adjusting the Type II ratio does not do the same. In each ratio instance the Type I error is *usually* equal or less than the type II error (on average). **The best *overall* classifier method is bagging and decision stump with the user supplied test data set.**

Iterations (meta learner) = 25

Default Cost ratio (1 – 1)

10-fold -> Ratio (1-1)	Type I %	Type II %
bagging and J48	8.8%	27.3%
bagging and Decision Stump	11.7%	18.2%
boosting (AdaBoostM1) and J48	8.0%	20.0%
boosting (AdaBoostM1) and Decision Stump	10.2%	32.7%

User Test Set -> Ratio (1-1)	Type I %	Type II %
bagging and J48	7.4%	21.4%
bagging and Decision Stump	14.7%	10.7%
boosting (AdaBoostM1) and J48	5.9%	25.0%
boosting (AdaBoostM1) and Decision Stump	10.3%	21.4%

Custom Cost ratio (1 – 0.5)

10-fold -> Ratio (1-0.5)	Type I %	Type II %
bagging and J48	4.4%	30.9%
bagging and Decision Stump	8.0%	36.4%
boosting (AdaBoostM1) and J48	8.0%	21.8%
boosting (AdaBoostM1) and Decision Stump	8.8%	32.7%

User Test Set -> Ratio (1-0.5)	Type I %	Type II %
bagging and J48	5.9%	28.6%
bagging and Decision Stump	11.8%	14.3%
boosting (AdaBoostM1) and J48	7.4%	39.3%
boosting (AdaBoostM1) and Decision Stump	11.8%	39.3%

Custom Cost ratio (1 – 1.5)

10-fold -> Ratio (1-1.5)	Type I %	Type II %
bagging and J48	9.5%	20.0%
bagging and Decision Stump	16.1%	14.5%
boosting (AdaBoostM1) and J48	5.8%	20.0%
boosting (AdaBoostM1) and Decision Stump	8.8%	25.5%

User Test Set -> Ratio (1-1.5)	Type I %	Type II %
bagging and J48	8.8%	17.9%
bagging and Decision Stump	19.1%	7.1%
boosting (AdaBoostM1) and J48	5.9%	35.7%
boosting (AdaBoostM1) and Decision Stump	10.3%	25.0%

Summary

Analyze the results (comparing 10 iterations to 25).

With 25 iterations the optimal models are the same as the 10 iteration classifications. (**Bagging and decision stump with the user supplied test data set).** However, it seems that with 25 iterations the optimal models are more balanced or have a lower type I /II error.

Optimal Models

10 - iterations -> Bagging and Decision Stump	Type I %	Type II %
10-fold (1-1 ratio)	N/A	N/A
User Supplied (1-1 ratio)	11.8%	14.3%
10-fold (1-0.5 ratio)	N/A	N/A
User Supplied (1-0.5 ratio)	11.8%	14.3%
10-fold (1-1.5 ratio)	16.1%	16.4%
User Supplied (1-1.5 ratio)	22.1%	7.1%

25 - iterations -> Bagging and Decision Stump	Type I %	Type II %
10-fold (1-1 ratio)	N/A	N/A
User Supplied (1-1 ratio)	14.7%	10.7%
10-fold (1-0.5 ratio)	N/A	N/A
User Supplied (1-0.5 ratio)	11.8%	14.3%
10-fold (1-1.5 ratio)	16.1%	14.5%
User Supplied (1-1.5 ratio)	19.1%	7.1%

Raw Weka Data

Due to sheer amount of data, the Raw Weka Info will be **omitted**. (Would be 100+ pages to include each model)