MLRF Lecture 05

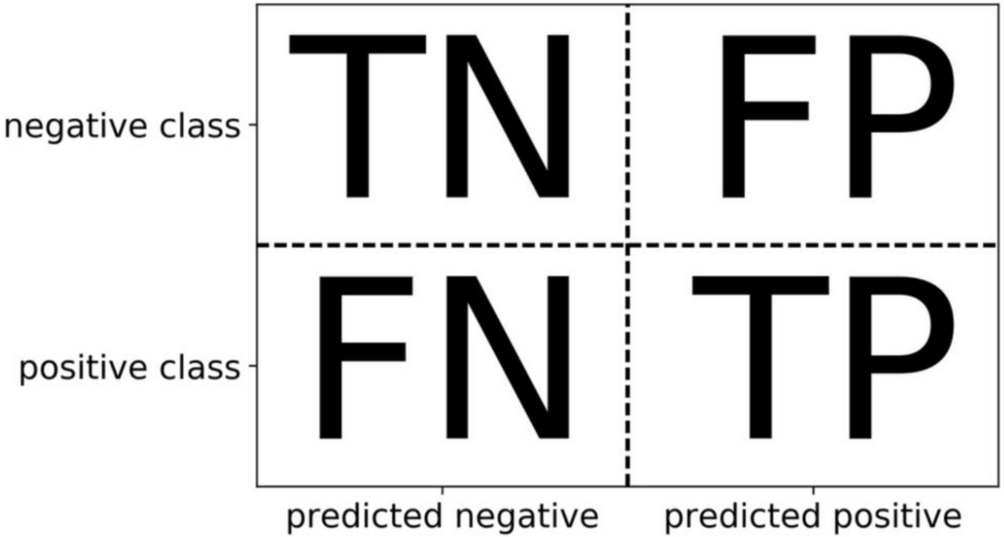
J. Chazalon, LRDE/EPITA, 2019

Classifier evaluation

Lecture 05 part 04

Metrics

Confusion matrix and Accuracy

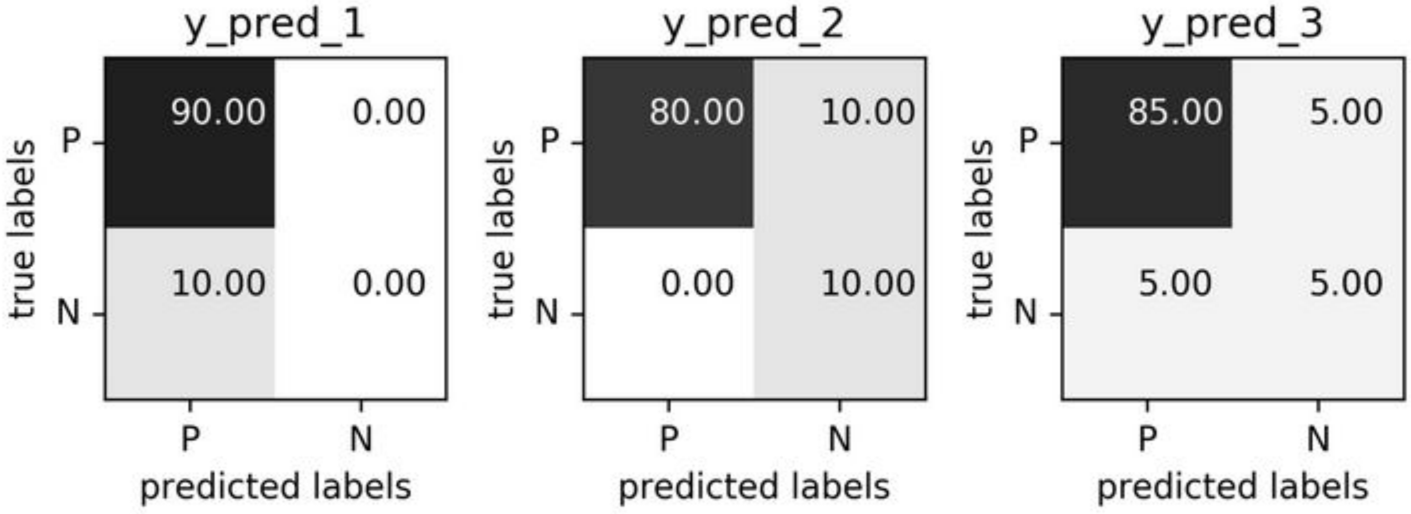


$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

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.

