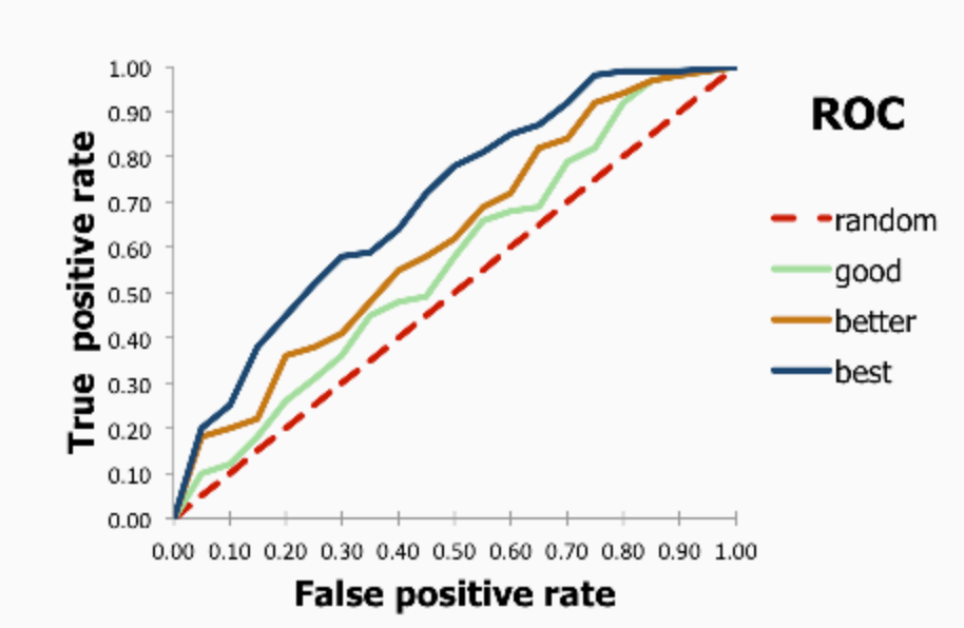
Unit 6–3
Evaluating a classification model



