Model Selection

CSC 461: Machine Learning

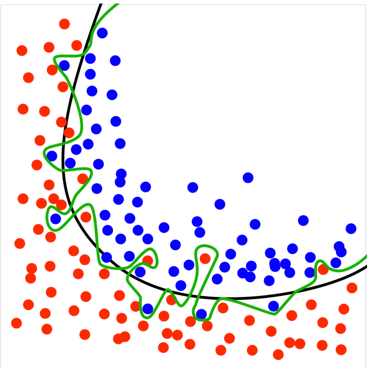
Fall 2020

Prof. Marco Alvarez
University of Rhode Island

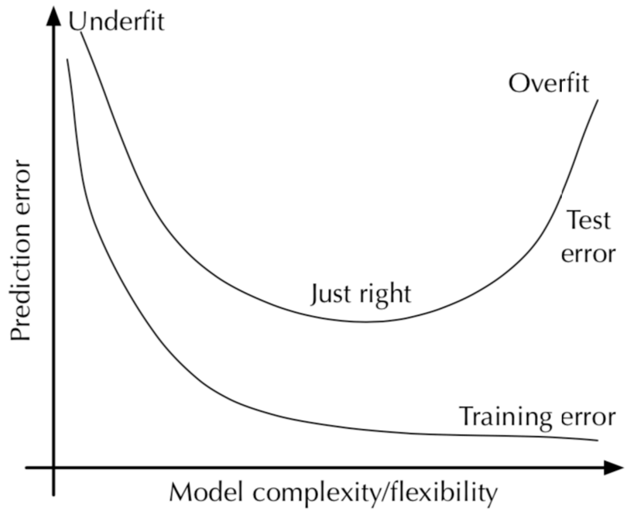
Overfitting

Overfitting

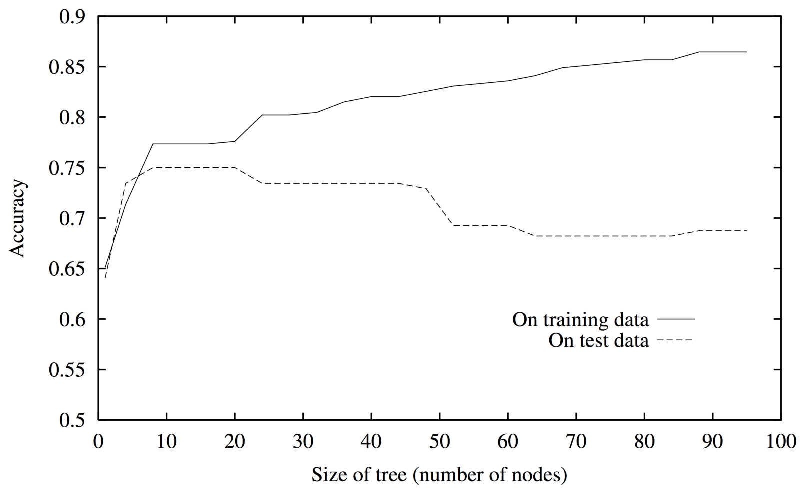
- Learning a **model** that "knows" the training data very well but does not **generalize**



Model complexity



Model complexity (DTs)

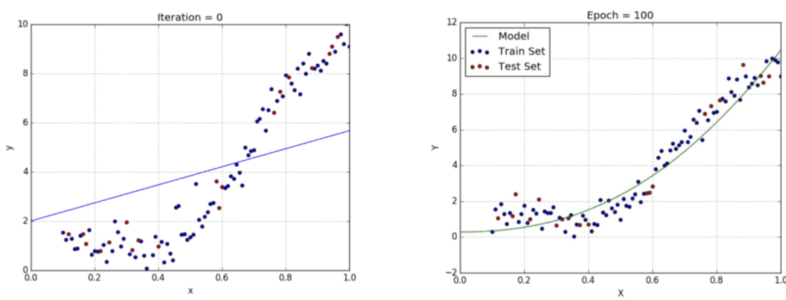


Machine Learning, Tom Mitchell, McGraw Hill, 1997

Overfitting

- Reasons
  - ✓ model is too **complex**
  - ✓ model is fitting **noise** present in the training data
  - ✓ training data is not a representative sample of the distribution
- How to prevent?
  - ✓ use more training data
  - ✓ use fewer features
  - ✓ regularize your model