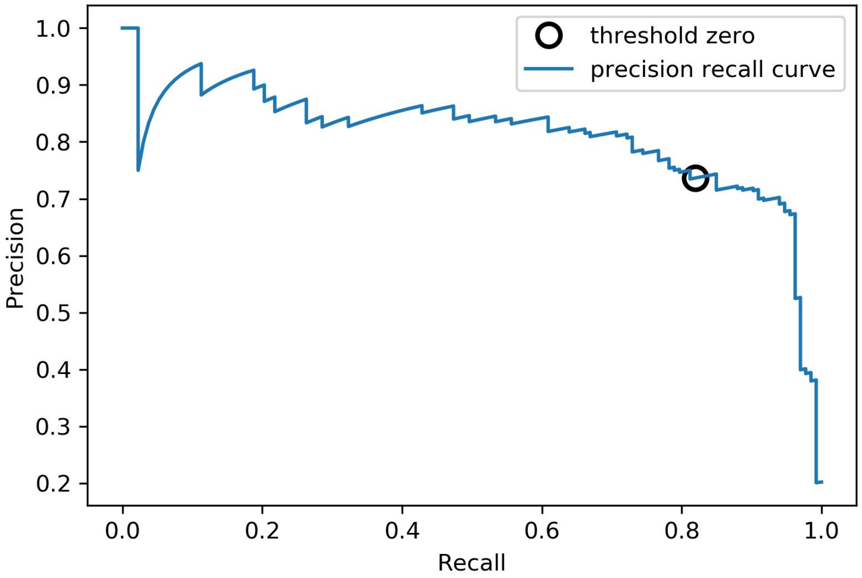
$$\text{F} = 2\,\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Harmonic mean of precision and recall