Context & Recap

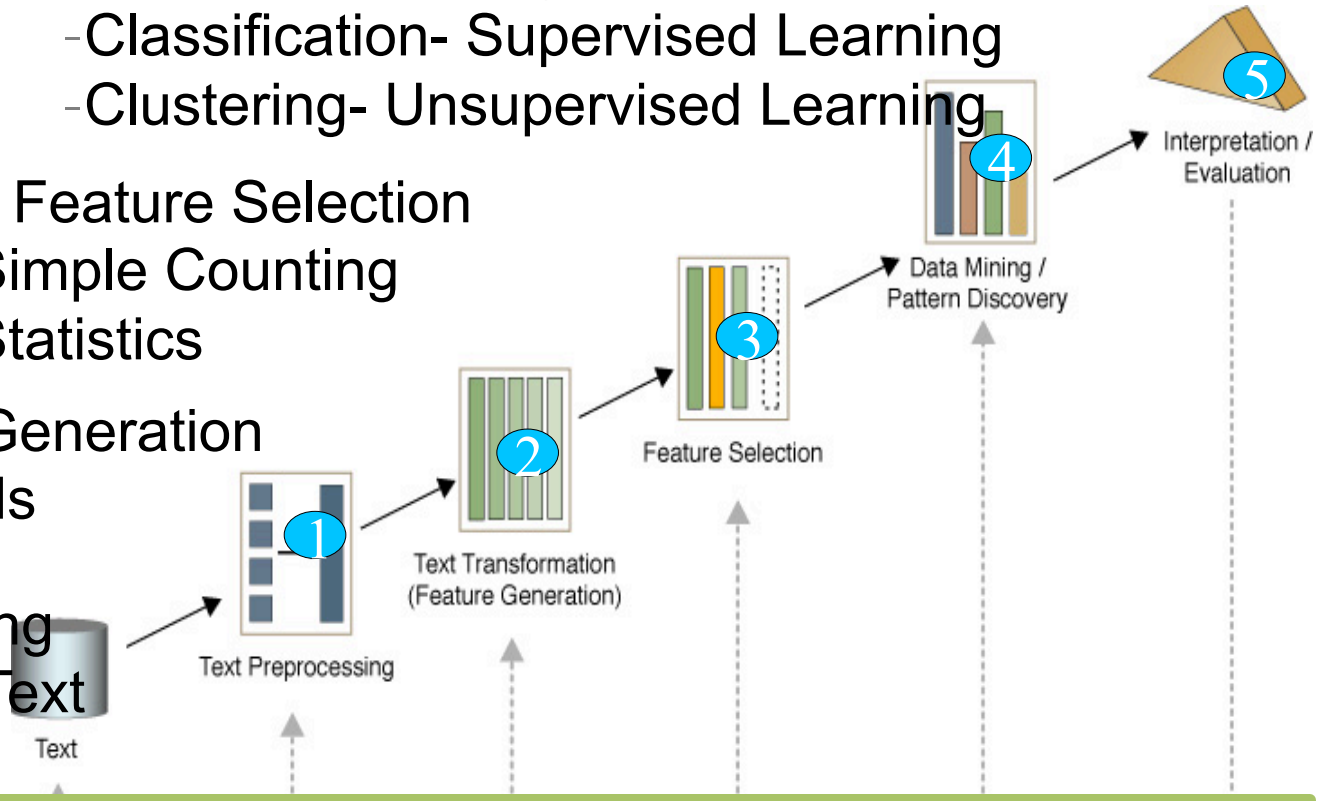
5. Analyzing Results

4. Text/Data Mining
- Classification- Supervised Learning
- Clustering- Unsupervised Learning

3. Feature Selection
- Simple Counting
- Statistics

2. Features Generation
- Bag of Words

1. Text Preprocessing
- Syntactic/Semantic Text Analysis



label	confidence(crime)	confidence(kenya)	confidence(healthcare)	prediction(label)	metada...
unknown	0.010	0.013	0.978	healthcare	U1.txt
unknown	0.001	0.975	0.024	kenya	U2.txt
unknown	0.029	0.020	0.951	healthcare	U3.txt
unknown	0.964	0.006	0.030	crime	U4.txt

Evaluating Algorithms

Algorithms can be evaluated based on figures in the confusion matrix as follows:

		Predicted class	
		Yes (+)	No (-)
Actual Class	Yes (+)	50	5
	No (-)	10	40

		Predicted class	
		Yes (+)	No (-)
Actual Class	Yes (+)	True Positive (TP)	False Negative(FN)
	No (-)	False Positive (FP)	True Negative(TN)

True Positive Rate (TPR) is the percentage of positive samples predicted correctly. TPR = TP / (TP+FN) = 50/55 **(same as recall)**

$$\frac{\text{Positive correctly classified}}{\text{Total positives}}$$

True Negative Rate(TNR) is the percentage of negative samples correctly predicted, i.e. TN / (TN+FP) = 40 / 50

False Negative Rate(FNR) is the percentage of positive samples incorrectly classified. i.e. FN / (TP+FN) = 5/55

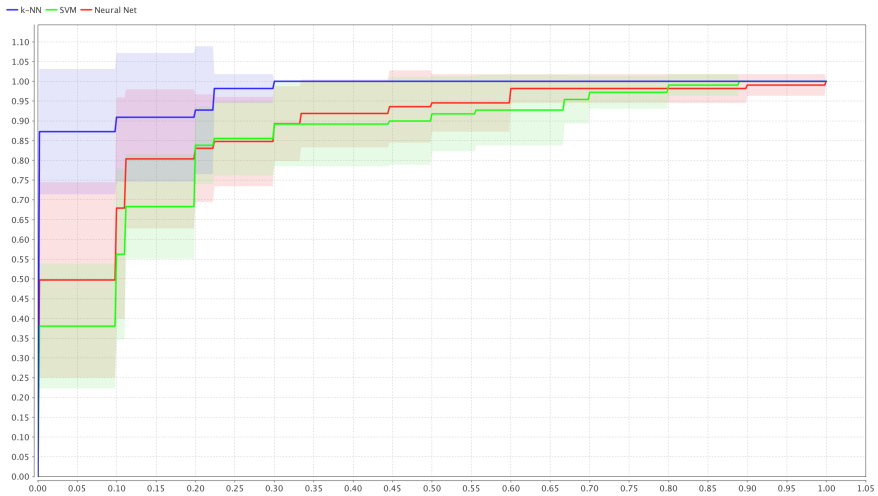
False Positive Rate (FPR) is percentage of negative samples incorrectly classified. i.e. FP / (TN+FP) = 10/50

$$\frac{\text{Negatives incorrectly classified}}{\text{Total negatives}}$$

Precision is percentage of rows predicted as postive that were classified correctly. i.e. TP /(TP + FP) = 50/60

Evaluating Algorithms

- The confusion matrix is one popular way to measure algorithm performance. Another popular method is to plot the Receiver Operator Characteristic (ROC) curve of a number of algorithms as illustrated below. The algorithm with the highest area under the curve (AUC) is the best; *k*-NN in this case (blue curve)
  - Needs a binary class label



The following slides explain how to generate and ROC curve, and how to read it.

