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

Due date: March 16, '19

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

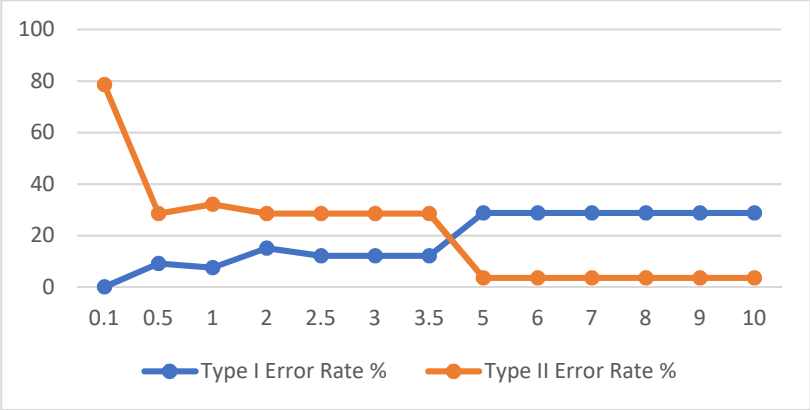
Using the same methods as above (10-fold cross-validation on the fit data), determine the preferred (optimal cost ratio) model, and evaluate the models using the test dataset for:

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.

Part 4 of Assignment II cost sensitive classifier combined with J48

Cost of Type I Error	Cost of Type II Error	Evaluate on train set					Re-evaluate on test set				
		Type I	Type 2	Type I Error Rate %	Type II Error Rate %	Misclassification Rates %	Type I	Type 2	Type I Error Rate %	Type II Error Rate %	Misclassification Rates %
1	0.1	5	38	3.76	69.09	22.87	0	22	0	78.57	23.40
1	0.5	7	23	5.26	41.82	15.95	6	8	9.09	28.57	14.89
1	1	12	11	9.02	20.00	12.23	5	9	7.57	32.14	14.89
1	2	17	6	12.78	10.91	12.23	10	8	15.15	28.57	19.14
1	2.5	17	8	12.78	14.55	13.29	8	8	12.12	28.57	17.02
1	3	18	8	13.53	14.55	13.82	8	8	12.12	28.57	17.02
1	3.5	18	8	13.53	14.55	13.82	8	8	12.12	28.57	17.02
1	5	31	6	23.31	10.91	19.68	19	1	28.78	3.57	21.27
1	6	32	5	24.06	9.09	19.68	19	1	28.78	3.57	21.27
1	7	31	6	23.31	10.91	19.68	19	1	28.78	3.57	21.27
1	8	35	5	26.32	9.09	21.27	19	1	28.78	3.57	21.27
1	9	35	5	26.32	9.09	21.27	19	1	28.78	3.57	21.27
1	10	35	4	26.32	7.27	20.74	19	1	28.78	3.57	21.27



According to Type I error and Type II error rate in the chart, I would prefer the models when Cost of Type II Error goes to 1:2, it gives you a balance between type error rates.

Here we pick the 1:2 cost ratio to do the following comparation.

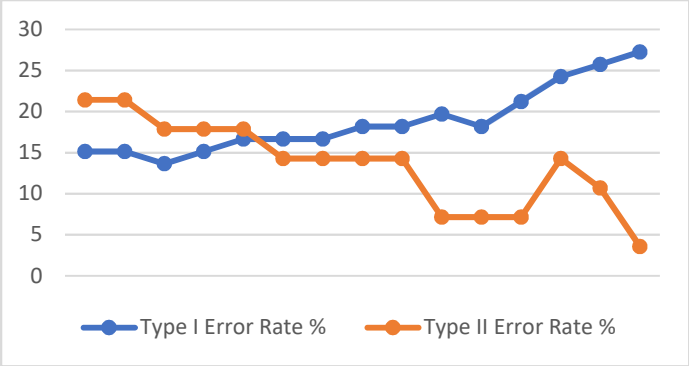
How do the results of each classifier compare to the cost-sensitive tree obtained in Part 4 of Assignment II? Comment.

