PSY9511: Seminar 3

Regularization and variable selection

Esten H. Leonardsen

07.09.23

Outline

1. Assignment 1
2. Assignment 2
3. Regularization
    - Variable selection
    - Shrinkage (+ live coding 🥳)
    - Dimensionality reduction

Assignment 1

UNIVERSITETET
I OSLO

Assignment 1: Coding

- Create a vector of 100 standard normally distributed numbers and visualize them with a histogram.
- Show rows 5, 8, 9, and 10 of the Auto dataset.
- Show the last three columns of the Auto dataset.
- Show all cars with five cylinders in the Auto dataset.

Assignment 1: Coding

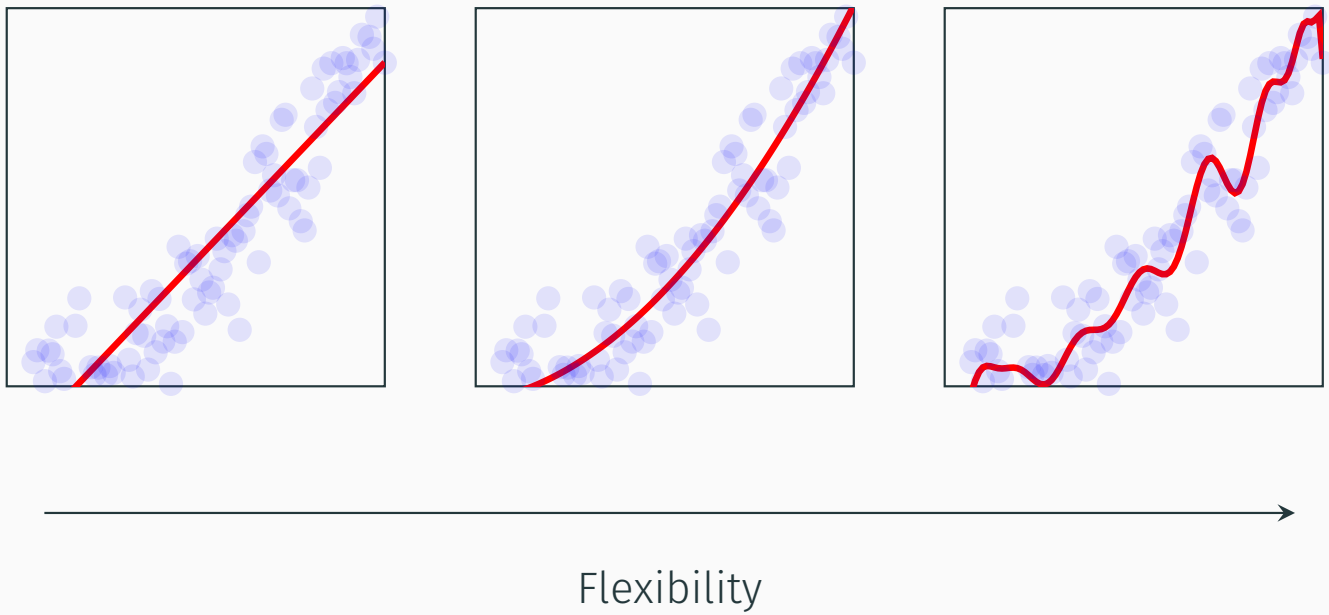
- Create a vector of 100 standard normally distributed numbers and visualize them with a histogram.
- Show rows 5, 8, 9, and 10 of the Auto dataset.
- Show the last three columns of the Auto dataset.
- Show all cars with five cylinders in the Auto dataset.