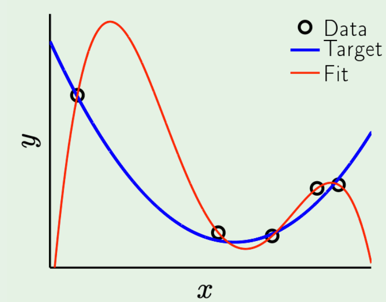
e.g. Curve Fitting



https://towardsdatascience.com/introduction-to-linear-regression-and-polynomial-regression-f8adc96f31cb

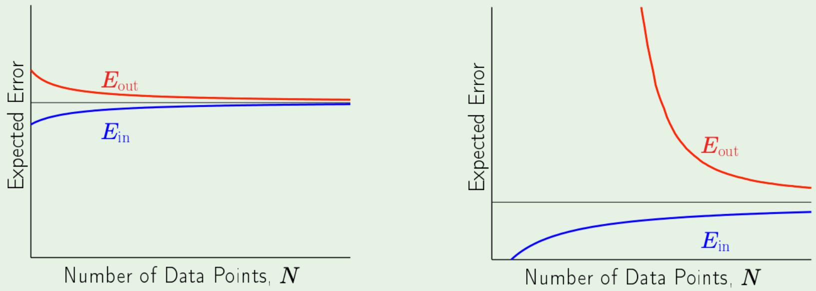
Overfitting

- Imagine the target function below …
  - ✓ 5 noisy data points and a 4th order polynomial fit
  - ✓ what can you say about training error? test error?



https://work.caltech.edu/lectures.html

Learning curves

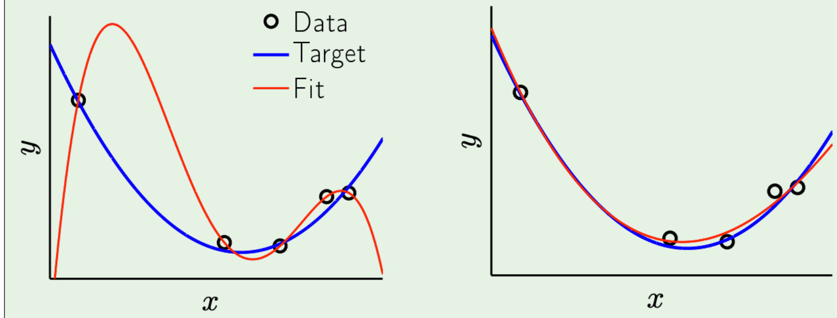


Simple Model

Complex Model

https://work.caltech.edu/lectures.html

Restricting the model



free fit

restrained fit

https://work.caltech.edu/lectures.html

Model Evaluation

Confusion matrix (2 classes)

	POSITIVE (1)	NEGATIVE (0)
POSITIVE (1)	TP	FN
NEGATIVE (0)	FP	TN

Actual values? Predicted values?

Evaluation metrics (2 classes)

sensitivity, recall, hit rate, or true positive rate (TPR)

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} = 1 - \mathrm{FNR}$$

specificity, selectivity or true negative rate (TNR)

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{N}} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} = 1 - \mathrm{FPR}$$

precision or positive predictive value (PPV)

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} = 1 - \mathrm{FDR}$$

negative predictive value (NPV)

$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FN}} = 1 - \mathrm{FOR}$$

miss rate or false negative rate (FNR)

$$\mathrm{FNR} = \frac{\mathrm{FN}}{\mathrm{P}} = \frac{\mathrm{FN}}{\mathrm{FN}+\mathrm{TP}} = 1 - \mathrm{TPR}$$

fall-out or false positive rate (FPR)

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{N}} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} = 1 - \mathrm{TNR}$$

https://en.wikipedia.org/wiki/Confusion_matrix

Evaluation metrics (2 classes)

accuracy (ACC)

$$\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{P}+\mathrm{N}} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$$

