

Overfitting, Model Selection

CSC 461: Machine Learning

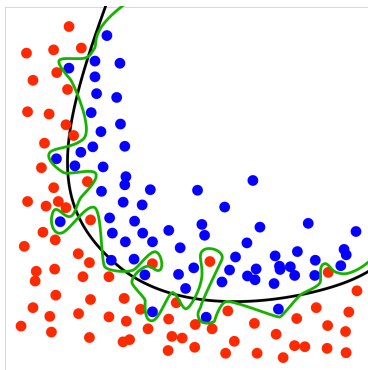
Fall 2021

Prof. Marco Alvarez
University of Rhode Island

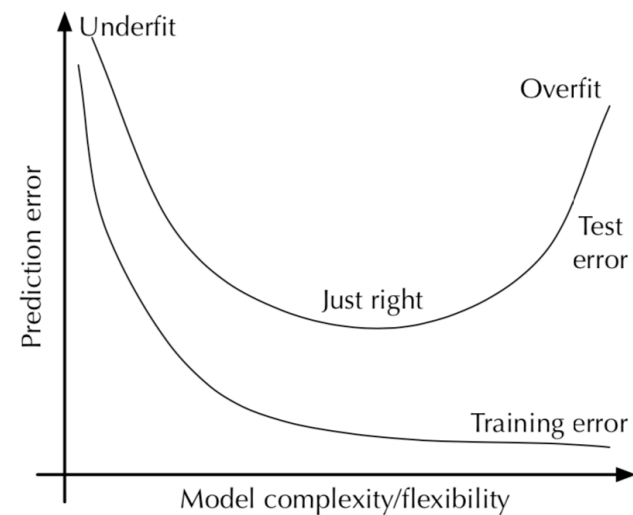
Overfitting

Overfitting

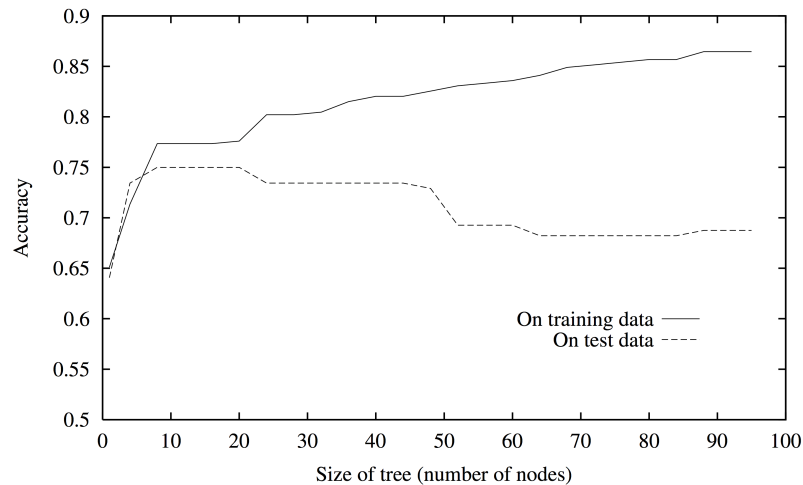
- Learning a **model** that “knows” the training data very well but does not **generalize**



Model complexity



Model complexity (DTs)



Overfitting

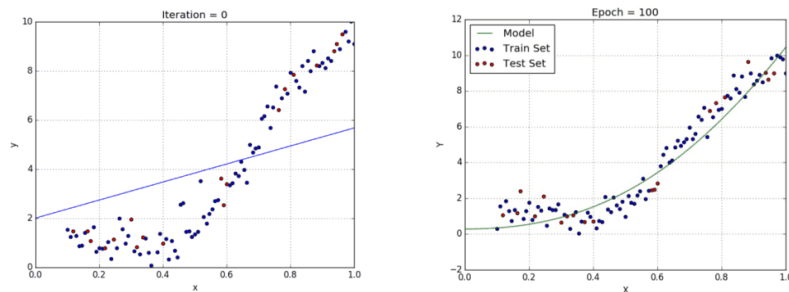
► Reasons

- ✓ 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

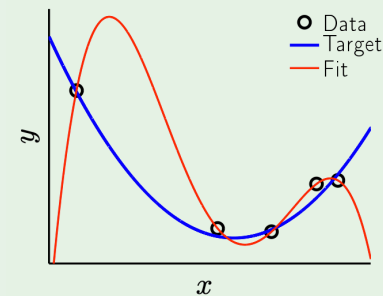
e.g. Curve Fitting



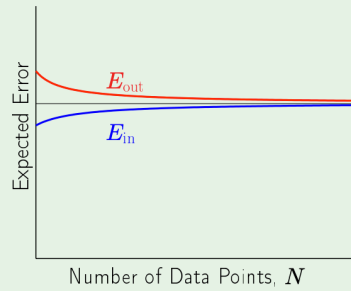
Overfitting

► Imagine the target function below ...

