MLRF Lecture 05

J. Chazalon, LRDE/EPITA, 2019

Classifier evaluation

Lecture 05 part 04

Data to test generalization

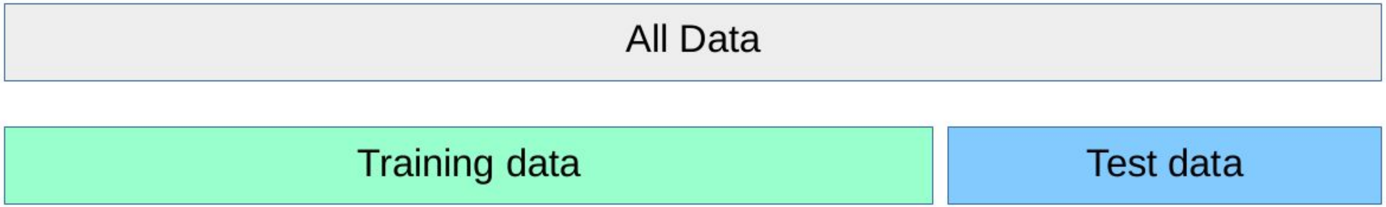
Bootstrap

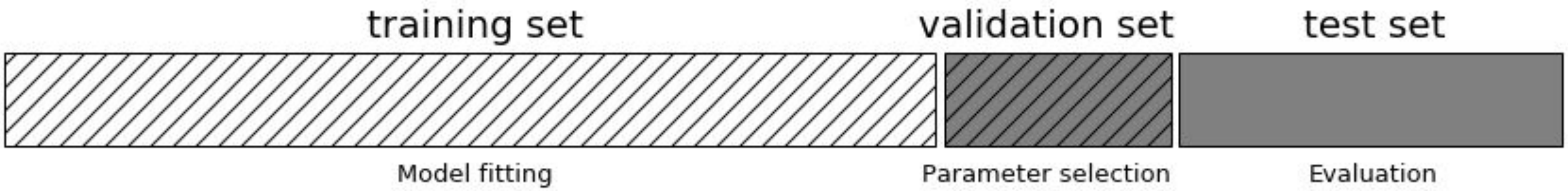
Draw randomly with replacement samples of n units from the training set.

We can estimate the variance of estimators we use in the classification rule.