Plotting a Precision/Recall for classification data

For binary classification

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



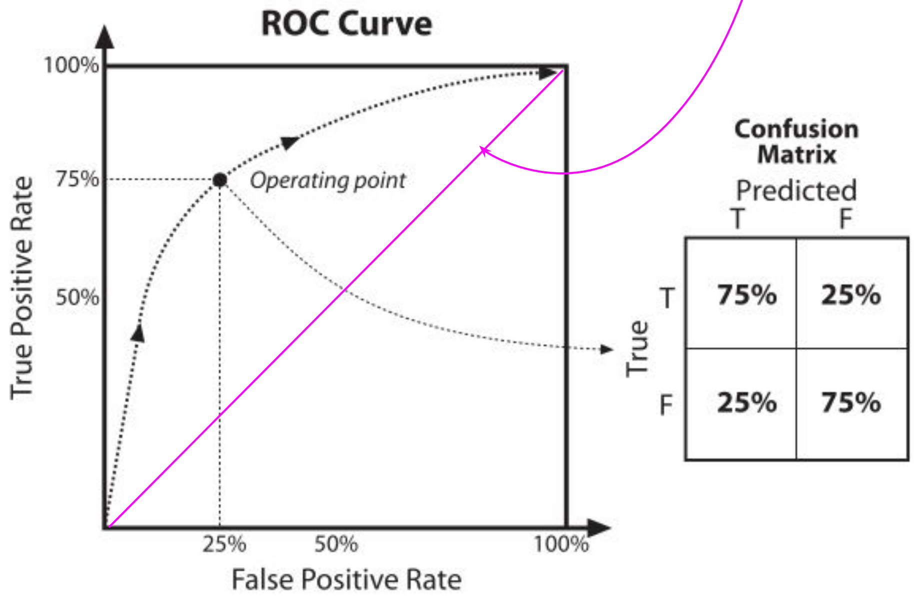
TPR, FPR, ROC

Random results

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

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

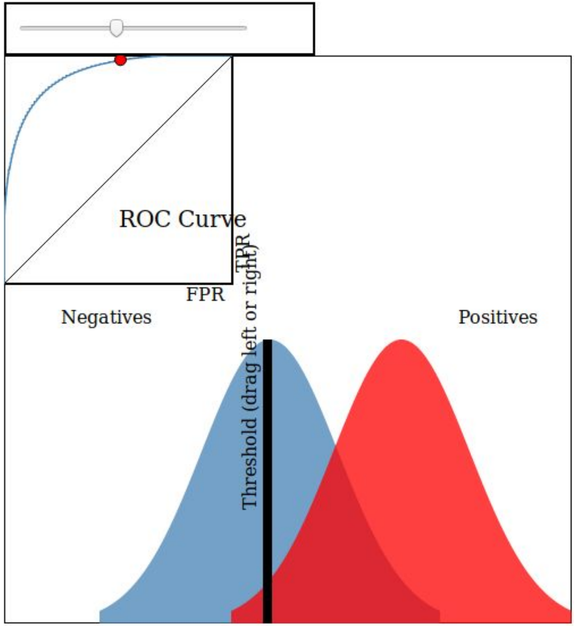
$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} = \mathrm{noise}$$

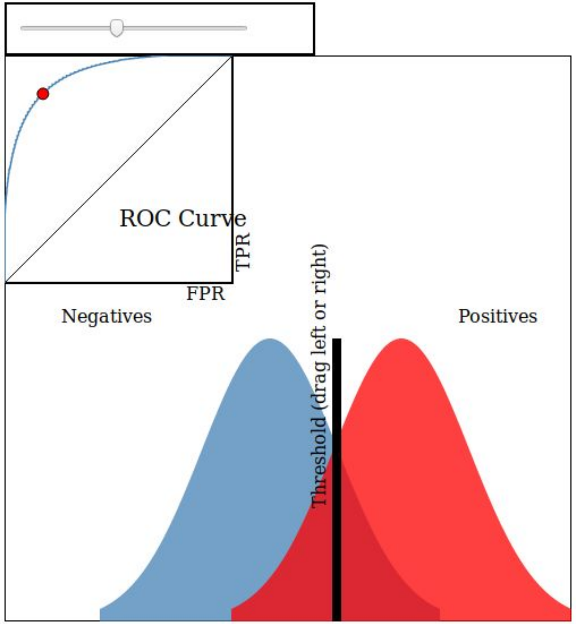


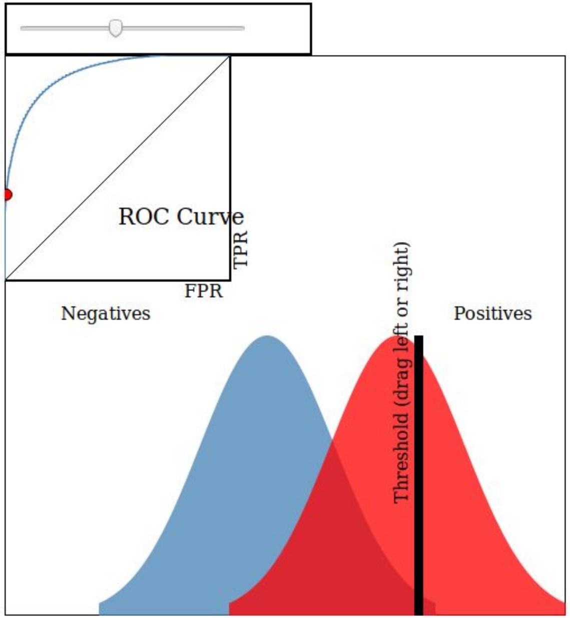
Common measure:
Area under the curve (AUC)

