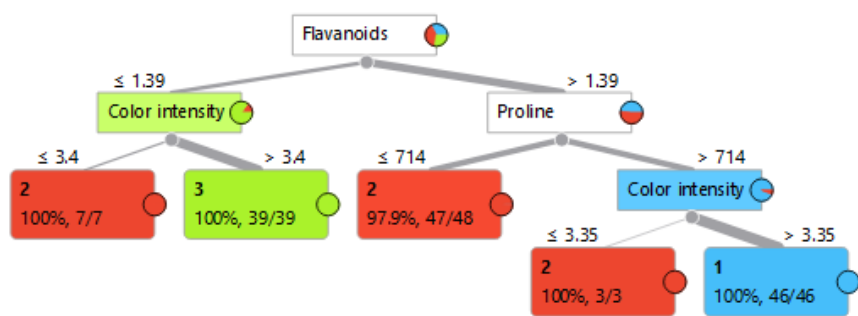


The features of the most informative projection were: Alcohol and Flavanoids

The wine from the second region has a much larger spread with a lower alcohol content. The wine from the Third region has much lower flavanoids than the other two. While the first region has much higher alcohol and

Tree size: 9 nodes, 5 leaves  
Edge widths: Relative to parent  
Target class: None



The root of the tree is the Flavonoids, this correspond with the large difference from the above scatter plot.

Test and Score

Sun Aug 03 25, 19:00:34

Settings

Sampling type: Stratified 5-fold Cross validation  
Target class: None, show average over classes

Scores

Model	AUC	CA	F1	Prec	Recall	MCC
Tree	0.933	0.909	0.908	0.911	0.909	0.865
kNN	0.846	0.720	0.718	0.718	0.720	0.578

Classification accuracies of models:  
kNN: 0.720  
Tree: 0.909

Confusion Matrix

Sun Aug 03 25, 19:05:32

Confusion matrix for kNN (showing number of instances)

		Predicted			
		1	2	3	Σ
Actual	1	42	2	3	47
	2	6	38	13	57
	3	4	12	23	39
	Σ	52	52	39	143

Region 2 is misclassified more frequently in both models:

Predicted as 1: 6 times

Predicted as 3: 13 times

Total misclassified: 19 instances out of 57

Both models classify the difference between Region 1 and 3 well.

Confusion Matrix

Sun Aug 03 25, 19:05:38

Confusion matrix for Tree (showing number of instances)

		Predicted			
		1	2	3	Σ
Actual	1	45	2	0	47
	2	6	47	4	57
	3	0	1	38	39
	Σ	51	50	42	143

Predicted as 1: 6 times

Predicted as 3: 4 times

Total misclassified: 10 instances out of 57

Both models classify the difference between Region 1 and 3 well.

The kNN misclasifies a greater amount.

We split the data into separate training and test sets to make sure that it would work for not just the training set and would work for any set produced from the same source.