Holdout

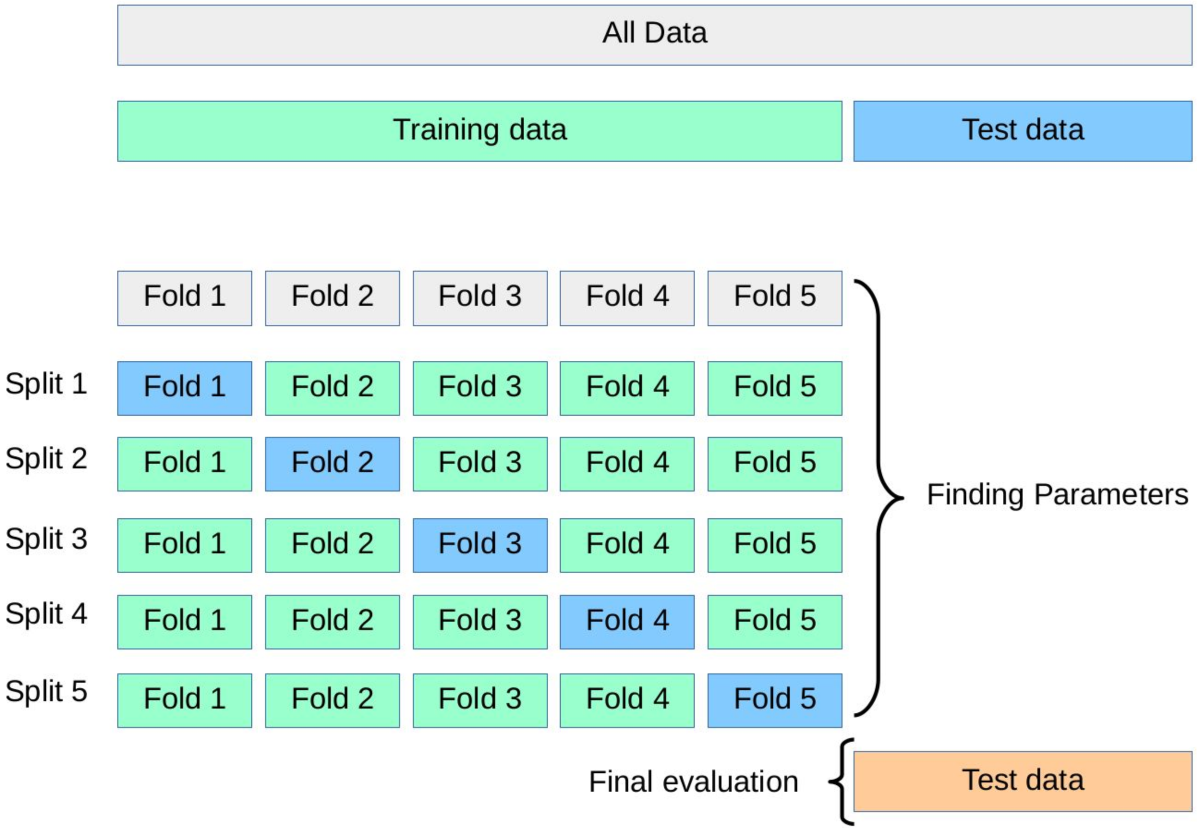
Just keep a part of the dataset for later validation/testing.