ROC curves

label	confidence(crime)	confidence(kenya)	confidence(healthcare)	prediction(label)	metada...
unknown	0.010	0.013	0.978	healthcare	U1.txt
unknown	0.001	0.975	0.024	kenya	U2.txt
unknown	0.029	0.020	0.951	healthcare	U3.txt
unknown	0.964	0.006	0.030	crime	U4.txt

- ROC curves also tells us how confident an algorithm should be in its prediction before we accept it, for example:
  - Should we accept a prediction if the algorithm is 50% confident or higher?
  - Or should an algorithm be 70% confident before we accept a prediction?
  - How would we recognise a document that is not about one of our three topics? How confident would the predictions be?

The value chosen will effect class precision and recall as explained in the next slide:

Evaluating Algorithms

Precision & Recall (for one class, e.g Kenya)

A model that declares every document to be about Kenya will have perfect recall for Kenya (it found all the Kenya docs) but poor precision.

i.e. accept all predictions of Kenya, with a high or low confidence levels

Conversely a model that declares very few docs to be about Kenya will have a very high precision for that class (a doc predicted to be about Kenya was always a correct prediction) but poor recall. This would happen, for example, if a model only classified as positive instances those that had an exact term matches with the training data set.

i.e. accept predictions with a high confidence level only

What is the optimal confidence level? The aim of a classifier is to optimise both precision and recall.

Evaluating Algorithms

An ROC (Receiver Operator Characteristics) curve indicates the optimal confidence level from which to accept the outcome from a classifier. It is a plot of TPR (true positive rate) against FPR (false positive rate) and is constructed as follows:

1. List all predictions from a classifier, and the confidence level of that prediction.
2. Sort the predictions in order of confidence level.

   The next slide shows a classifier that classified 10 instances, ordered by confidence level. The top row is the actual class of the instance. The second row, **confidence**, is how confident the classifier is that this instance is positive.

   The first instance is in the positive class, but the classifier was only 25% confident that is was positive.

   The second instance is in the negative class, and the classifier is 43% confident that it is a positive instance.
3. Having sorted the instances, work from left to right as follows:

Evaluating Algorithms

4. To start with, assume all instances are positive. Essentially you are starting off by accepting positive predictions, regardless of confidence level. Therefore the classifier has classified five instance correctly, and 5 instances incorrectly. Calculate TP, FP, TN, FN, TPR & FPR.

Note: at this point recall is 100%, but precision is only 50%

5. Next, reject the first prediction and accept all the others as positive. In other words accept any positive prediction where the confidence level is higher than 0.25. Again calculate TP, FP, TN, FN, TPR & FPR. At this point TP goes down to 4 because the first instance, which is actually positive is being classed as negative.

Note: at this point recall is down to 80%, & precision is also down to 44%

6. Next reject the first two predictions, and accept all others. In other words accept any positive prediction where the confidence level is higher than 0.43. Again calculate TP, FP, TN, FN, TPR & FPR. This time FP goes down to 4 because the second instance, which had been incorrectly classed as positive, is now correctly classed as negative. That leaves only four negatives incorrectly classed as positive.