F1 score

is the harmonic mean of precision and sensitivity

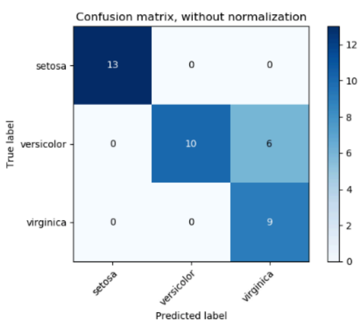
$$\mathrm{F}_1 = 2 \cdot \frac{\mathrm{PPV} \cdot \mathrm{TPR}}{\mathrm{PPV}+\mathrm{TPR}} = \frac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$$

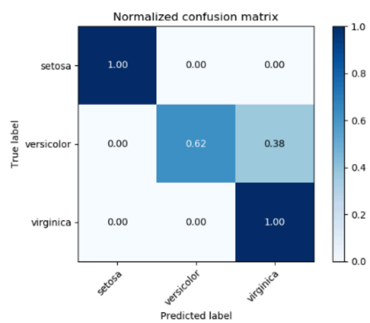
Matthews correlation coefficient (MCC)

$$\mathrm{MCC} = \frac{\mathrm{TP}\times\mathrm{TN} - \mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$$

https://en.wikipedia.org/wiki/Confusion_matrix

Confusion matrix

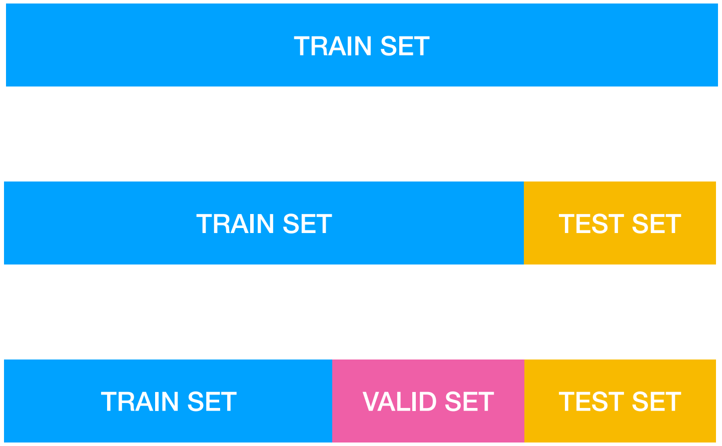


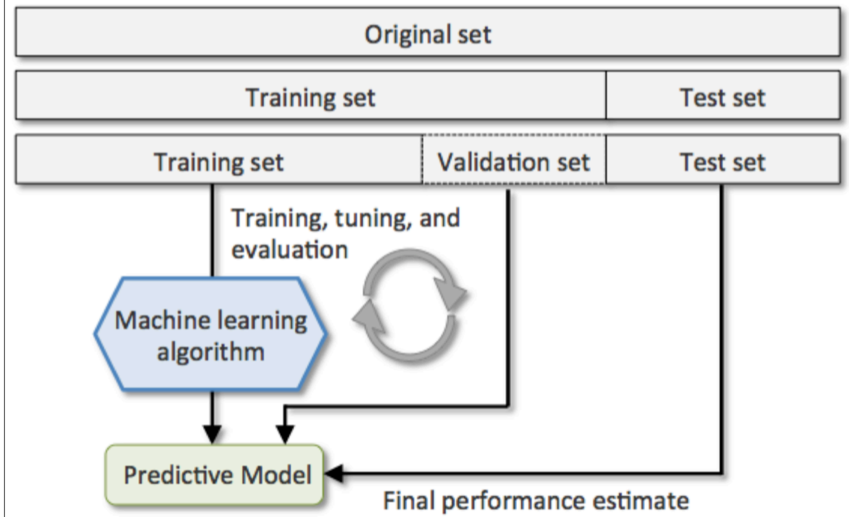


https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html

Cross Validation

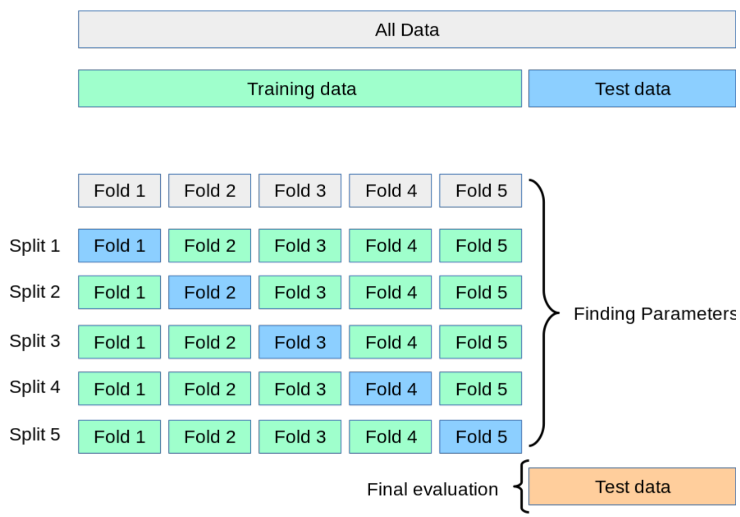
Train, validation, and test sets