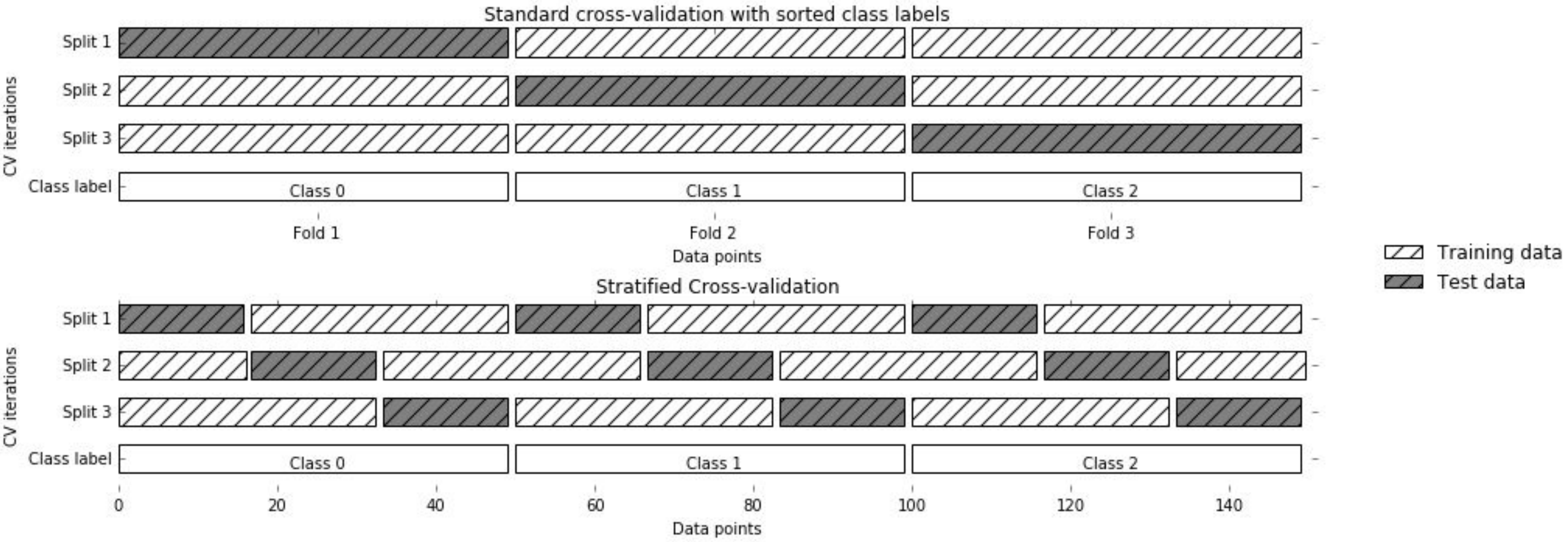


Cross validation

Cross validation with meta parameter tuning



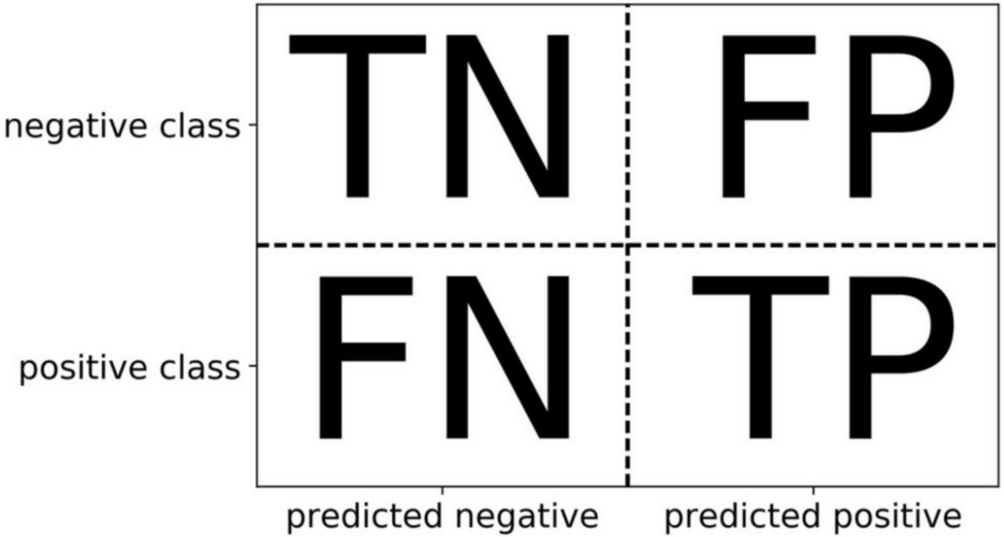
StratifiedKFold (best)



Stratified: Ensure relative class frequencies in each fold reflect relative class frequencies on the whole dataset.