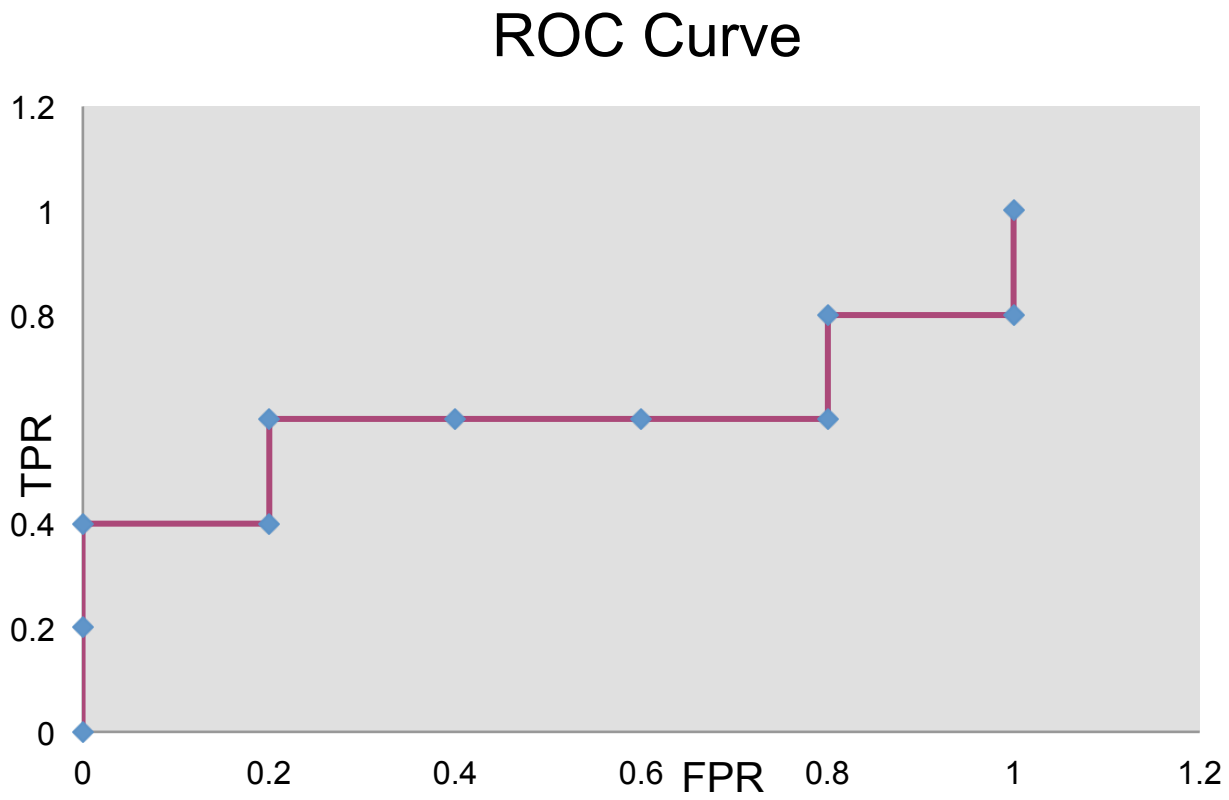
Note: at this point recall is still at 80%, & precision is back up to 50%

Evaluating Algorithms

7. Continue in this way across the table, calculating TP, FP, TN, FN, TPR & FPR at each stage.
8. The ROC curve is TPR plotted against FPR. The figures below give the following curve:

	Instances										
Actual class	+	-	+	-	-	-	+	-	+	+	
Confidence it's in + class	0.25	0.43	0.53	0.76	0.80	0.82	0.85	0.87	0.93	0.95	1
TP	5	4	4	3	3	3	3	2	2	1	0
FP	5	5	4	4	3	2	1	1	0	0	0
TN	0	0	1	1	2	3	4	4	5	5	5
FN	0	1	1	2	2	2	2	3	3	4	5
TPR	1	0.8	0.8	0.6	0.6	0.6	0.6	0.4	0.4	0.2	0
FPR	1	1	0.8	0.8	0.6	0.4	0.2	0.2	0	0	0
Precision	0.5	0.44	0.5	0.43	0.5	0.6	0.75	0.67	1	1	
Recall	1	0.8	0.8	0.6	0.6	0.6	0.6	0.4	0.4	0.2	

ROC Curve from previous table



Interpreting an ROC curve

There are three critical points along an ROC curve that have the following interpretations:

- (TPR=1, FPR=1): Model predicts EVERY instance is in the positive class.
- (TPR=0, FPR=0): Model predicts that NO instance is in the positive class.
- (TPR=1, FPR=0): The ideal model. 100% true postives, and 0% false positives, giving 100% precision and recall.