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

Solution:

1. Cost sensitive classifier combined with **bagging and J48** *(default settings)*

Cost of Type I Error	Cost of Type II Error	Evaluate on train set					Re-evaluate on test set				
		Tpye1	Type 2	Type I Error Rate %	Type II Error Rate%	Misclassification Rates %	Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclassification Rates %
1	1	11	11	8.27	20	11.70	10	6	15.15	21.43	17.02
1	1.5	13	10	9.77	18.18	12.23	10	6	15.15	21.43	17.02
1	2	19	10	14.28	18.18	15.42	9	5	13.64	17.86	14.89
1	2.5	19	6	14.28	10.91	13.29	10	5	15.15	17.86	15.95
1	3	21	7	15.79	12.72	14.89	11	5	16.67	17.86	17.02
1	3.4	21	7	15.79	12.72	14.89	11	4	16.67	14.28	15.95
1	3.5	21	5	15.79	9.01	13.82	11	4	16.67	14.28	15.95
1	3.6	22	6	16.54	10.91	14.89	12	4	18.18	14.28	17.02
1	4	21	6	15.79	10.91	14.36	12	4	18.18	14.28	17.02
1	5	26	6	19.55	10.91	17.02	13	2	19.69	7.14	15.96
1	6	27	7	20.30	12.72	18.08	12	2	18.18	7.14	14.89
1	7	29	4	21.80	7.27	17.55	14	2	21.21	7.14	17.02
1	8	30	4	22.55	7.27	18.08	16	4	24.24	14.28	21.27
1	9	31	4	23.30	7.27	18.61	17	3	25.75	10.71	21.27
1	10	31	4	23.30	7.27	18.61	18	1	27.27	3.57	20.21



According to Type I error and Type II error rate in the chart, I would prefer the models when Cost of Type II Error goes to 1:3~1:4, it gives you a balance between type error rates. Although the type II error and the misclassification rate is relatively lower for 1:3.5, the type errors are more balance when the cost ratio goes to 1:3.

I'd pick 1:3 as the best cost ratio for bagging and J48.

Comparison:

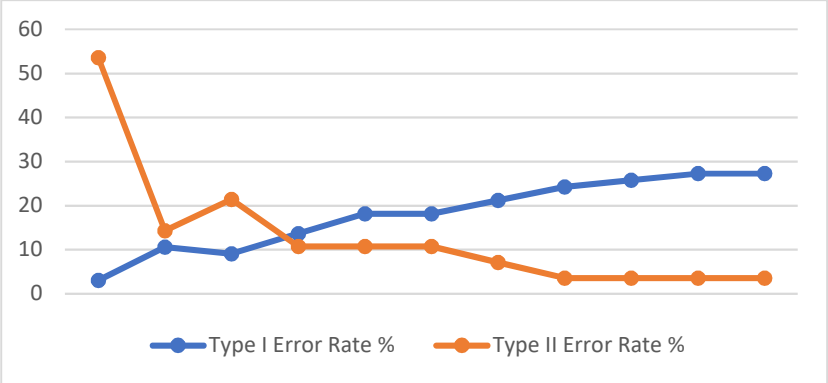
Classifier	Cost of Type I Error	Cost of Type II Error	Re-evaluate on test set				
			Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclassification Rates %
J48	1	2	10	8	15.15	28.57	19.14
Bagging J48	1	3	11	5	16.67	17.86	17.02
Bagging J48	1	3. 5	11	4	16.67	14.28	15.95

Comments: The performance of the data sets using the bagging and J48 base learner performed significantly better than that of J48 only. Although the type I error rate is a bit higher, the type II error rate and total misclassification rate decrease significantly.