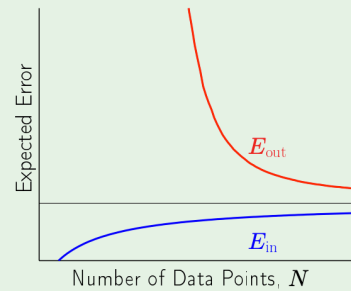
- ✓ 5 noisy data points and a 4th order polynomial fit
- ✓ what can you say about training error? test error?



Number of data instances



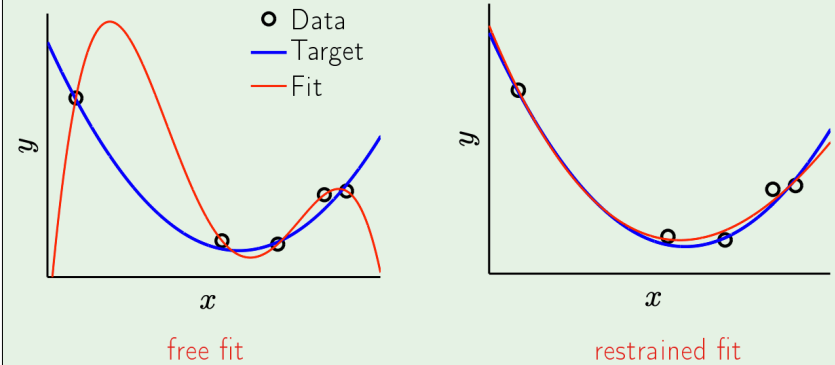
Simple Model



Complex Model

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

Restricting the model



<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)

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

specificity, selectivity or true negative rate (TNR)

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

precision or positive predictive value (PPV)

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

negative predictive value (NPV)

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

miss rate or false negative rate (FNR)

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

fall-out or false positive rate (FPR)

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

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

Evaluation metrics (2 classes)

accuracy (ACC)

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

F1 score

is the **harmonic mean** of **precision** and **sensitivity**

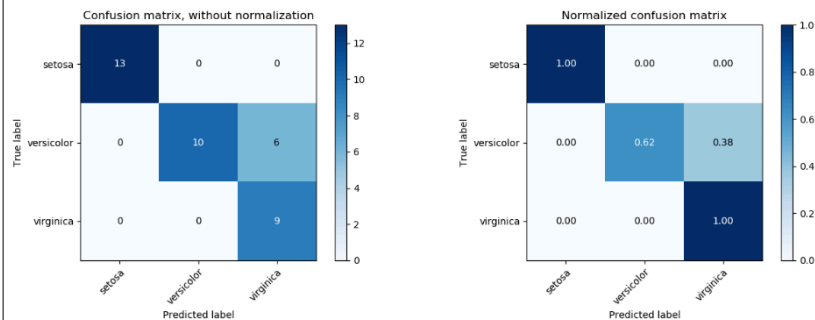
$$F_1 = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$$

Matthews correlation coefficient (MCC)

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

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

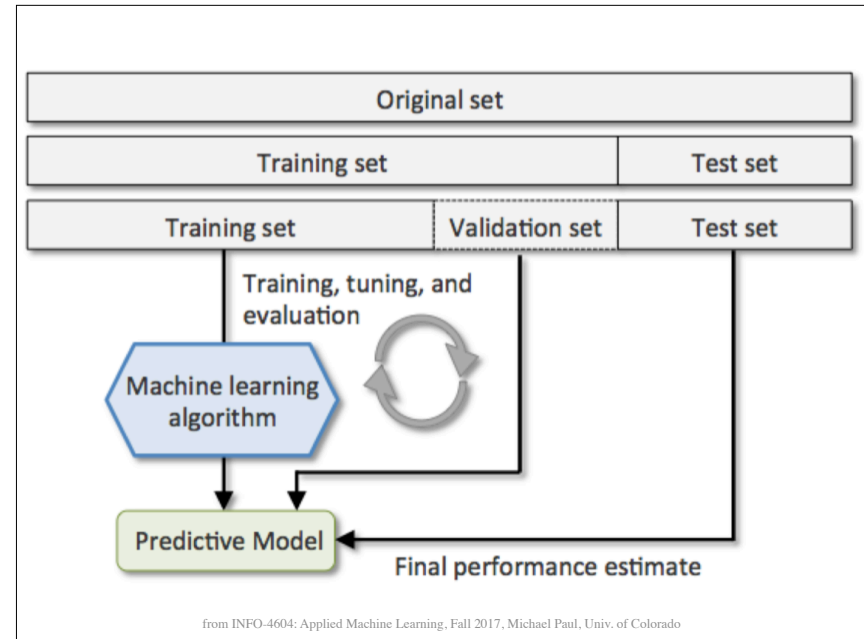
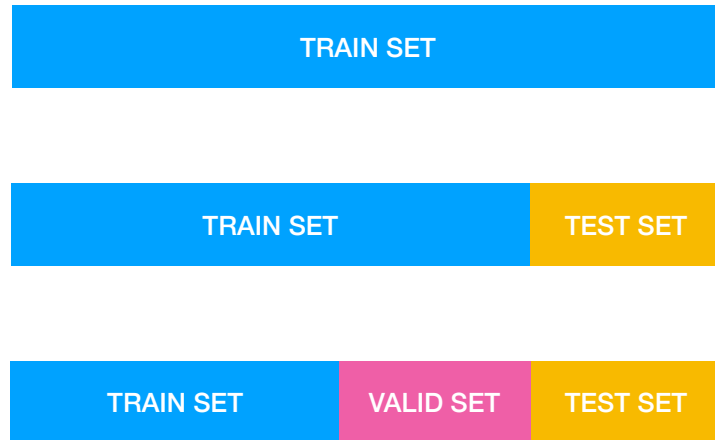
Confusion matrix