http://localhost:8889/notebooks/notebooks%2FAssignment%201.ipynb

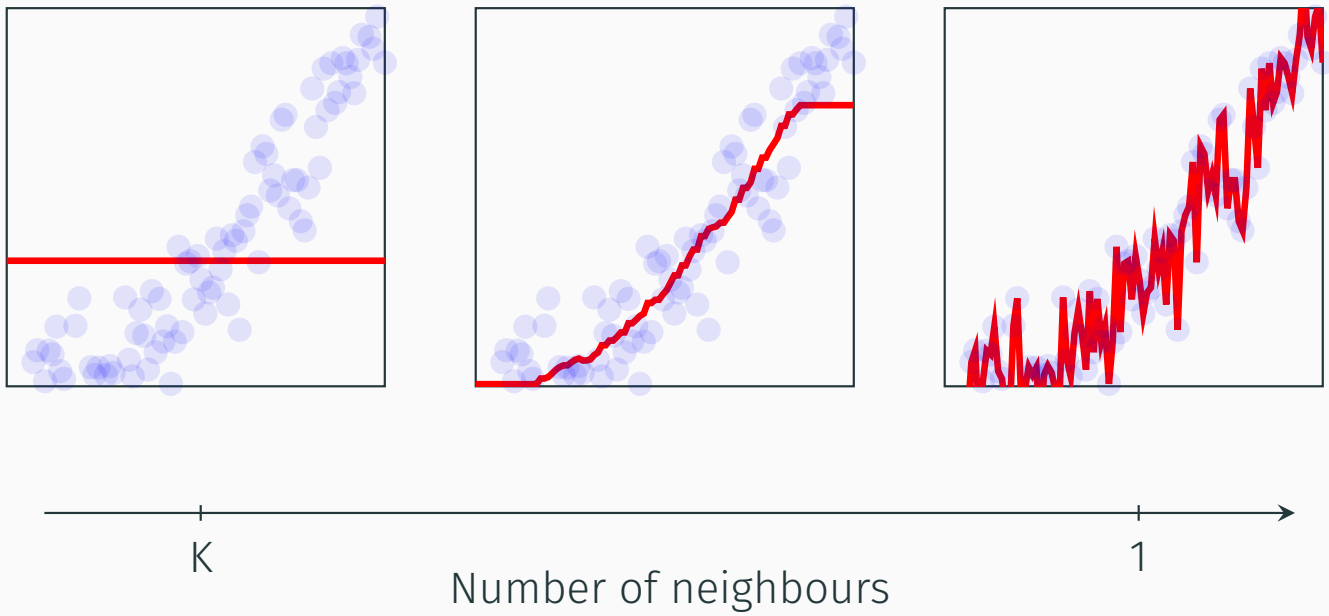
Assignment 2: Bias-variance trade-off

Assignment 2: Bias-variance trade-off

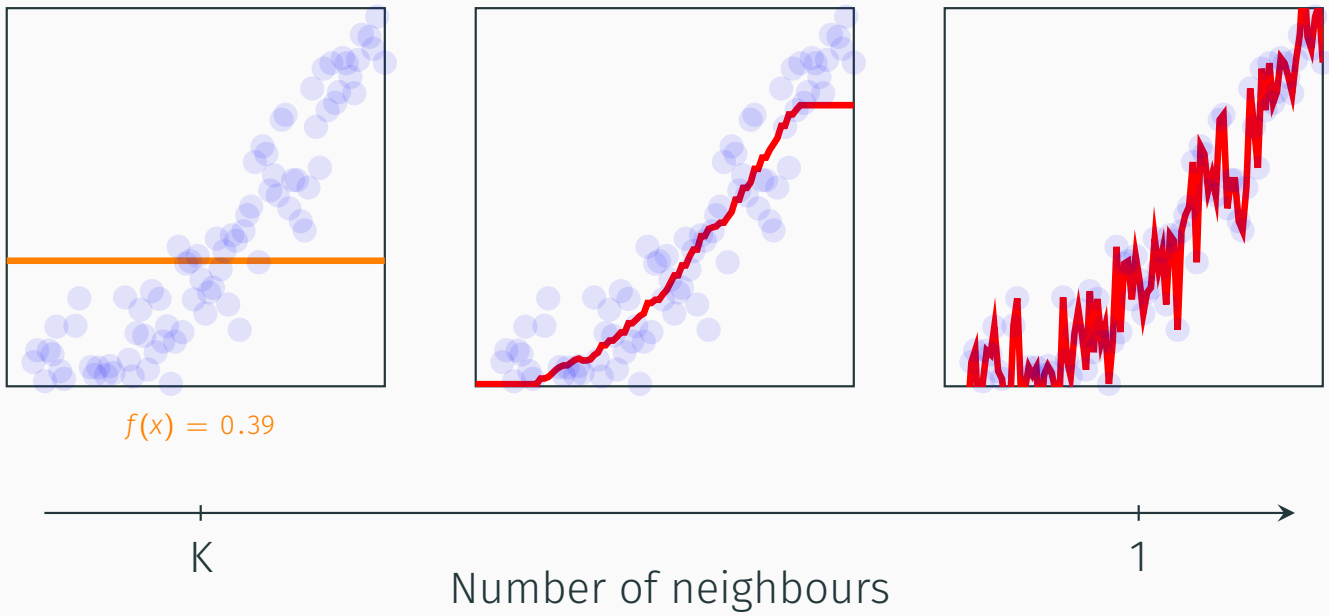
Assignment 2: Bias-variance trade-off



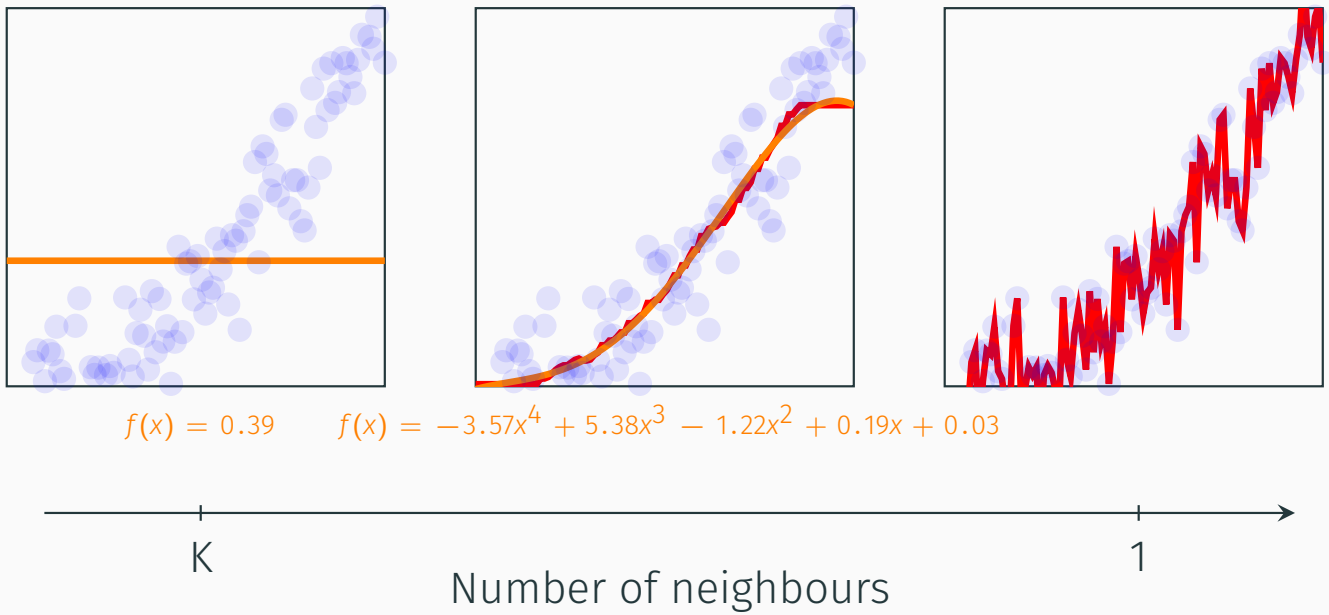
Assignment 2: Bias-variance trade-off



Assignment 2: Bias-variance trade-off



Assignment 2: Bias-variance trade-off



Assignment 2: Bias-variance trade-off

$f(x) = 0.39$ $f(x) = -3.57x^4 + 5.38x^3 - 1.22x^2 + 0.19x + 0.03$