2. Cost sensitive classifier combined with **bagging and Decision Stump** *(default settings)*

Cost of Type I Error	Cost of Type II Error	Evaluate on train set					Re-evaluate on test set				
		Tpye1	Type 2	Type I Error Rate %	Type II Error Rate%	Misclassification Rates %	Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclassification Rates %
1	0.5	11	21	8.27	38.18	17.02	2	15	3.03	53.57	18.08
1	0.6	15	16	11.27	29.09	16.48	7	4	10.60	14.28	11.70
1	0.7	16	13	12.03	23.63	15.42	6	6	9.09	21.42	12.76
1	0.8	17	15	12.78	27.27	17.02	9	3	13.63	10.71	12.76
1	0.9	19	12	14.28	21.81	16.48	12	3	18.18	10.71	15.95
1	1	22	8	16.54	14.54	15.95	12	3	18.18	10.71	15.95
1	2	28	5	21.05	9.09	17.55	14	2	21.21	7.14	17.02
1	3	30	4	22.56	7.27	18.08	16	1	24.24	3.57	18.08
1	4	30	3	22.56	5.45	17.55	17	1	25.75	3.57	19.14

1	5	31	3	23.31	5.45	18.08	18	1	27.27	3.57	20.21
1	6	32	3	24.06	5.45	18.61	18	1	27.27	3.57	20.21



According to Type I error and Type II error rate in the chart, I would pick the models when Cost of Type II Error goes to 1:0.6 and 1:0.8, it gives a balance between type error rates, with the lowest misclassification rate. However, the Type II Error Rate is lower of 1:0.8.

I'll pick 1:0.8 as the optimal cost ratio for bagging and decision stump.

Comparison:

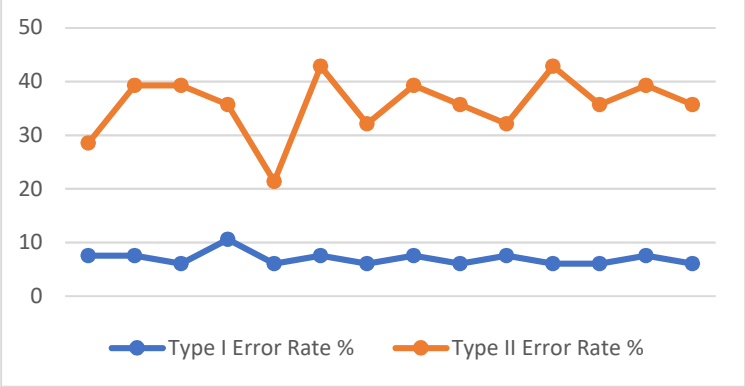
Classifier	Cost of Type I Error	Cost of Type II Error	Re-evaluate on test set				
			Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclassification Rates %
J48	1	3	8	8	12.12	28.57	17.02
bagging Decision Stump	1	0. 8	9	3	13.63	10.71	12.76

Comments: The performance of the data sets using the bagging and Decision Stump base learner performed significantly better than that of J48 only, even better than bagging and J48. The Type II error rate drop significantly.

3. Cost sensitive classifier combined with **boosting (AdaBoostM1) and J48** *(default settings)*

Cost of Type I Error	Cost of Type II Error	Evaluate on train set					Re-evaluate on test set				
		Tpye1	Type 2	Type I Error Rate %	Type II Error Rate%	Misclass ification Rates %	Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclass ification Rates %
1	0.1	12	12	9.02	21.81	12.76	5	8	7.57	28.57	13.82
1	0.2	13	13	9.77	23.63	13.82	5	11	7.57	39.28	17.02
1	0.3	15	14	11.27	25.45	15.42	4	11	6.06	39.28	15.97
1	0.4	10	11	7.51	20	11.17	7	10	10.61	35.71	18.08
1	0.5	14	11	10.52	20	13.29	4	6	6.06	21.42	10.63
1	0.6	15	12	11.27	21.81	14.36	5	12	7.57	42.85	18.08