Metrics

Confusion matrix and Accuracy

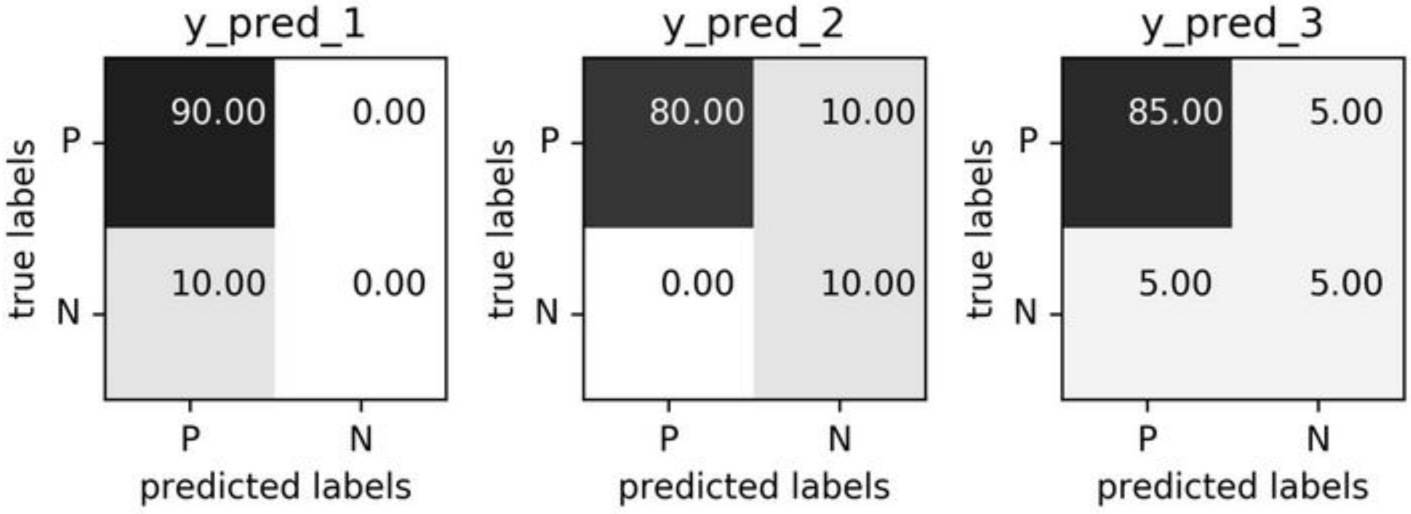


$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

From A. Müller

Problems with Accuracy

All the following classifiers have a 90% accuracy



Do all errors have the same cost?

From A. Müller

Precision, recall, F-score

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Positive Predicted Value (PPV)

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Sensitivity, coverage, true positive rate.

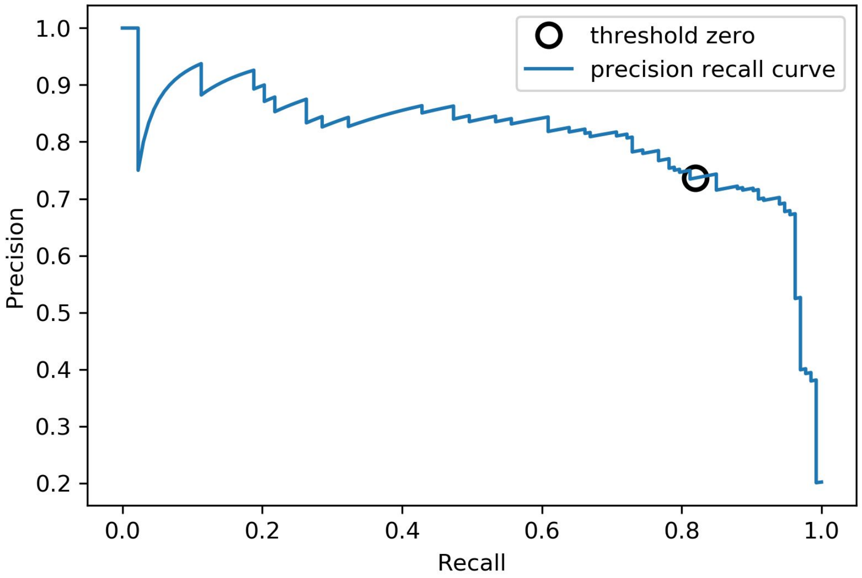
$$F = 2\,\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Harmonic mean of precision and recall

Plotting a Precision/Recall for classification data

For binary classification

Instead of $\hat{y} = \mathrm{argmax}_y\ p(y|x)$, take all possible thresholds for $p(y|x)$.