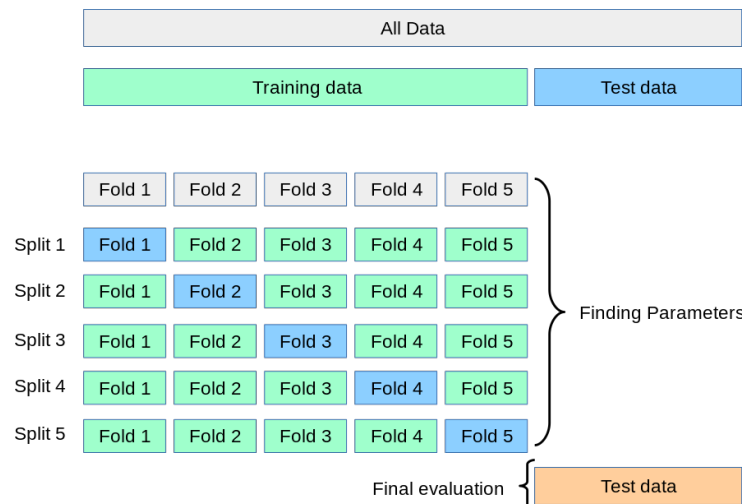
https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html

Cross Validation

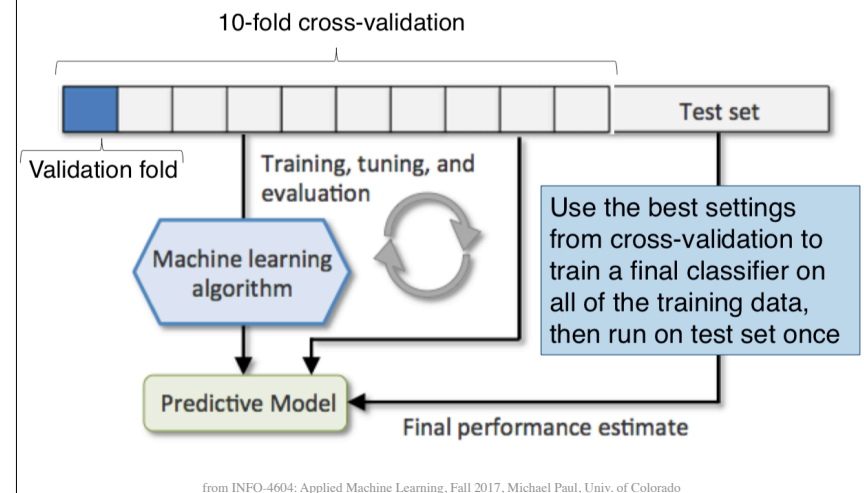
Train, validation, and test sets



What is k-fold Cross Validation?

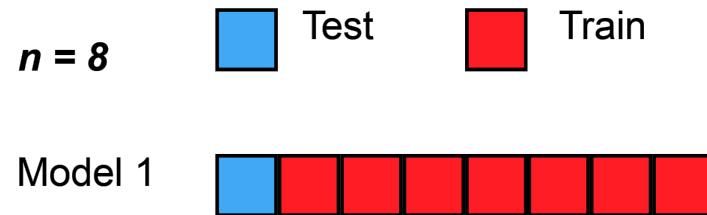


Using Cross-Validation



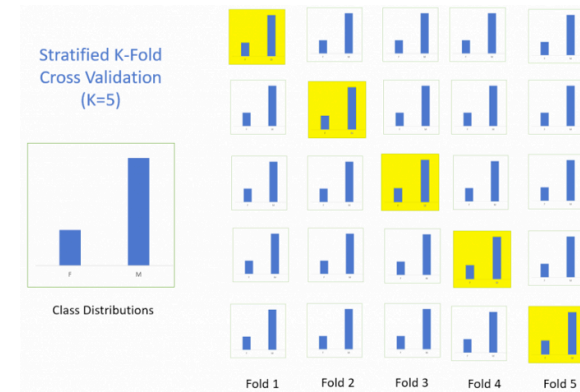
Leave-One-Out CV

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



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

Stratified cross validation



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

<https://towardsdatascience.com/cross-validation-explained-evaluating-estimator-performance-e51e5430f85>