1	0.7	14	12	10.52	21.81	13.82	4	9	6.06	32.14	13.82
1	0.8	12	13	9.02	23.63	13.29	5	11	7.57	39.28	17.02
1	0.9	11	11	8.27	20	11.70	4	10	6.06	35.71	14.89
1	1	13	16	9.77	29.09	15.42	5	9	7.57	32.14	14.89
1	2	16	10	12.03	18.18	13.82	4	12	6.06	42.85	17.02
1	3	15	12	11.27	21.81	14.36	4	10	6.06	35.71	14.89
1	4	14	15	10.53	27.27	15.42	5	11	7.57	39.28	17.02
1	5	15	7	11.27	12.72	11.70	4	10	6.06	35.71	14.89



According to Type I error and Type II error rate in the chart, I would pick the models when Cost of Type II Error goes to 0.5, it gives a relatively balance between type error rates, with the lowest misclassification rate.

I'll pick 1:0.5 as the optimal cost ratio for boosting and J48.

Comparison:

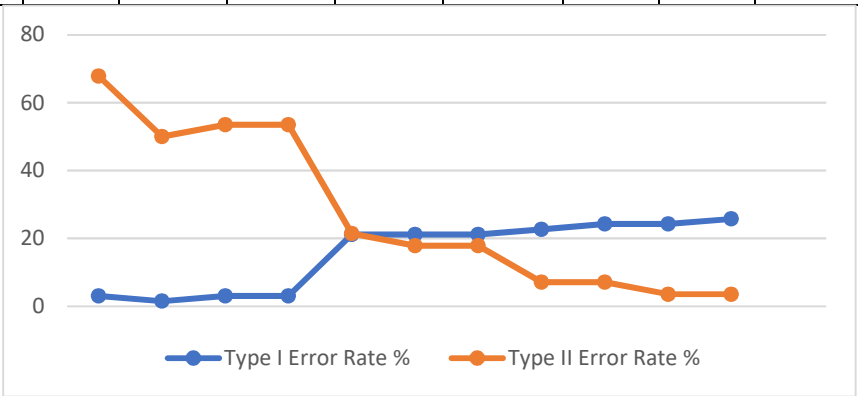
Classifier	Cost of Type I Error	Cost of Type II Error	Re-evaluate on test set				
			Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclassification Rates %
J48	1	3	8	8	12.12	28.57	17.02
boosting and J48	1	0. 5	4	6	6.06	21.42	10.63

Comments: The performance of the data sets using the boosting and J48 base learner performed better than that of J48 only. The type error and misclassification error all decreased, especially the type I error.

4. Cost sensitive classifier combined with **boosting (AdaBoostM1) and Decision Stump** *(default settings)*

Cost of Type I Error	Cost of Type II Error	Evaluate on train set					Re-evaluate on test set				
		Tpye1	Type 2	Type I Error Rate %	Type II Error Rate%	Misclass ification Rates %	Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclass ification Rates %
1	0.1	2	28	1.50	50.91	15.95	2	19	3.03	67.85	22.34
1	0.5	11	25	8.27	45.45	19.14	1	14	1.51	50	15.95
1	0.8	18	20	13.53	36.36	20.21	2	15	3.03	53.57	18.08

1	0.9	18	19	13.53	34.54	19.68	2	15	3.03	53.57	18.08
1	1	21	12	15.78	21.81	17.55	14	6	21.21	21.42	21.27
1	1.5	24	10	18.04	18.18	18.08	14	5	21.21	17.85	20.21
1	2	23	9	17.29	16.36	17.02	14	5	21.21	17.85	20.21
1	3	28	6	21.05	10.91	18.08	15	2	22.72	7.14	18.08
1	4	28	6	21.05	10.91	18.08	16	2	24.24	7.14	19.14
1	5	29	7	21.80	12.72	19.14	16	1	24.24	3.57	18.08
1	10	29	5	21.80	9.09	18.08	17	1	25.75	3.57	19.14



According to Type I error and Type II error rate in the chart, I would pick the models when Cost of Type II Error goes to 1:1~1:2. it gives a good balance between type error rates. Although the Type II Error Rate and total misclassification rate of 1:2 is relatively lower than 1:1, I pick 1:1 as the optimal cost ratio since it has the most balance type error rate.