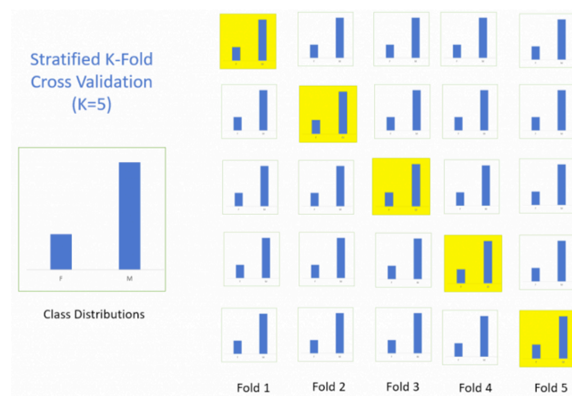
from INFO-4604: Applied Machine Learning, Fall 2017, Michael Paul, Univ. of Colorado

What is k-fold Cross Validation?



https://scikit-learn.org/stable/modules/cross_validation.html

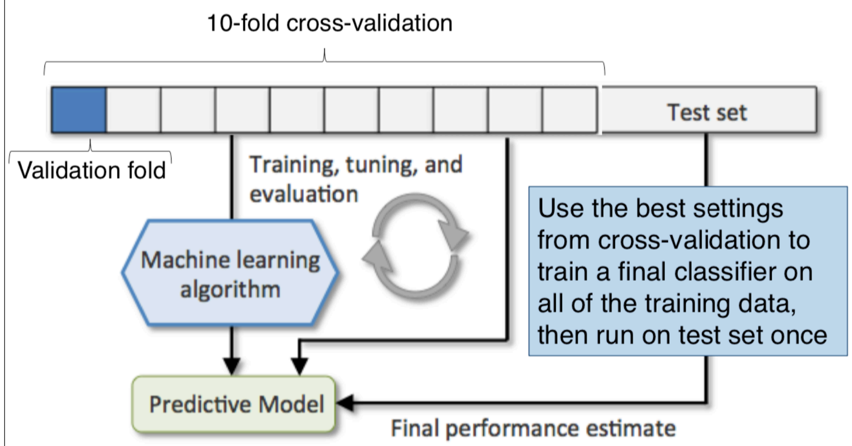
Stratified cross validation



Stratified cross validation aims at having the same class distribution in each fold

https://towardsdatascience.com/cross-validation-explained-evaluating-estimator-performance-e51e5430ff85

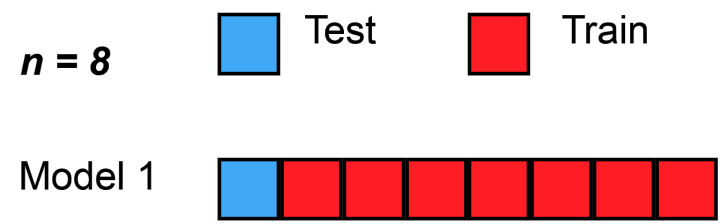
Using Cross-Validation



from INFO-4604: Applied Machine Learning, Fall 2017, Michael Paul, Univ. of Colorado

Leave-One-Out CV

- Special case of CV when **k = n**
- Can be expensive for large **n**



https://en.wikipedia.org/wiki/Cross-validation_(statistics)