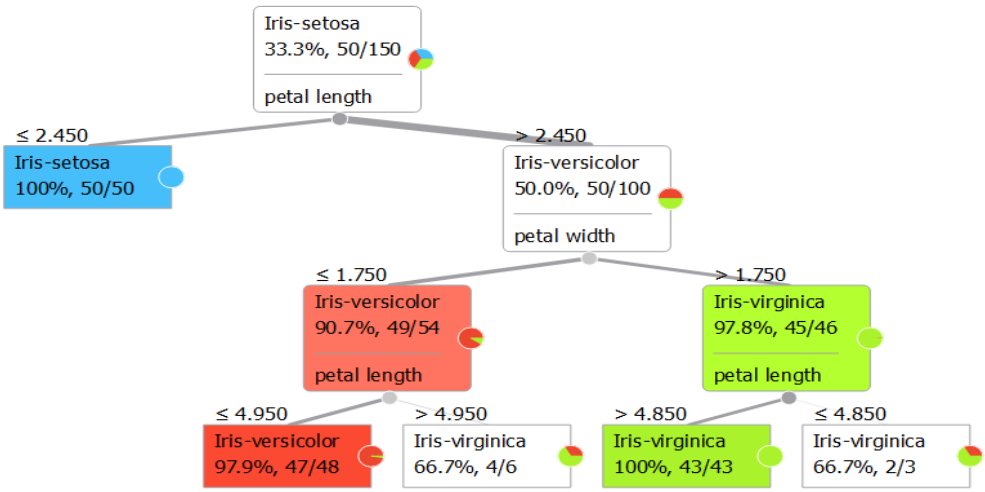
2.4 Problem 4
Build two classification trees using the iris sample dataset within the Orange application. Keep all parameters for both classifiers the same (Feature Selection, Pruning), and modify the Limit Depth parameter to a smaller value than the default (e.g., from 10 to 2). How does this affect the Precision and Recall of the classifier? What types of flowers are misclassified? Why? What does Tan refer to as the border where these misclassifications occur?
Answer:

For the above problem, depth parameters where assumed as 3 and 10. As the depth increases the precision and recall decreases.

Iris-Versicolor and Iris-Verginica are misclassified. The flowers are misclassified as only the petal length and petal width are considered while classifying.

Tan refers to the border where these misclassifications occur as the decision boundary.

Depth=3



Confusion Matrix

Wed Sep 21 16, 23:06:42

Confusion matrix for Classification Tree (showing proportion of predicted)

		Predicted			
		Iris-setosa	Iris-versicolor	Iris-virginica	Σ
Actual	Iris-setosa	100.0 %	0.0 %	0.0 %	50
	Iris-versicolor	0.0 %	93.8 %	9.6 %	50
	Iris-virginica	0.0 %	6.2 %	90.4 %	50
	Σ	50	48	52	150

Test & Score

Wed Sep 21 16, 23:06:58

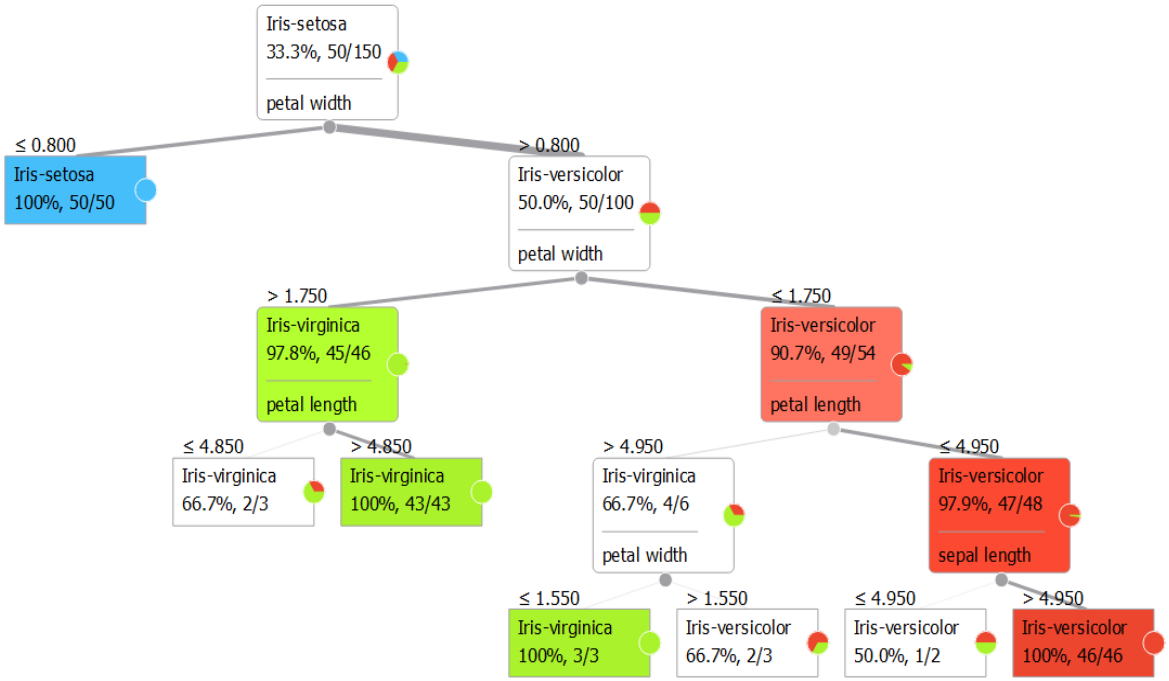
Settings

Sampling type: Stratified 3-fold Cross validation
Target class: Average over classes

Scores

Method	AUC	CA	F1	Precision	Recall
Classification Tree	0.959	0.947	0.947	0.947	0.947

Depth=10



Confusion Matrix Wed Sep 21 16, 22:57:06

Confusion matrix for Classification Tree (showing proportion of actual)

		Predicted			
		Iris-setosa	Iris-versicolor	Iris-virginica	Σ
Actual	Iris-setosa	100.0 %	0.0 %	0.0 %	50
	Iris-versicolor	0.0 %	92.0 %	8.0 %	50
	Iris-virginica	0.0 %	12.0 %	88.0 %	50
	Σ	50	52	48	150

Test & Score Wed Sep 21 16, 22:58:29

Settings

Sampling type: Stratified 3-fold Cross validation
Target class: Average over classes

Scores

Method	AUC	CA	F1	Precision	Recall
Classification Tree	0.949	0.933	0.933	0.934	0.933