Tutorial 1

Bias and Variance

ISLR 1-2, ESL 1-2

UNIVERSITÄT DES SAARLANDES

CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

Lecture Recap

- Common Paradigms
  - Supervised vs. Unsupervised Learning
  - Regression vs. Classification
  - Prediction vs. Inference
  - Accuracy vs. Interpretability
- Bias-Variance Tradeoff
- K-nearest neighbor classifier

Lecture Recap

- Common Paradigms
  - Supervised vs. Unsupervised Learning
  - Regression vs. Classification
  - Prediction vs. Inference
  - Accuracy vs. Interpretability

- Bias-Variance Tradeoff

- K-nearest neighbor classifier

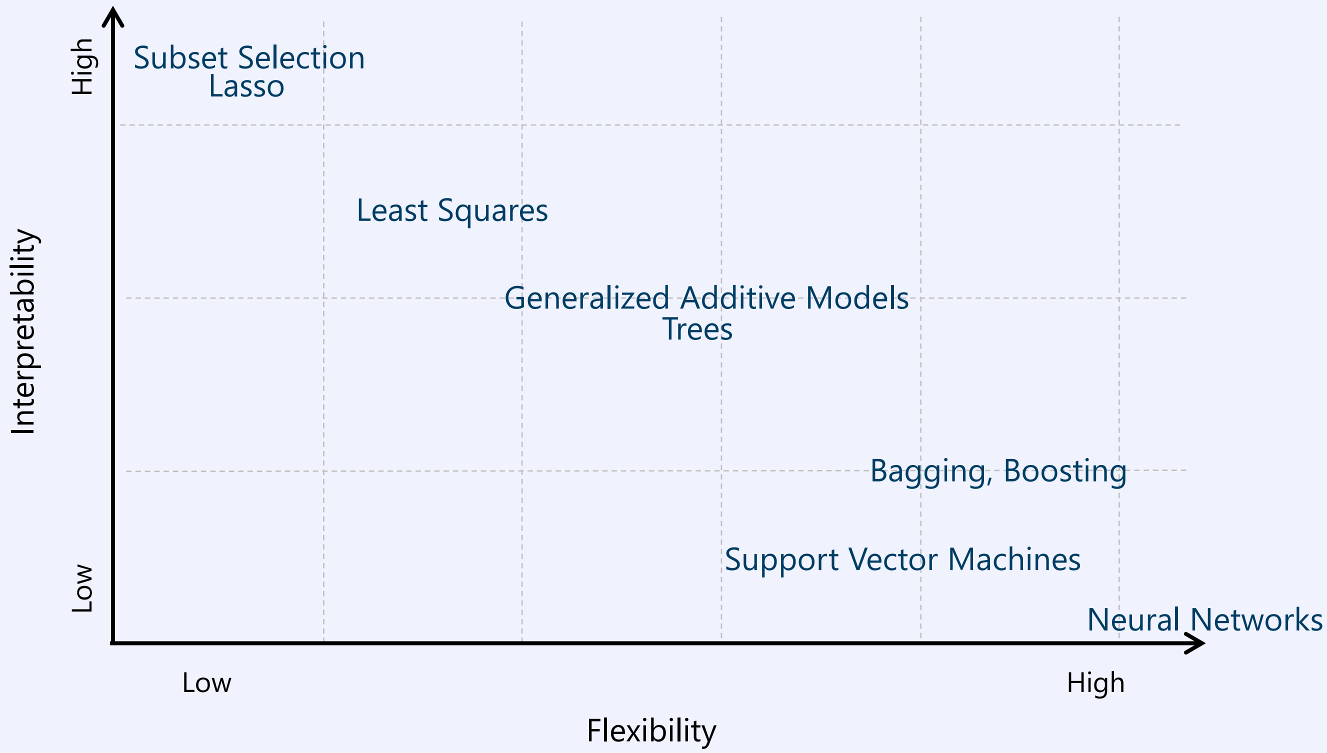
Lecture Recap

- Common Paradigms
  - Supervised vs. Unsupervised Learning
  - Regression vs. Classification
  - Prediction vs. Inference
  - Accuracy vs. Interpretability

- Bias-Variance Tradeoff

- K-nearest neighbor classifier

Accuracy vs. Interpretability



Assessing model accuracy

In regression, we assess the quality of fit by mean squared error (MSE)