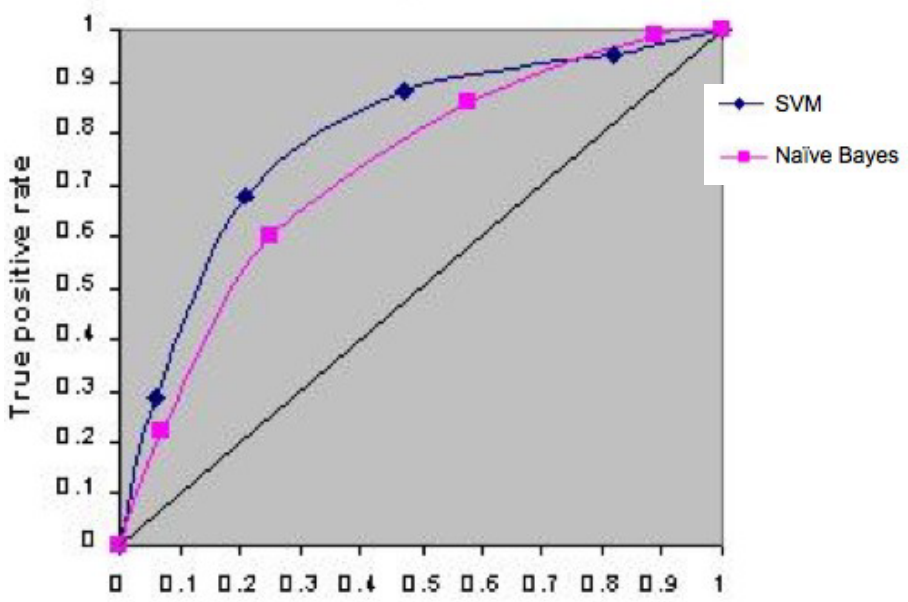
A good classifier model should have an ROC curve that comes close to the upper left hand corner of the diagram. The closest point represent the optimal confidence level.

Therefore the optimal model from the data above is (TPR=0.6, FPR=0.2). From slide 7 this is at the 0.85 confidence level. Therefore, for the data given above, the optimal model is got by classifying an instance as positive if it has a confidence >= 0.85, otherwise class it as a negative. (i.e. recall=0.6, precision=0.75).

A diagonal line from (0,0) to (1,1) represents random guessing.

