CSC 461: Machine Learning

Fall 2024

Model Selection

Prof. Marco Alvarez, Computer Science
University of Rhode Island

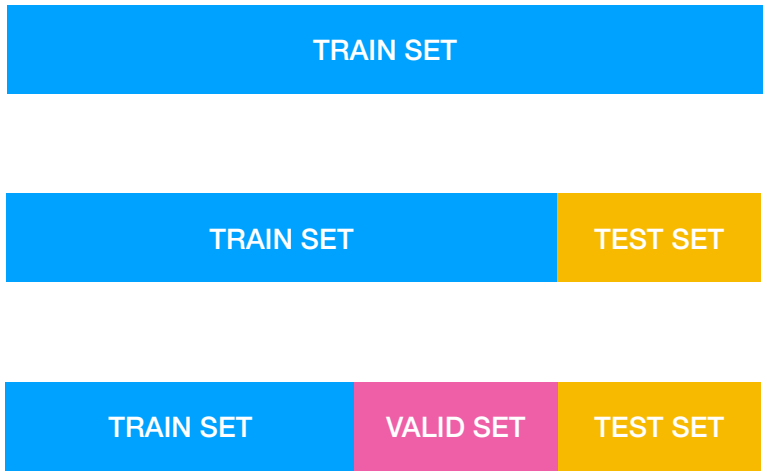
Model selection

- Model selection
  - process of choosing the best model from a set of candidate models
  - process involves evaluating and comparing different models based on their performance (use evaluation metrics)
- Hyperparameter tuning
  - optimizing model-specific parameters
  - techniques: grid search, random search
- Goals of selecting the best model
  - enhancing generalization
  - preventing overfitting/underfitting

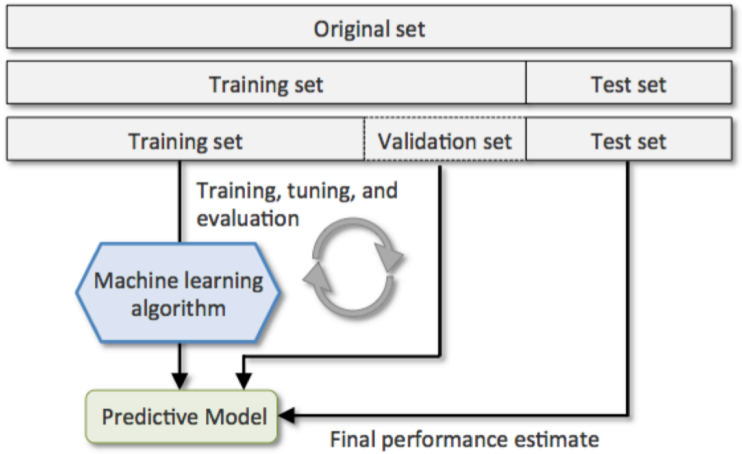
Overfitting and underfitting

- Overfitting
  - a model learns the training data too well, leading to poor generalization performance on unseen data
- Underfitting
  - a model is too simple to capture the underlying patterns in the data

Train, validation, and test

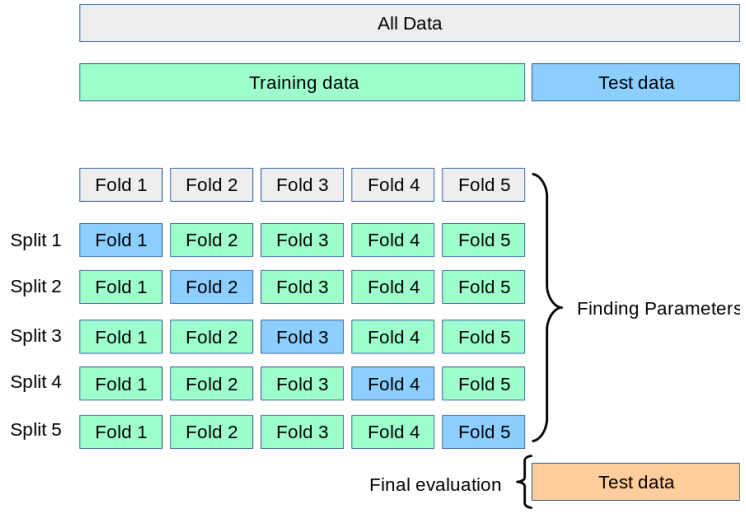


Holdout validation



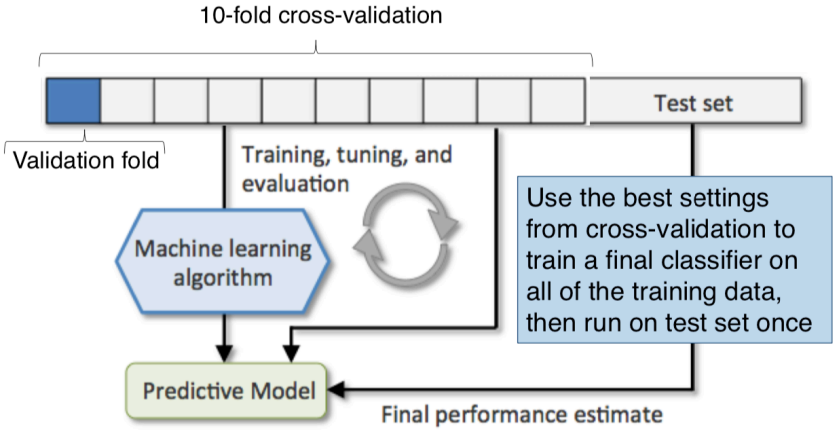
from INFO-4604: Applied Machine Learning, Fall 2017, Michael Paul, Univ. of Colorado

k-Fold cross validation



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

k-fold cross-validation



Use the best settings from cross-validation to train a final classifier on all of the training data, then run on test set once

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

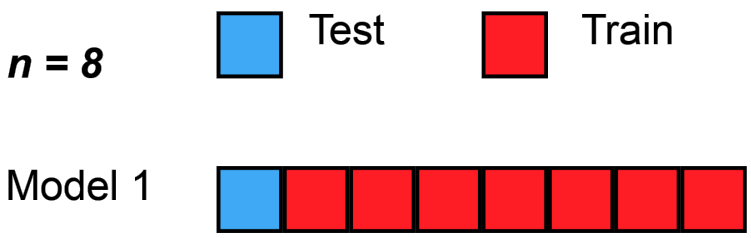
Stratified k-fold cross-validation

Stratified CV aims at keeping the same class distribution within each fold

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

Leave-one-out cross-validation

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



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