

Assignment 7

Jingyuan Liu

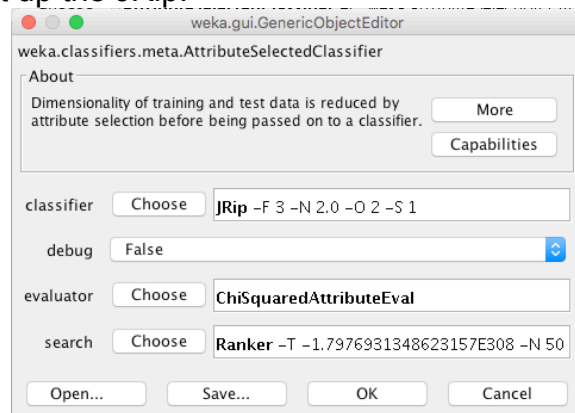
AndrewId: jingyual email: jingyual@andrew.cmu.edu

Step-by-Step Guide:

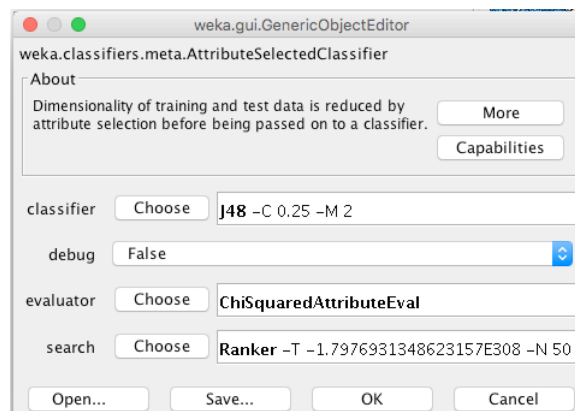
1. Complete the readings through week 10.
2. Use the experimenter to determine whether you get significantly different performance from J48 (tree based learning) and JRIP (rule based learning) when you use a feature selection wrapper that selections the top 50 features on each fold. Give screen shots and explain your results.

Solution:

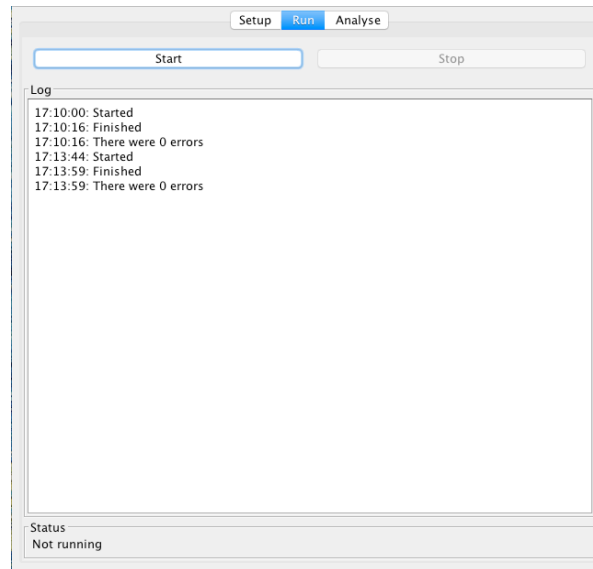
At first, we can set up the JRip:



Then we can set up the J48:



Then we can run the experimenter and got the result:



So we see that the training process was conducted with zero errors. Then we could turn to the Analysis panel to see the result:

Test output				
Tester:	weka.experiment.PairedCorrectedTTester			
Analysing:	Percent_correct			
Datasets:	1			
Resultsets:	2			
Confidence:	0.05 (two tailed)			
Sorted by:	-			
Date:	15-11-11 下午5:17			
Dataset	(1) meta.Att (2) meta.			
'training set'	(100)	67.04		62.75
		(v/ /*)		(0/1/0)
Key:				
(1) meta.AttributeSelectedClassifier	'-E \"ChiSquaredAttributeEval \" -S \"Rank			
(2) meta.AttributeSelectedClassifier	'-E \"ChiSquaredAttributeEval \" -S \"Rank			

3. Troubleshoot your results so that you understand why performance was or was not different between tree and rule based learning. The point here is not to identify where the feature space is weak but to investigate why the algorithms did or didn't perform differently. This can be thought of as a more advanced version of what you did with the Titanic dataset earlier in the semester. It should be doable since you will only be considering around 50 features. You may need to be creative. Now explain your results using what you understand about tree and rule based learning and what you found in your error analysis.

Solution:

At first, we could review the analysis result:

Test output			
Tester: weka.experiment.PairedCorrectedTTester			
Analysing: Percent_correct			
Datasets: 1			
Resultsets: 2			
Confidence: 0.05 (two tailed)			
Sorted by: -			
Date: 15-11-11 下午5:17			
Dataset	(1) meta.Att (2) meta.		
<hr/>			
'training set'	(100)	67.04	62.75
<hr/>			
	(v/ /*) (0/1/0)		
Key:			
(1) meta.AttributeSelectedClassifier '-E \"ChiSquaredAttributeEval \" -S \"Rank			
(2) meta.AttributeSelectedClassifier '-E \"ChiSquaredAttributeEval \" -S \"Rank			

We could see that this is not significant important. Besides, in this case, we could see that the JRip is better than the Tree model.

To saw the reason behind, we need to figure out why and when rule-based classifier is better than tree-based model. Therefore, we could try different amount of features.

We could set the feature to 20, and got the result:

Test output

Tester: weka.experiment.PairedTTester
Analysing: Percent_correct
Datasets: 1
Resultsets: 2
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 15-11-11 下午5:40

Dataset	(1) meta.Att (2) meta.	
'training set'	(100) 66.99 65.50	

(v/ /*) | (0/1/0)

Key:
(1) meta.AttributeSelectedClassifier '-E \"ChiSquaredAttributeEval \" -S \"Rank
(2) meta.AttributeSelectedClassifier '-E \"ChiSquaredAttributeEval \" -S \"Rank

From this figure, we could found that the result is different from result before. The tree-based classifier got a better result, while the rule-based got worse.

Then we could turn to the feature size of 100:

Test output			
Tester:	weka.experiment.PairedTTester		
Analysing:	Percent_correct		
Datasets:	1		
Resultsets:	2		
Confidence:	0.05 (two tailed)		
Sorted by:	-		
Date:	15-11-11 下午5:46		
Dataset	(1) meta.Att (2) meta.		
'training set'	(100)	66.65	62.14 *
	(v/ /*) (0/0/1)		
Key:			
(1) meta.AttributeSelectedClassifier	'-E \"ChiSquaredAttributeEval \" -S \"Rank		
(2) meta.AttributeSelectedClassifier	'-E \"ChiSquaredAttributeEval \" -S \"Rank		

We could see that rule-based model is almost the same, while the tree-based model got worse with the increase of the feature space.

We could see that tree-based models would be likely to perform worse with the increase of the feature space. This could be caused by the reason of over-fitting.

On the other hand, rule-based model is not that feature-space sensitive to the feature space size.

4. Are the results surprising given the discussion about tree and rule based learning in the book? Why or why not?

Solution:

I think it is reasonable for these two models to perform as above. Based on the knowledge from book, we would know that tree-based models are very likely to be over-fitting, especially with the increase of feature space increase. On the other hand, rule learning is more prone to over-fitting, because it tends to make decisions based on more local information.