Assignment 2: Bias-variance trade-off

$f(x) = 0.39$ $f(x) = -3.57x^4 + 5.38x^3 - 1.22x^2 + 0.19x + 0.03$

1

Assignment 2: Bias-variance trade-off

$$f(x) = 0.39 \qquad f(x) = -3.57x^4 + 5.38x^3 - 1.22x^2 + 0.19x + 0.03$$

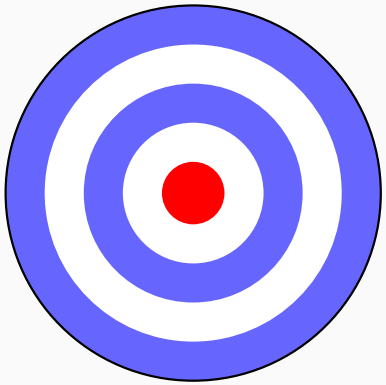
1 1 2 3 4 5

Assignment 2: Bias-variance trade-off

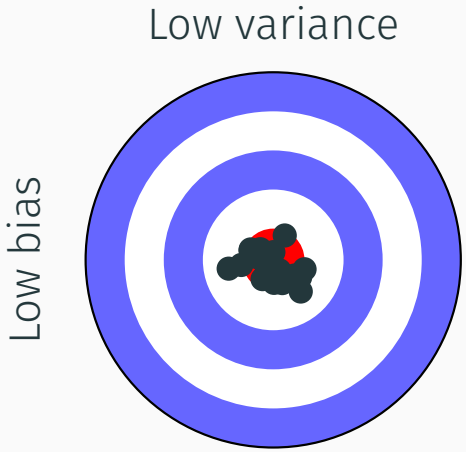
Model flexibility: Denotes the complexity of the approximated function $\hat{y} = \hat{f}(x)$.

- Informally: Wigglyness of the line
- Formally: Number of parameters in the function (degrees of freedom)