More about ROC curves: adjusting the threshold

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

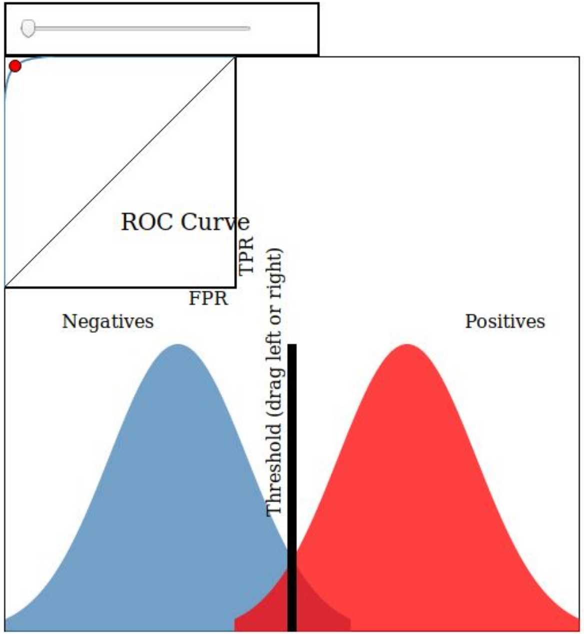


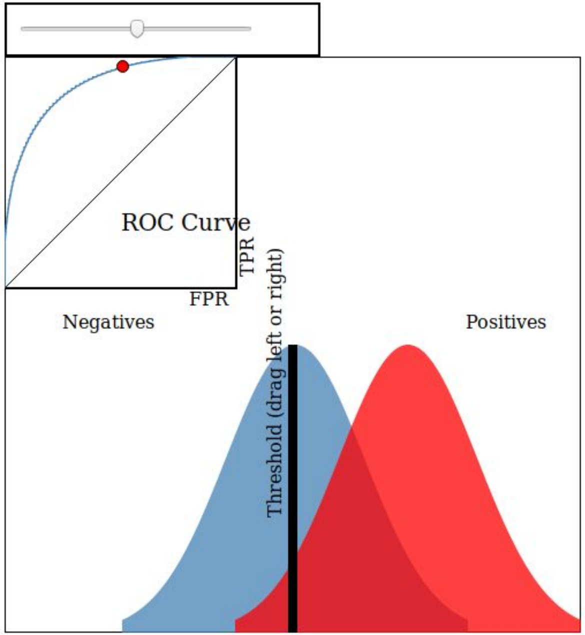


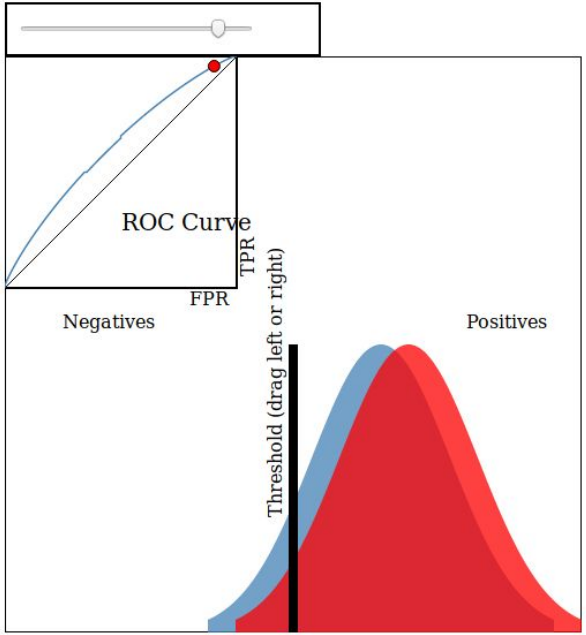


More about ROC curves: class overlap

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

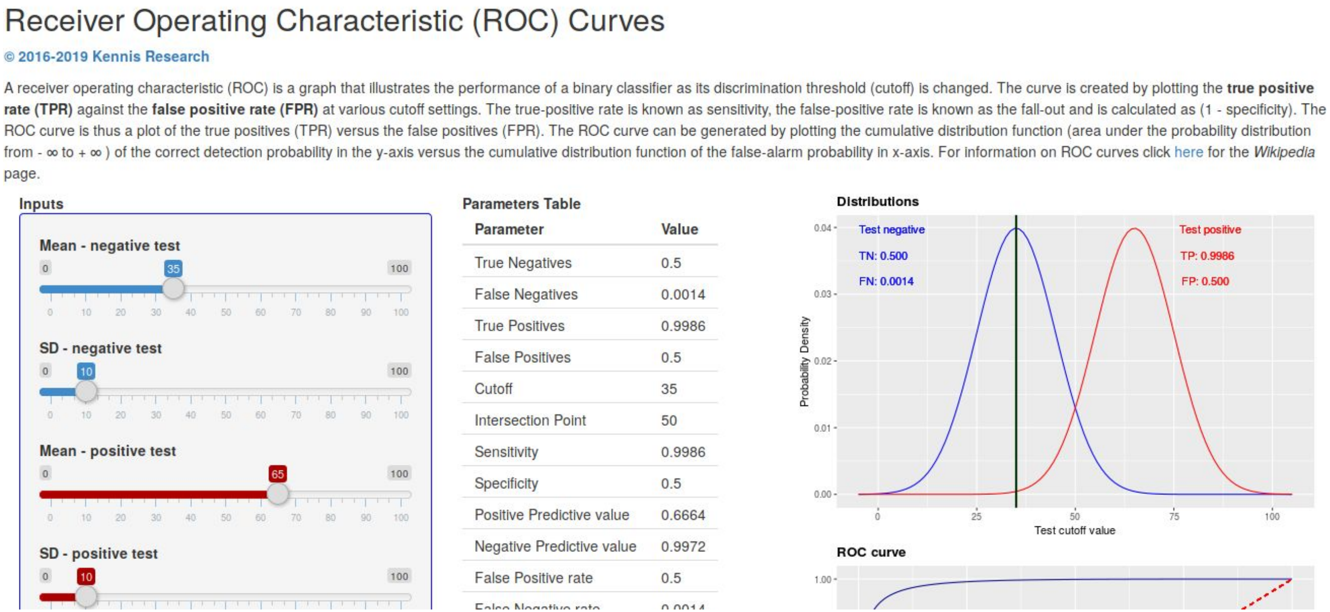






More about ROC curves

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



Receiver Operating Characteristic (ROC) Curves

© 2016-2019 Kennis Research

A receiver operating characteristic (ROC) is a graph that illustrates the performance of a binary classifier as its discrimination threshold (cutoff) is changed. The curve is created by plotting the **true positive rate (TPR)** against the **false positive rate (FPR)** at various cutoff settings. The true-positive rate is known as sensitivity, the false-positive rate is known as the fall-out and is calculated as (1 - specificity). The ROC curve is thus a plot of the true positives (TPR) versus the false positives (FPR). The ROC curve can be generated by plotting the cumulative distribution function (area under the probability distribution from - ∞ to + ∞) of the correct detection probability in the y-axis versus the cumulative distribution function of the false-alarm probability in x-axis. For information on ROC curves click here for the *Wikipedia* page.

Inputs

Mean - negative test: 35 (0–100)

SD - negative test: 10 (0–100)

Mean - positive test: 65 (0–100)

SD - positive test: 10 (0–100)

Parameters Table

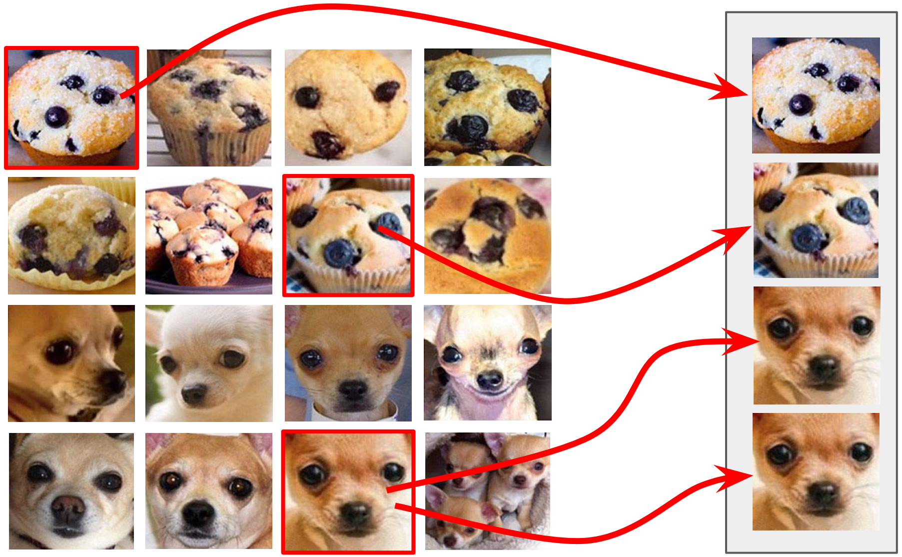
Parameter	Value
True Negatives	0.5
False Negatives	0.0014
True Positives	0.9986
False Positives	0.5
Cutoff	35
Intersection Point	50
Sensitivity	0.9986
Specificity	0.5
Positive Predictive value	0.6664
Negative Predictive value	0.9972
False Positive rate	0.5
False Negative rate	0.0014

Split the dataset to assess generalization performance

Bootstrap

Draw **randomly**, **with replacement** samples from the training set.