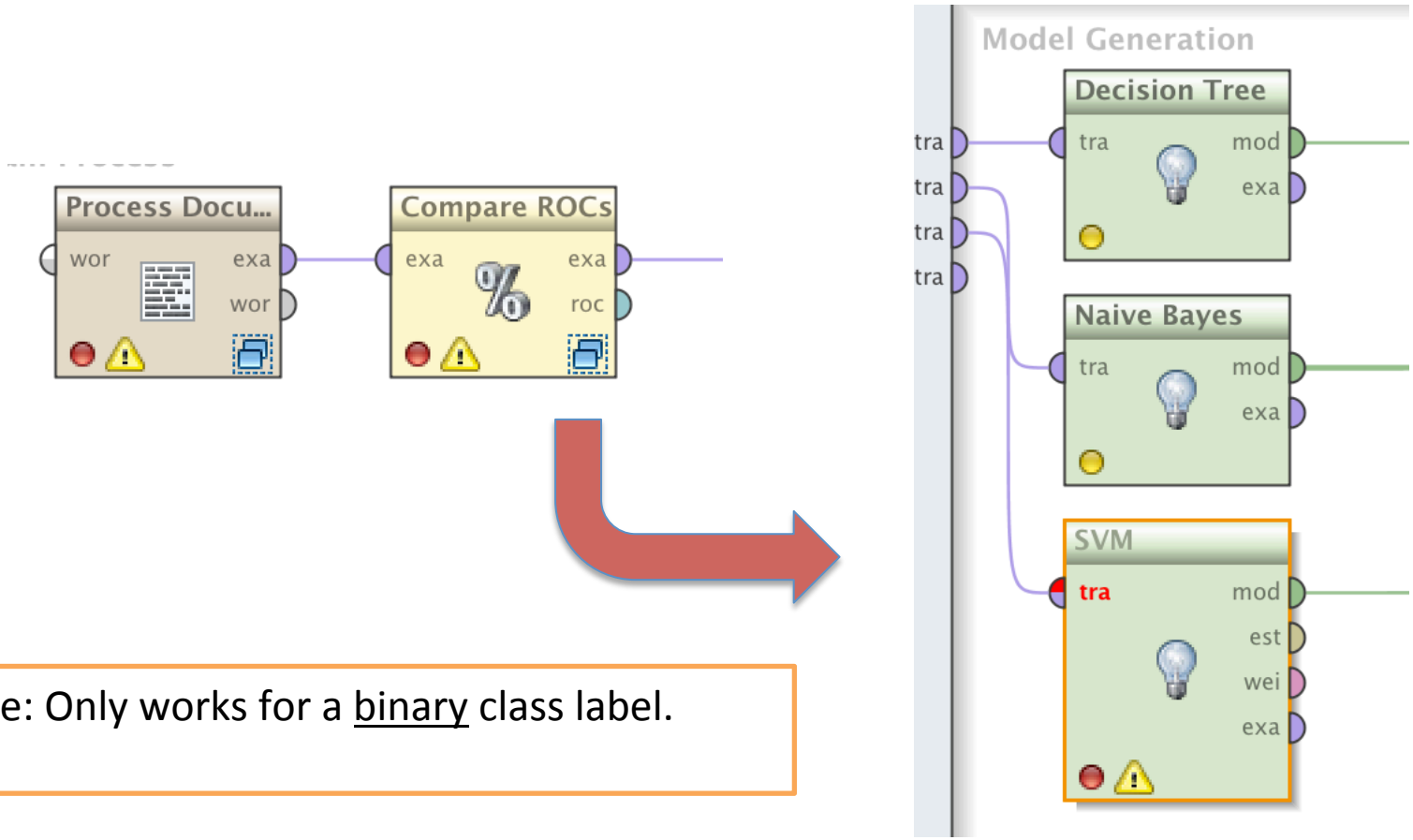
Comparing classifiers

The performance of classifiers can be compared by drawing their ROC curves. The curve with the largest Area Under the Curve (AUC) is the better classifier.



ROC curves in RM

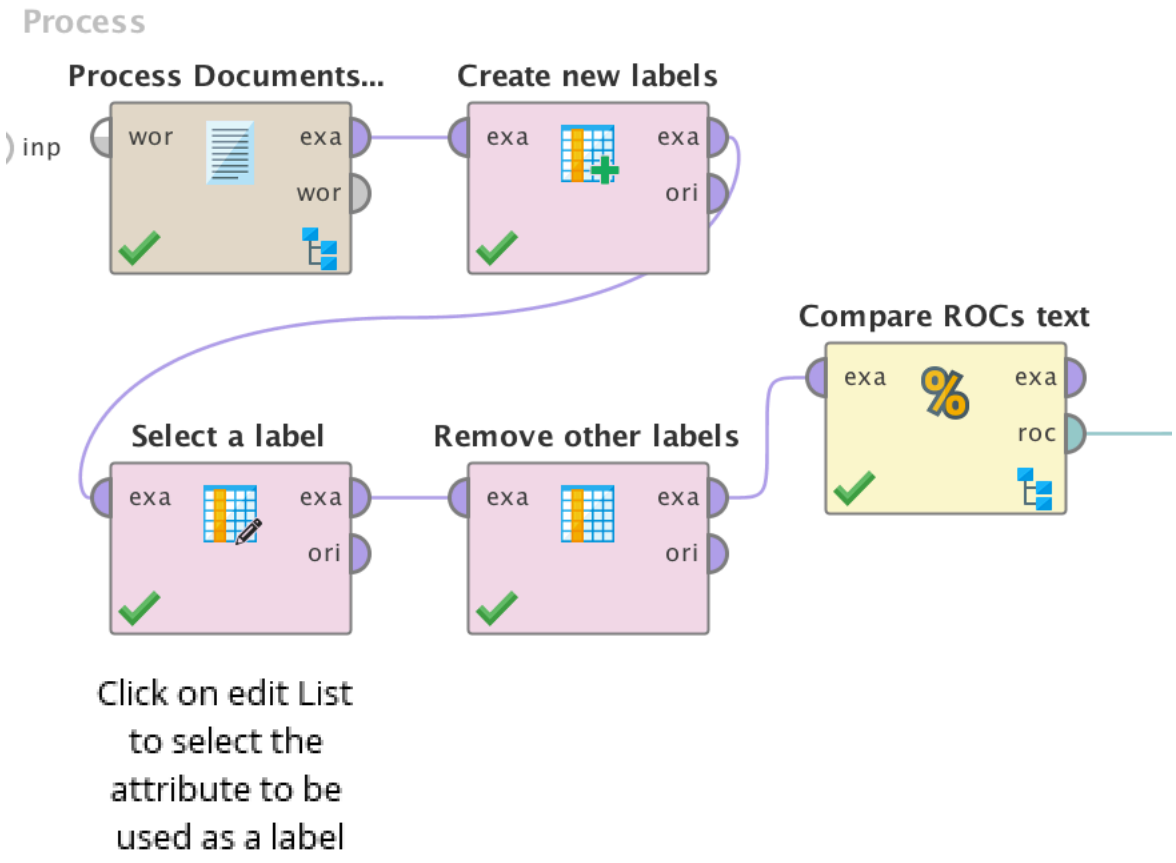
- ROC curves for a number of classifiers can be compared in Rapidminer used the operator Compare ROCs with classification algorithms nested within the operator:



Note: Only works for a binary class label.

ROC curves in RM

- See lect6-ROC.rmp on Moodle for an example of generating ROCs curve for our 15 texts



Exercise

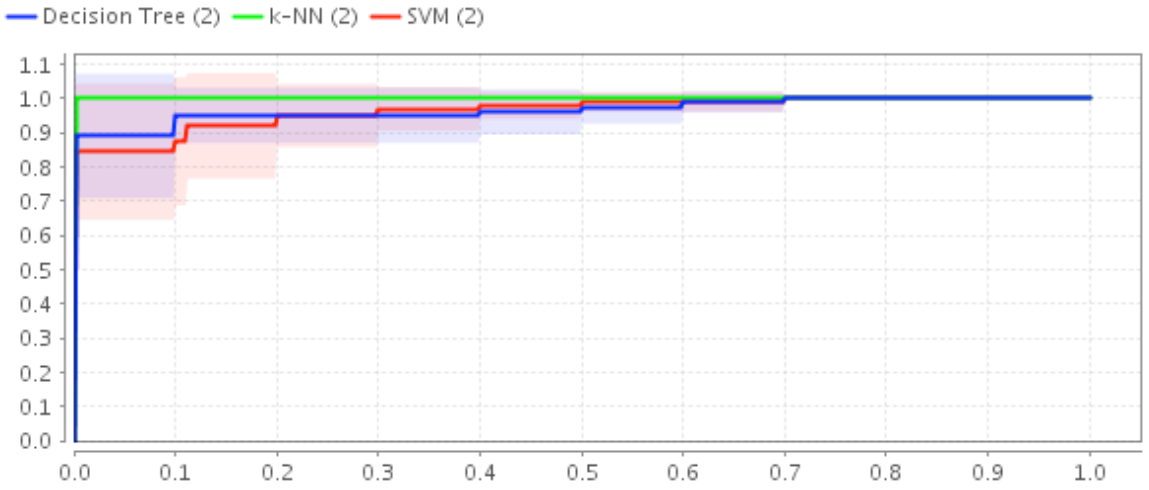
Fill the table below, draw the corresponding ROC curve, and interpret the results.

	Instances										
Actual class	+	+	+	-	-	-	+	-	-	+	
Confidence	0.3	0.4	0.5	0.8	0.8	0.8	0.9	0.9	0.9	1	1
TP	5										0
FP	5										0
TN	0										5
FN	0										5
TPR	1										0
FPR	1										0

What would the table be like for a perfect classifier?

Past exam question – Summer 2013

The following diagram compares the ROC curves for three text classifiers. Explain how to interpret the diagram. Your answer should make reference to the role of True Positive Rate and False Negative Rate in generating an ROC curve, and how that informs your interpretation of the curve. *10 marks*



Answer scheme for previous questions:

True positive rate: a true positive is a row that is classified correctly. True positive rate is the percentage or rows that should be classified in the positive class that are actually classified in the positive class. (**3 marks**)

False positive rate: a false positive is a row that is classified by the learner as being positive when in fact it is negative. False positive rate is the percentage or rows that should be classified in the negative class that are actually mis-classified as in the positive class. (**3 marks**)

How to interpret: best plot goes from (0,0) to (1,0) to (1,1), false positive rate remains at zero while true positive rises to 100%; A diagonal line from (0,0) to (1,1) is equivalent to random guessing; any curve below diagonal line is worse than a random guess; lines above that are an improvement on a random guess. The curve with the largest area under the curve is the best classifier, i.e. k-NN (**4 marks**)