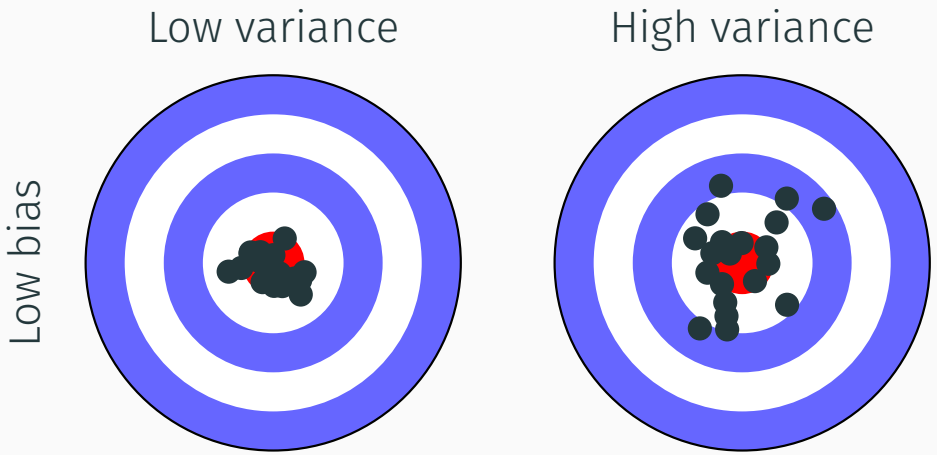
Assignment 2: Bias-variance trade-off



Assignment 2: Bias-variance trade-off

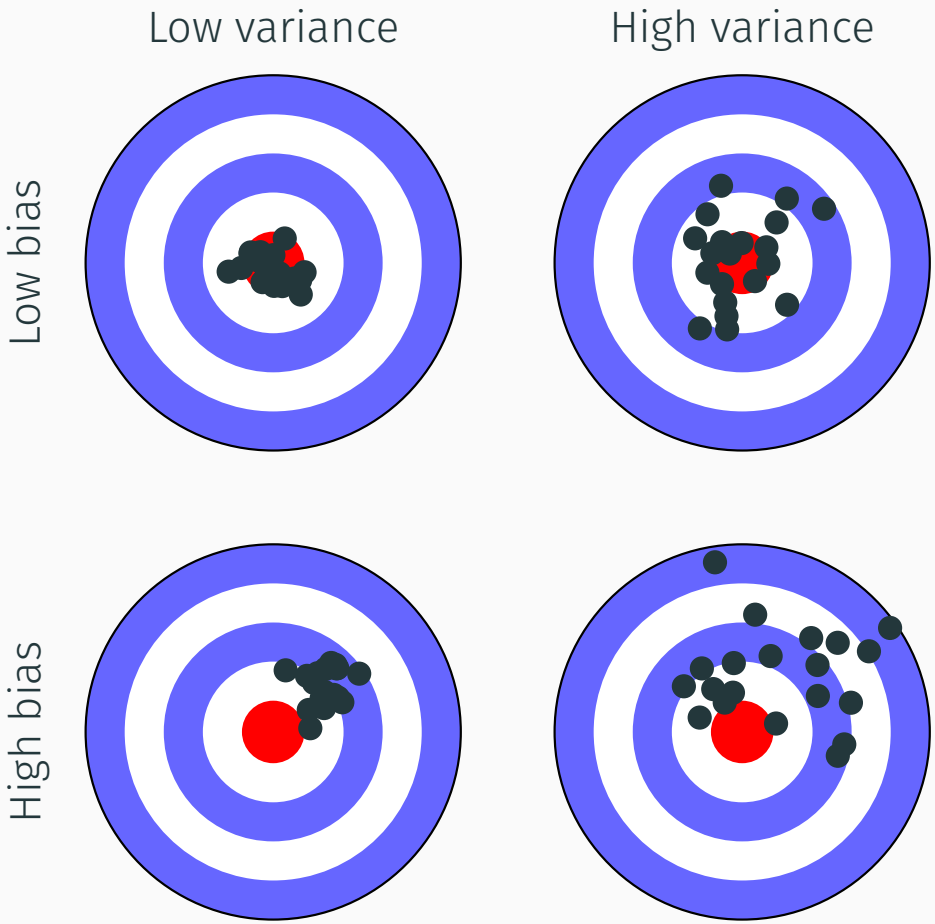


Assignment 2: Bias-variance trade-off



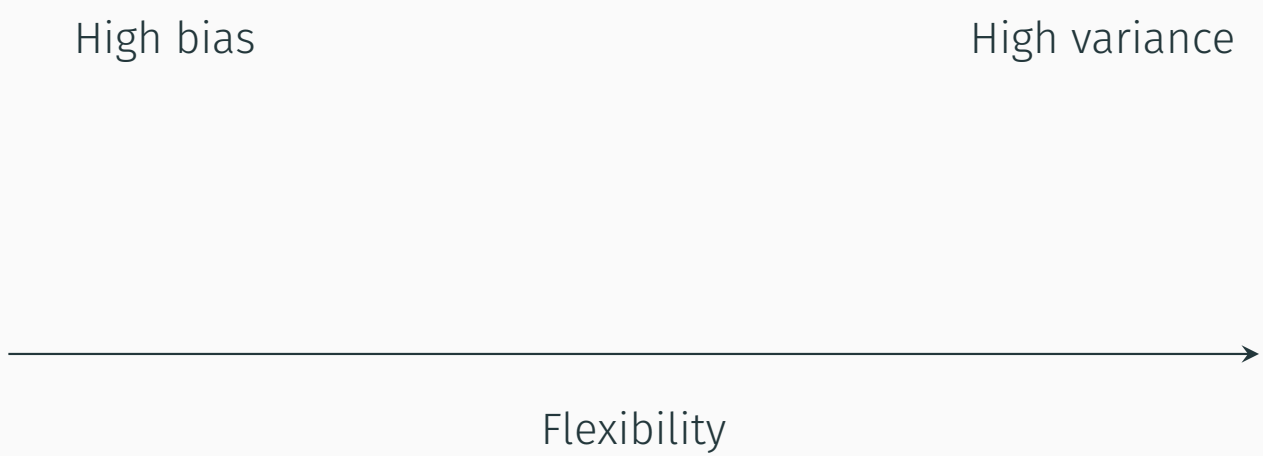
Assignment 2: Bias-variance trade-off

Assignment 2: Bias-variance trade-off

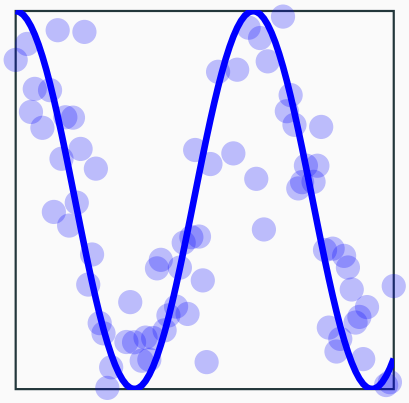


Assignment 2: Bias-variance trade-off

Assignment 2: Bias-variance trade-off

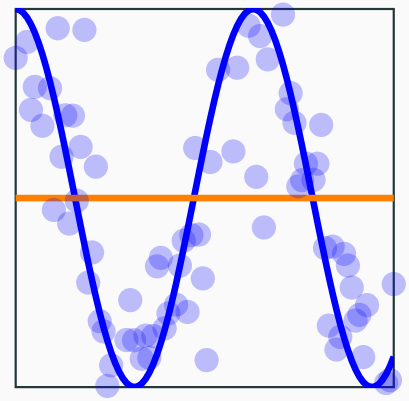


Assignment 2: Bias-variance trade-off



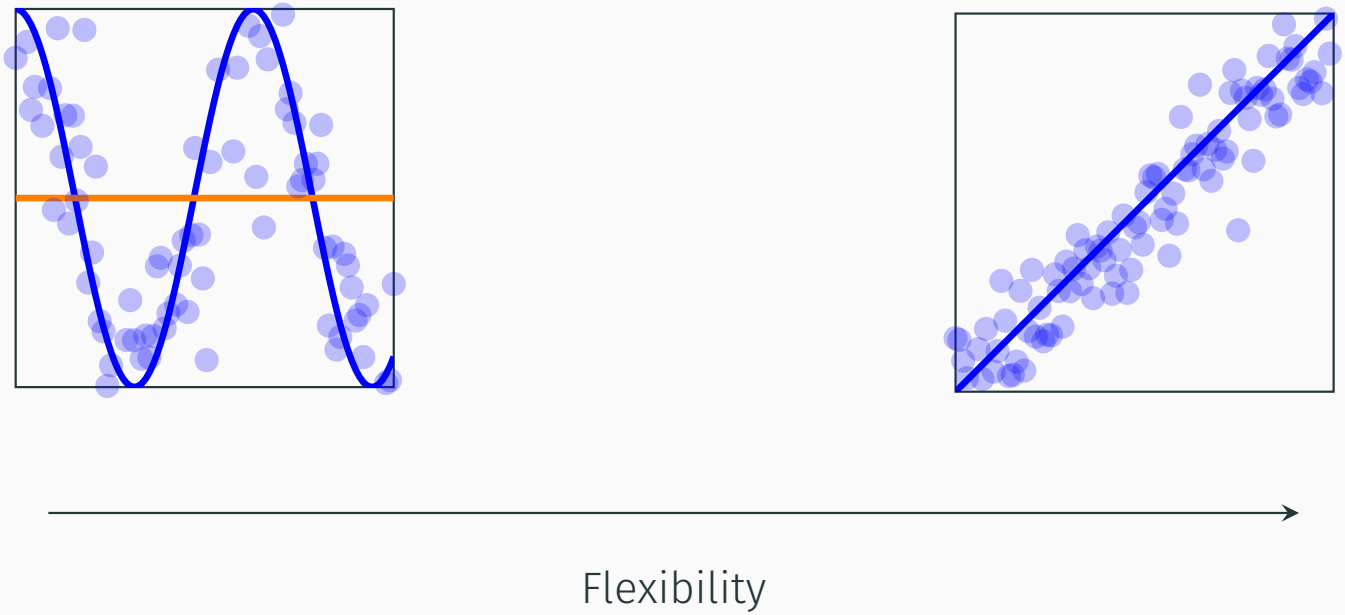
Flexibility

Assignment 2: Bias-variance trade-off

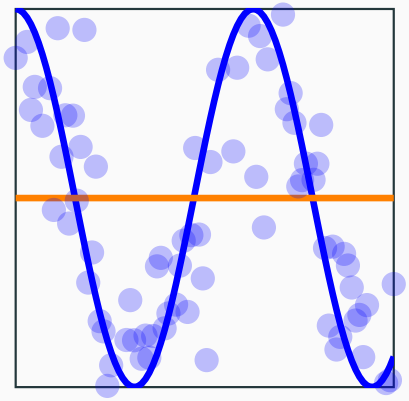