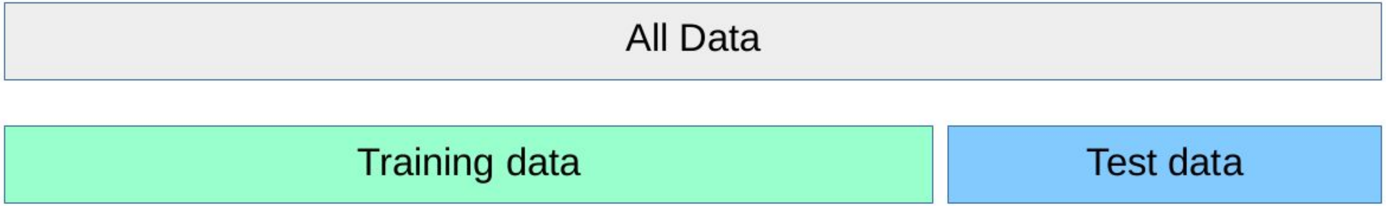
Enables us to estimate the variance of estimators we use in the classification rule.

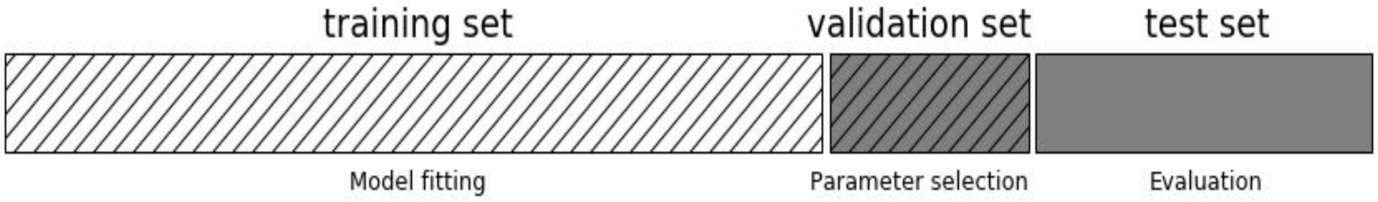


Efron and Tibshirani 1994

Holdout

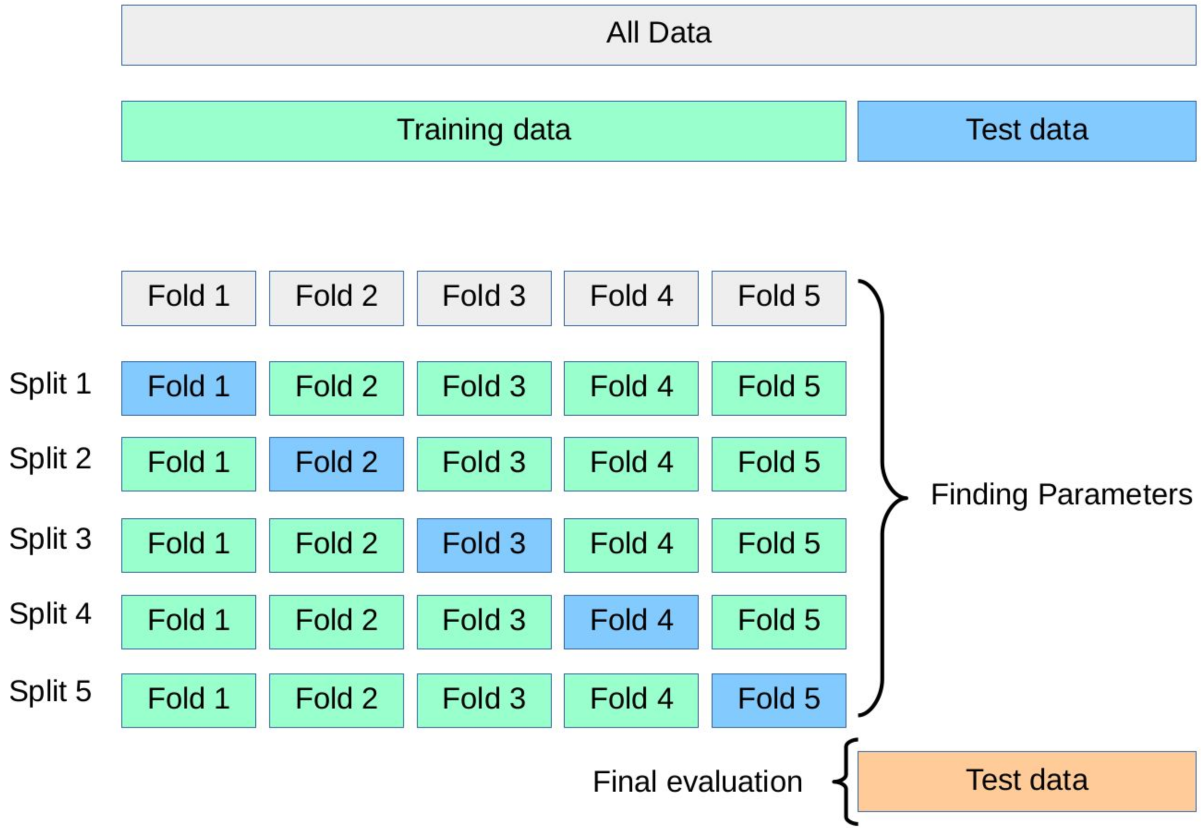
Just keep a part of the dataset for later validation/testing.