I'll pick 1:1 as the optimal cost ratio for boosting and decision stump.

Comparison:

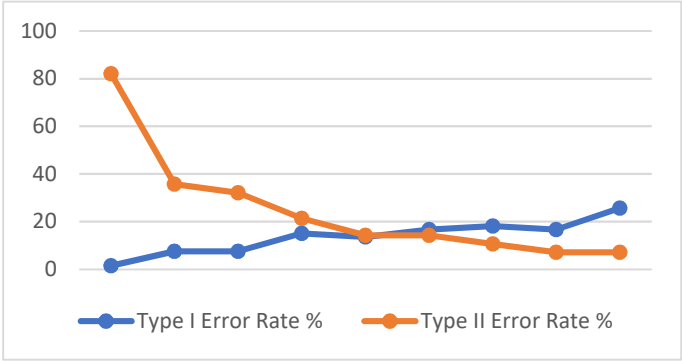
Classifier	Cost of Type I Error	Cost of Type II Error	Re-evaluate on test set				
			Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclassification Rates %
J48	1	3	8	8	12.12	28.57	17.02
boosting Decision Stump	1	1	14	6	21.21	21.42	21.27
boosting Decision Stump	1	2	14	5	21.21	17.85	20.21

Comments: The performance of the data sets using the boosting and Decision Stump base learner performed worse than that of J48 only. The type I error and misclassification error all increased, despite the type II error.

1. Cost sensitive classifier combined with **bagging and J48** *(25 iterations)*

Cost of Type I Error	Cost of Type II Error	Evaluate on train set					Re-evaluate on test set				
		Tpye1	Type 2	Type I Error Rate %	Type II Error Rate%	Misclassification Rates %	Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclassification Rates %
1	0.1	2	43	1.50	78.18	23.93	1	23	1.515	82.14	25.53
1	0.5	9	17	6.76	30.90	13.82	5	10	7.575	35.71	15.95
1	1	12	11	9.02	20	12.23	5	9	7.575	32.14	14.89
1	2	18	8	13.53	14.54	13.82	10	6	15.15	21.42	17.02
1	2.5	19	6	14.28	10.90	13.29	9	4	13.63	14.28	13.82
1	3	21	6	15.78	10.90	14.36	11	4	16.66	14.28	15.95
1	4	24	4	18.04	7.272	14.89	12	3	18.18	10.71	15.95
1	5	28	3	21.05	5.454	16.48	11	2	16.66	7.142	13.82
1	10	35	4	26.31	7.272	20.74	17	2	25.75	7.142	20.21

Bagging with 25 iterations and base learner



According to Type I error and Type II error rate in the chart, I would pick the models when Cost of Type II Error goes to 1:2.5. it gives a balance between type error rates, it also has the lowest misclassification rate.

Comparison:

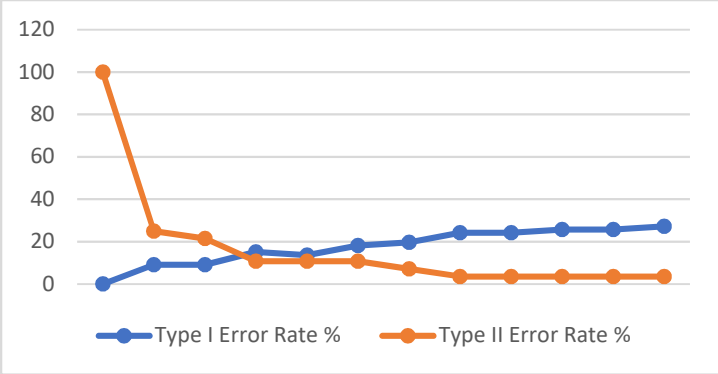
Classifier	Cost of Type I Error	Cost of Type II Error	Re-evaluate on test set				
			Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclassification Rates %
J48	1	3	8	8	12.12	28.57	17.02
Bagging J48	1	3. 5	11	4	16.67	14.28	15.95
Bagging J48(25)	1	2. 5	9	4	13.63	14.28	13.82