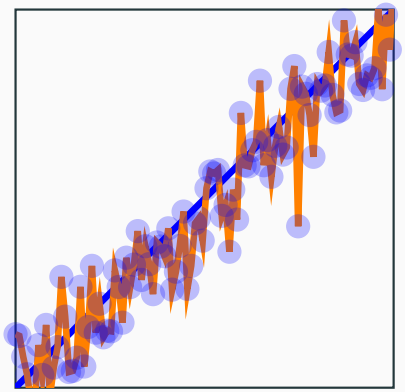
Flexibility

Assignment 2: Bias-variance trade-off



Assignment 2: Bias-variance trade-off





Flexibility

Assignment 2: Bias-variance trade-off

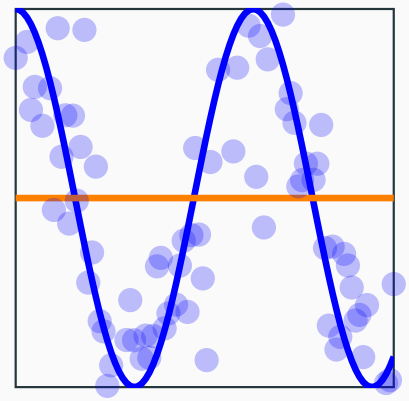
Model flexibility: Denotes the complexity of the approximated function $\hat{y} = \hat{f}(x)$.

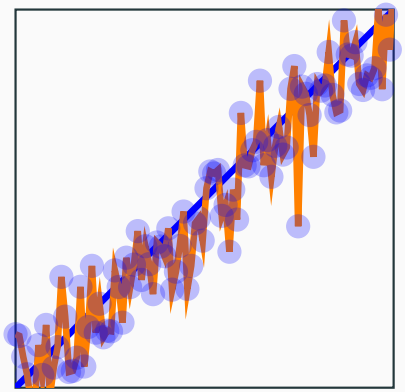
- Informally: Wigglyness of the line
- Formally: Number of parameters in the function (degrees of freedom)

Bias and variance: Two ways the model can be bad

- High bias: The model misses in *systematic* ways
- High variance: The model misses in *unsystematic* ways