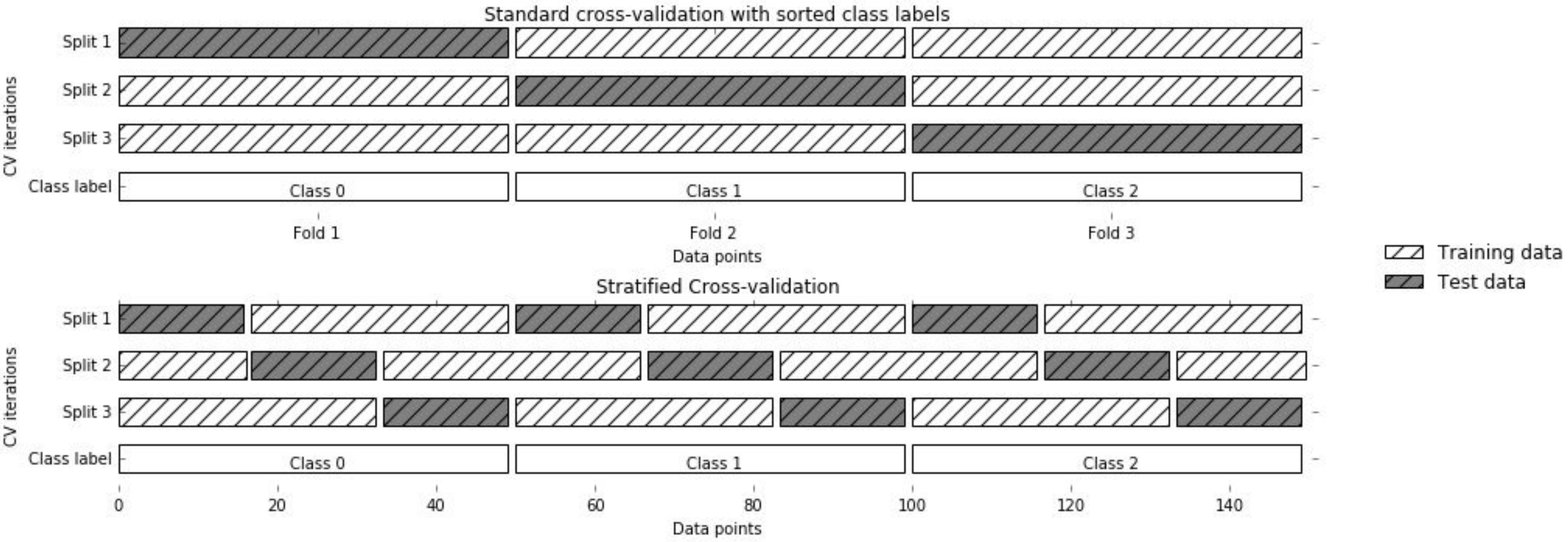


Cross validation

Cross validation with meta parameter tuning



StratifiedKFold (best)



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

Missing things

Missing things

Cost of misclassification

Multiclass classification evaluation

...