Comments: The performance of the data sets using the Bagging J48 (25 iterations) base learner performed better than that of J48 only or bagging J48 (10 iterations). As the iteration step increased, the type I error rate decreased while type II error stay the same. Increasing the iterations helps to find a better type error balance.

2. Cost sensitive classifier combined with **bagging and Decision Stump** *(25 iterations)*

Cost of Type I Error	Cost of Type II Error	Evaluate on train set					Re-evaluate on test set				
		Tpye1	Type 2	Type I Error Rate %	Type II Error Rate%	Misclass ification Rates %	Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclass ification Rates %
1	0.1	1	51	0.75	92.72	27.65	0	28	0	100	29.78
1	0.5	11	20	8.27	36.36	16.48	6	7	9.09	25	13.82
1	0.6	13	16	9.77	29.09	15.42	6	6	9.09	21.42	12.76
1	0.7	15	16	11.27	29.09	16.48	10	3	15.15	10.71	13.82
1	0.8	17	14	12.78	21.21	16.48	9	3	13.63	10.71	12.76
1	0.9	21	9	15.78	16.36	15.95	12	3	18.18	10.71	15.95
1	1	22	9	16.54	16.36	16.48	13	2	19.69	7.14	15.95
1	2	28	6	21.05	10.90	18.08	16	1	24.24	3.57	18.08
1	3	29	4	21.80	7.27	17.55	16	1	24.24	3.57	18.08
1	4	29	3	21.80	5.45	17.02	17	1	25.75	3.57	19.14
1	5	29	3	21.80	5.45	17.02	17	1	25.75	3.57	19.14
1	10	34	3	25.56	5.45	19.68	18	1	27.27	3.57	20.21

Bagging with 25 iterations and base learner



According to Type I error and Type II error rate in the chart, I would pick the models when Cost of Type II Error goes to 1:0.8. it gives a balance between type error rates, it also has the lowest misclassification rate.

Comparison:

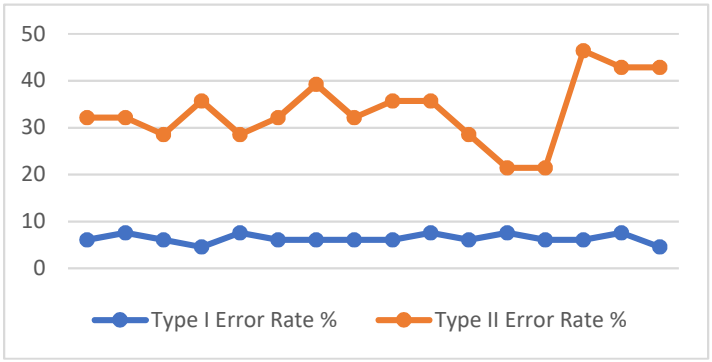
Classifier	Cost of Type I Error	Cost of Type II Error	Re-evaluate on test set				
			Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclassification Rates %
J48	1	3	8	8	12.12	28.57	17.02
Bagging Decision Stump	1	0. 8	9	3	13.63	10.71	12.76
Bagging Decision Stump (25)	1	0. 8	9	3	13.63	10.71	12.76

Comments: The performance of the data sets using the Bagging Decision Stump (25 iterations) base learner performed same as that of bagging Decision Stump (10 iterations), better than J48 only. As the iteration step increased, there's no change happens, it might be that already find a best regression with default iteration.