- over training data, it is defined as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2$$

  which we typically refer to as the training error

- we are generally more interested in the error over unseen data

$$avg\left(\hat{f}(x_0) - y_0\right)^2$$

  which we typically call the test error or generalization error

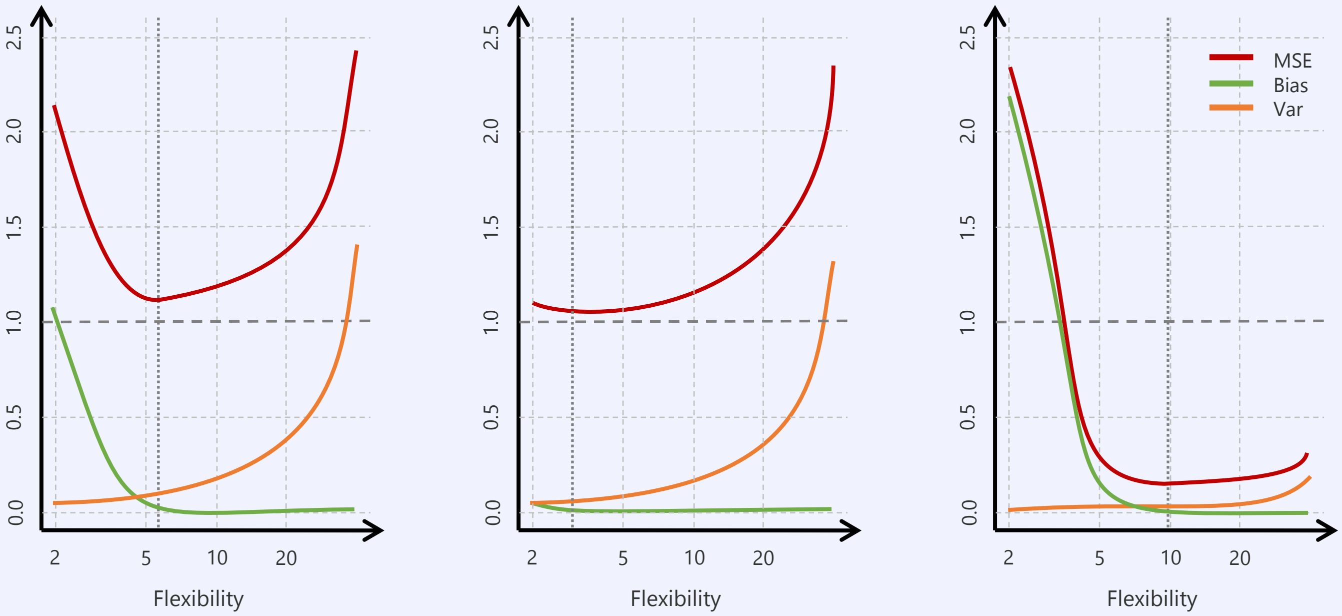
Lecture Recap

- Common Paradigms
  - Accuracy vs. Interpretability
  - Supervised vs. Unsupervised Learning
  - Regression vs. Classification
  - Prediction vs. Inference

- Bias-Variance Tradeoff

- K-nearest neighbor classifier

Bias-Variance Decomposition



Synthetic data example 1
Moderately nonlinear function

Synthetic data example 2
Almost linear function

Synthetic data example 3
Highly nonlinear function

Lecture Recap

- Common Paradigms
  - Accuracy vs. Interpretability
  - Supervised vs. Unsupervised Learning
  - Regression vs. Classification
  - Prediction vs. Inference
- Bias-Variance Tradeoff
- K-nearest neighbor classifier

Nearest Neighbors

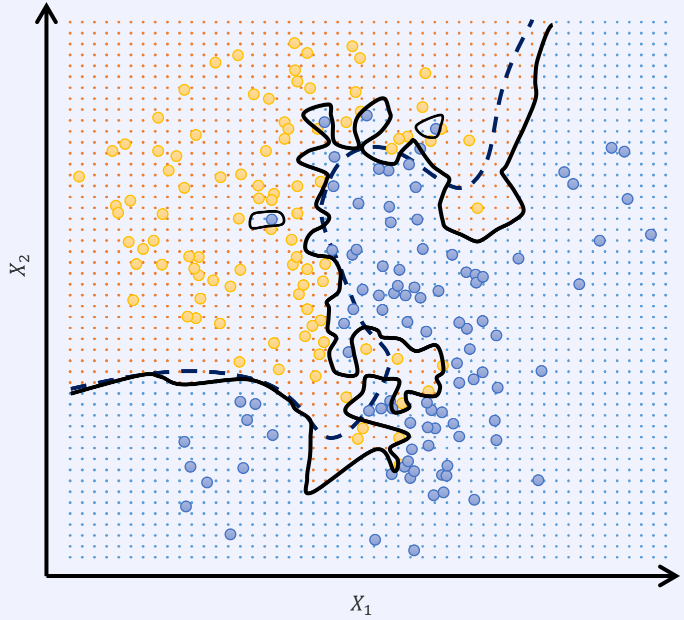
$\boldsymbol{k}$-nearest neighbors ($\boldsymbol{k}$NN)

Classifies each point to the majority class among its $\boldsymbol{k}$ nearest neighbors, i.e.

$$\arg\max_{j=1,\ldots,k} \frac{1}{k} \sum_{i \in \mathcal{N}_0} I(y_i = j)$$

where $\mathcal{N}_0$ are the $\boldsymbol{k}$ data points nearest to $\boldsymbol{x}_0$

kNN with $k = 1$



overtrained
model too complex