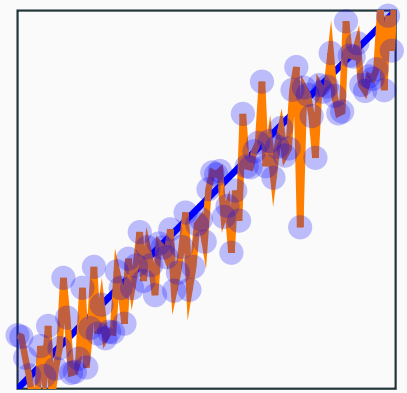
Assignment 2: Bias-variance trade-off





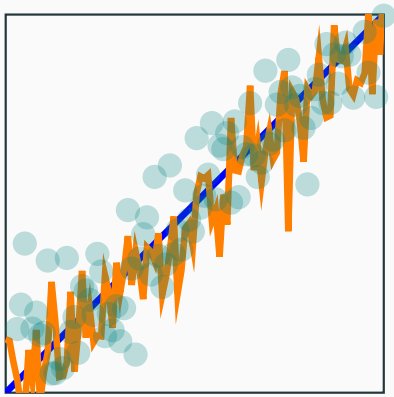
Flexibility

Assignment 2: Bias-variance trade-off



Flexibility

Assignment 2: Bias-variance trade-off



Flexibility

Assignment 2: Bias-variance trade-off

Model flexibility: Denotes the complexity of the approximated function $\hat{y} = \hat{f}(x)$.

- Informally: Wigglyness of the line
- Formally: Number of parameters in the function (degrees of freedom)

Bias and variance: Two ways the model can be bad

- High bias: The model misses in *systematic* ways
- High variance: The model misses in *unsystematic* ways

Underfitting and overfitting: Bias-variance trade-off in practice

- Underfitting: The model is equally bad on training and test data *due to not having captured the true relationship between inputs and outputs*
- Overfitting: The model is good on training data, but bad on test data *because it has found patterns in the noise during training*

Assignment 2

UNIVERSITETET
I OSLO

Assignment 2: Data splitting

Assignment 2: Random seeds

Assignment 2: Log-odds vs probability vs class

Assignment 2: Eye test

Regularization

UNIVERSITETET
I OSLO

Regularization: Motivation

$$y \sim \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3$$

Regularization: Motivation