3. Cost sensitive classifier combined with **boosting (AdaBoostM1) and J48** *(25 iterations)*

Cost of Type I Error	Cost of Type II Error	Evaluate on train set					Re-evaluate on test set				
		Tpye1	Type 2	Type I Error Rate %	Type II Error Rate%	Misclass ification Rates %	Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclass ification Rates %
1	0.1	13	14	9.77	25.45	14.36	4	9	6.06	32.14	13.82
1	0.5	11	10	8.27	18.18	11.17	5	9	7.57	32.14	14.89
1	1	11	13	8.27	23.63	12.76	4	8	6.06	28.57	12.7
1	2	15	10	11.27	18.18	13.29	3	10	4.54	35.71	13.82
1	3	14	10	10.52	18.18	12.766	5	8	7.57	28.57	13.82
1	4	12	10	9.02	18.18	11.70	4	9	6.06	32.14	13.82
1	5	13	11	9.77	20	12.76	4	11	6.06	39.28	13.82
1	6	16	12	12.03	21.81	14.89	4	9	6.06	32.14	13.82
1	7	15	13	11.27	23.63	14.89	4	10	6.06	35.71	14.89
1	8	13	10	9.77	12.23	12.23	5	10	7.57	35.71	15.95
1	9	15	10	11.27	18.18	13.29	4	8	6.06	28.57	12.76
1	10	11	10	8.27	18.18	11.17	4	6	6.06	21.42	10.63
1	11	13	11	9.77	20	12.76	5	12	7.57	42.85	18.08
1	12	13	10	9.77	18.18	12.23	3	12	4.54	42.85	15.95

Bagging with 25 iterations and base learner



According to Type I error and Type II error rate in the chart, I would pick the models when Cost of Type II Error goes to 1:9.5 or 1:10. it gives a balance between type error rates. Since the type I error and total misclassification error of 1:10 is relatively lower, and it also has the lowest misclassification rate. I'd pick 1:10 as the optimal cost ratio.

Comparison:

Classifier	Cost of Type I Error	Cost of Type II Error	Re-evaluate on test set				
			Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclassification Rates %

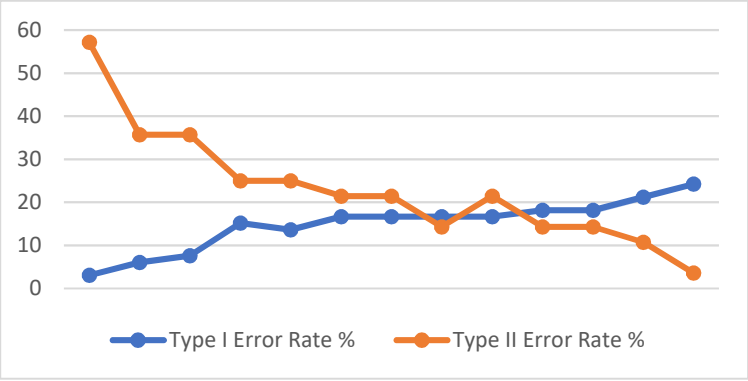
J48	1	3	8	8	12.12	28.57	17.02
boosting J48	1	0. 5	4	6	6.06	21.42	10.63
boosting J48 (25)	1	10	4	6	6.06	21.42	10.63

Comments: The performance of the data sets using the boosting J48 (25 iterations) base learner performed same as that of boosting J48 (10 iterations), better than J48 only. As the iteration step increased, there's no change happens, it might be that already find a best regression with default iteration.