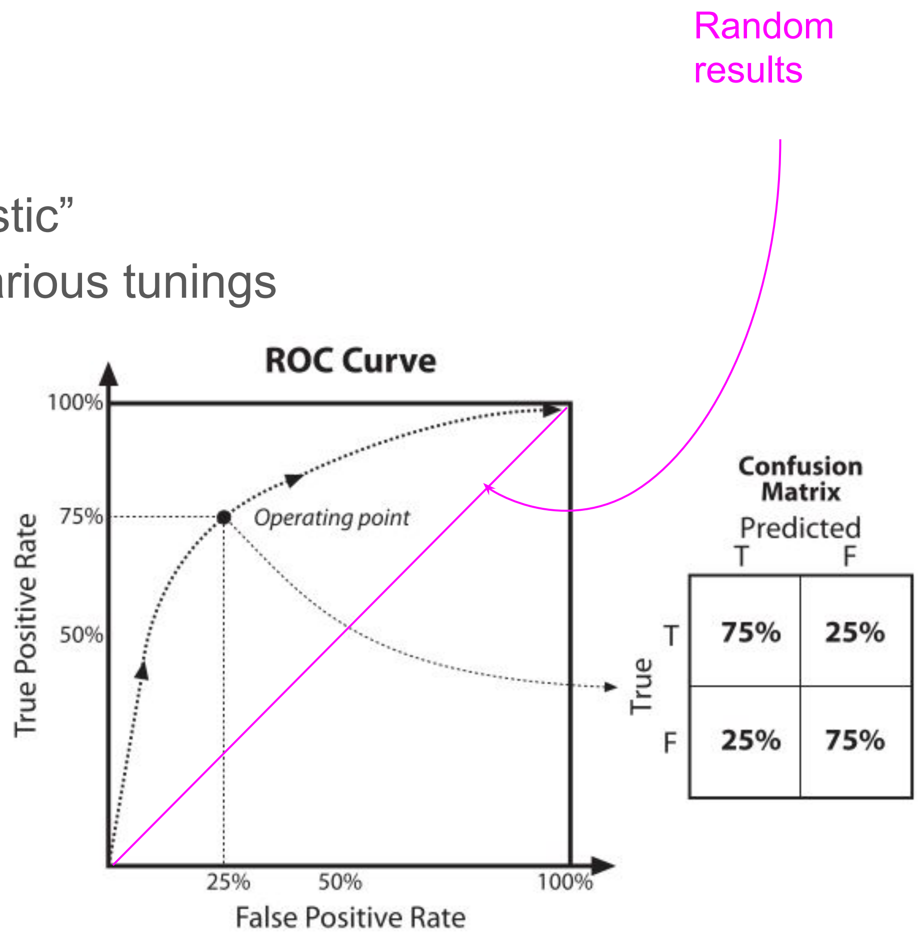
TPR, FPR, ROC

ROC: “Receiver Operating Characteristic”
Kind of signal/noise measure under various tunings

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} = \mathrm{recall}$$

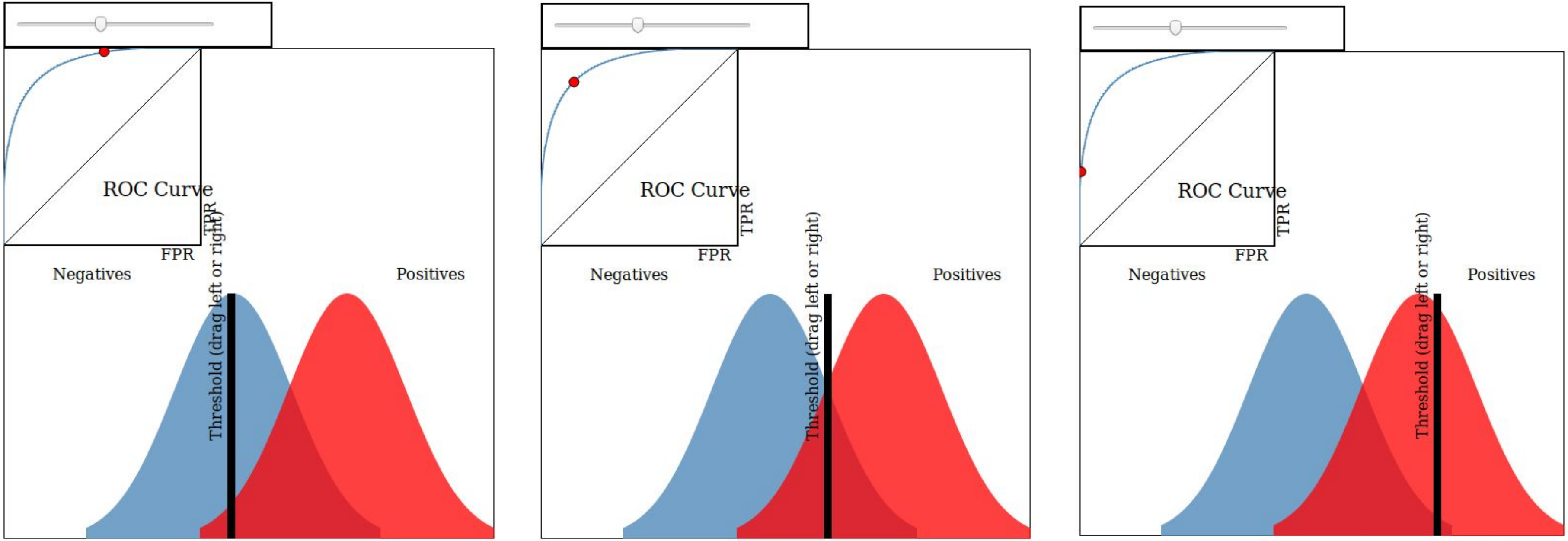
$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} = \mathrm{noise}$$