4. Cost sensitive classifier combined with **boosting (AdaBoostM1) and Decision Stump** *(25 iterations)*

Cost of Type I Error	Cost of Type II Error	Evaluate on train set					Re-evaluate on test set				
		Tpye1	Type 2	Type I Error Rate %	Type II Error Rate%	Misclass ification Rates %	Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclass ification Rates %
1	0.1	3	32	2.25	58.18	18.6	2	16	3.03	57.14	19.14
1	0.5	8	20	6.01	36.36	14.89	4	10	6.06	35.71	14.89
1	1	15	15	11.27	27.27	15.95	5	10	7.57	35.71	15.95
1	2	19	10	14.28	18.18	15.42	10	7	15.15	25	18.08
1	3	24	8	18.04	14.54	17.02	9	7	13.63	25	17.02
1	3.5	25	7	18.79	12.72	17.02	11	6	16.67	21.42	18.08
1	3.6	23	7	17.29	12.72	15.95	11	6	16.67	21.42	18.08
1	3.7	25	7	18.79	12.72	17.02	11	4	16.67	14.28	15.95
1	3.8	25	7	18.79	12.72	17.02	11	6	16.67	21.42	18.08
	3.9	24	7	18.04	12.72	16.48	12	4	18.18	14.28	17.02
1	4	25	7	18.79	12.72	17.02	12	4	18.18	14.28	17.02
1	5	24	6	18.04	10.90	15.95	14	3	21.21	10.71	18.08
1	10	29	5	21.80	9.090	18.08	16	1	24.24	3.571	18.08

Number of performed Iterations: 25



According to Type I error and Type II error rate in the chart, I would pick the models when Cost of Type II Error goes to 1:3.7. it gives a balance between type error rates, it also has the lowest misclassification rate.

Comparison:

Classifier	Cost of Type I Error	Cost of Type II Error	Re-evaluate on test set				
			Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclassification Rates %
J48	1	3	8	8	12.12	28.57	17.02
boosting Decision Stump	1	1	14	6	21.21	21.42	21.27
boosting Decision Stump (25)	1	3. 7	11	4	16.67	14.28	15.95

Comments: The performance of the data sets using boosting Decision Stump (25 iterations) base learner performed better than that of boosting Decision Stump (10 iterations), better than J48 only. As the iteration step increased, the type I & II error rate both decreased. Increasing the iterations helps to find a better type error balance.

Comparison of all set:

Classifier	Cost of Type I Error	Cost of Type II Error	Re-evaluate on test set				
			Tpye1	Tpye2	Type I Error Rate %	Type II Error Rate %	Misclassification Rates %
J48	1	3	8	8	12.12	28.57	17.02
J48(Bagging)	1	3. 5	11	4	16.67	14.28	15.95
J48(Bagging) (25)	1	2. 5	9	4	13.63	14.28	13.82
J48(Boosting)	1	0. 5	4	6	6.06	21.42	10.63
J48 (Boosting) (25)	1	10	4	6	6.06	21.42	10.63
Decision Stump (Bagging)	1	0. 8	9	3	13.63	10.71	12.76
Decision Stump (Bagging) (25)	1	0. 8	9	3	13.63	10.71	12.76
Decision Stump (Boosting)	1	1	14	6	21.21	21.42	21.27
Decision Stump (Boosting) (25)	1	3. 7	11	4	16.67	14.28	15.95

1: CompareJ48 and J48 with bagging and boosting:

Using techniques like Bagging and Boosting helps to decrease the variance and increased the robustness of the model. The type error rates are more balanced than that of J48 only. When using Bagging, J48 find the best type error rates; when using Boosting, the type error rates balance just improved a little, the type I error rates decreased significantly thus achieves the lowest total misclassification rate.

2: Compare Decision Stump with Bagging and with Boosting:

Decision stump with Bagging performs better than that of with Boosting, the type 1 error and type 2 error rate all lower than that with boosting.

3: Compare default iterations and 25 iterations:

When J48 with bagging, increasing the iterations help to find a more stable model. When J48 with boosting, increasing the iterations, the type error rates and misclassification rate all remain the same of each optimal cost ratio.

When increasing the iterations, the decision stump with bagging seems no change, it means the model already find the stable model with default iterations.