Common measure:
Area under the curve (AUC)



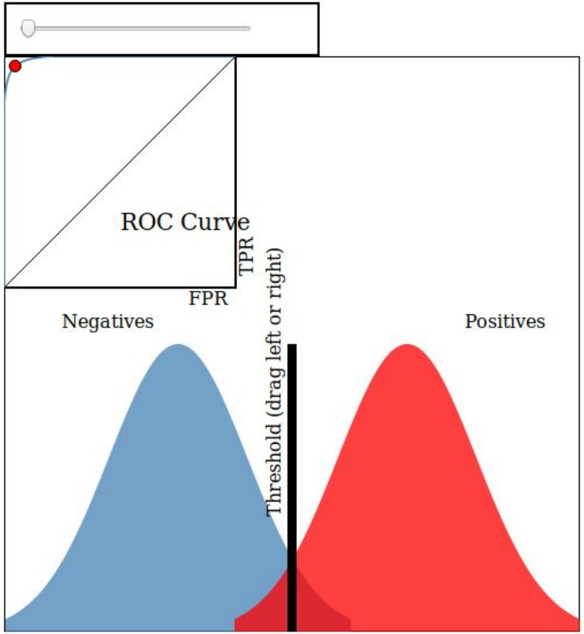
More about ROC curves: adjusting the threshold

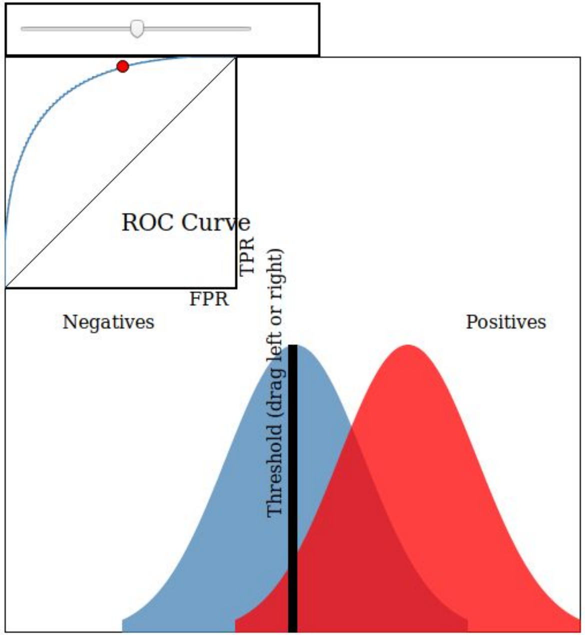
http://www.navan.name/roc/

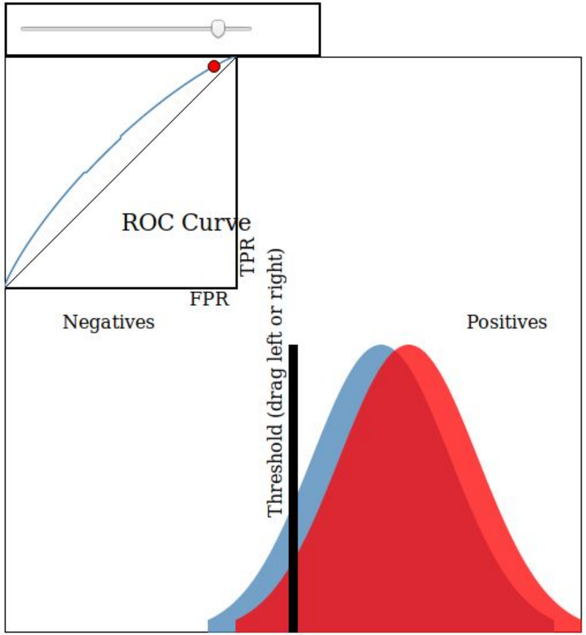


More about ROC curves: class overlap

http://www.navan.name/roc/

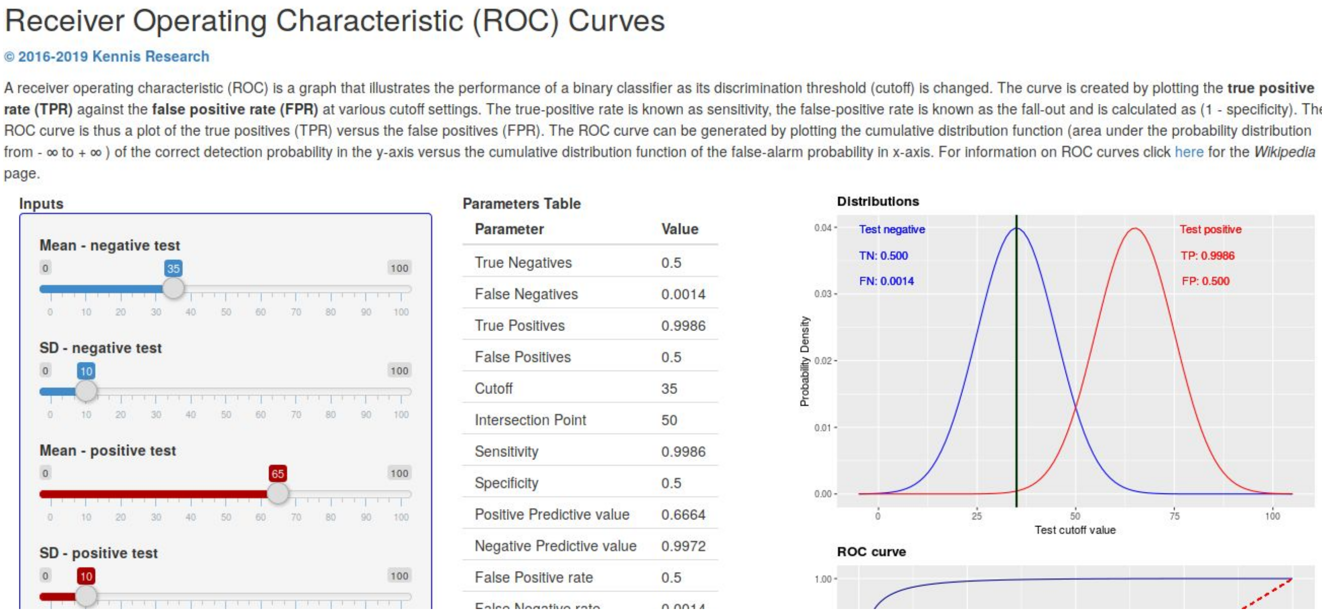






More about ROC curves

https://kennis-research.shinyapps.io/ROC-Curves/



Missing things

Missing things

Cost of misclassification

Multiclass classification evaluation

...