

# Exercises week 38

## Contents

- [Overarching aims of the exercises this week](#)

September 18-22, 2023

Date: **Deadline is Sunday September 24 at midnight**

## Overarching aims of the exercises this week

The aim of the exercises this week is to derive the equations for the bias-variance tradeoff to be used in project 1 as well as testing this for a simpler function using the bootstrap method. The exercises here can be reused in project 1 as well.

Consider a dataset  $\mathcal{L}$  consisting of the data  $\mathbf{X}_{\mathcal{L}} = \{(y_j, \mathbf{x}_j), j = 0 \dots n-1\}$ .

We assume that the true data is generated from a noisy model

$$\mathbf{y} = f(\mathbf{x}) + \epsilon.$$

Here  $\epsilon$  is normally distributed with mean zero and standard deviation  $\sigma^2$ .

In our derivation of the ordinary least squares method we defined an approximation to the function  $f$  in terms of the parameters  $\beta$  and the design matrix  $\mathbf{X}$  which embody our model, that is  $\tilde{\mathbf{y}} = \mathbf{X}\beta$ .

The parameters  $\beta$  are in turn found by optimizing the mean squared error via the so-called cost function

$$C(\mathbf{X}, \beta) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \tilde{y}_i)^2 = \mathbb{E}[(\mathbf{y} - \tilde{\mathbf{y}})^2].$$

Here the expected value  $\mathbb{E}$  is the sample value.

Show that you can rewrite this in terms of a term which contains the variance of the model itself (the so-called variance term), a term which measures the deviation from the true data and the mean value of the model (the bias term) and finally the variance of the noise. That is, show that

$$\mathbb{E}[(\mathbf{y} - \tilde{\mathbf{y}})^2] = (\text{Bias}[\tilde{\mathbf{y}}])^2 + \text{var}[\tilde{\mathbf{y}}] + \sigma^2,$$

with

$$(\text{Bias}[\tilde{\mathbf{y}}])^2 = (\mathbf{y} - \mathbb{E}[\tilde{\mathbf{y}}])^2,$$

and

$$\text{var}[\tilde{\mathbf{y}}] = \frac{1}{n} \sum_i (\tilde{y}_i - \mathbb{E}[\tilde{\mathbf{y}}])^2.$$

Explain what the terms mean and discuss their interpretations.

$$\mathbb{E}[(y - \tilde{y})^2] = \mathbb{E}[y^2] + \mathbb{E}[\tilde{y}^2] - 2\mathbb{E}[y\tilde{y}]$$

$$\mathbb{E}[y^2] = \mathbb{E}[(\mu + \varepsilon)^2] = \underbrace{\mathbb{E}[\mu^2]}_{=\mu^2 \text{ since } \mu \text{ is non-stochastic}} + \mathbb{E}[\varepsilon^2] + 2\mathbb{E}[\mu\varepsilon] = \mu^2 + \sigma^2 + 2\mu \underbrace{\mathbb{E}[\varepsilon]}_{=0} = \mu^2 + \sigma^2$$

$$\mathbb{E}[\tilde{y}^2] = \text{Var}(\tilde{y}) + (\mathbb{E}[\tilde{y}])^2$$

$$\mathbb{E}[y\tilde{y}] = \mathbb{E}[(\mu + \varepsilon)\tilde{y}] = \mathbb{E}[\mu\tilde{y} + \varepsilon\tilde{y}] = \mathbb{E}[\mu\tilde{y}] + \mathbb{E}[\varepsilon\tilde{y}] = \mu\mathbb{E}[\tilde{y}] + \underbrace{\mathbb{E}[\varepsilon]\mathbb{E}[\tilde{y}]}_{=0} = \mu\mathbb{E}[\tilde{y}]$$

$\nearrow \varepsilon \text{ and } \tilde{y} \text{ are independent}$

$$\begin{aligned} \Rightarrow \mathbb{E}[(y - \tilde{y})^2] &= \mu^2 + \sigma^2 + \text{Var}(\tilde{y}) + (\mathbb{E}[\tilde{y}])^2 - 2\mu\mathbb{E}[\tilde{y}] = \\ &= \mathbb{E}[(\mu - \mathbb{E}[\tilde{y}])^2] + \sigma^2 + \text{Var}(\tilde{y}) = \\ &= (\text{Bias}(\tilde{y}))^2 + \text{Var}(\tilde{y}) + \sigma^2 \end{aligned}$$

The bias of  $\tilde{y}$  is the difference between the expectation of  $\tilde{y}$  and the true data:

$$\text{Bias} \simeq \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \mathbb{E}[\tilde{y}])^2$$

The variance is the expected squared deviation of  $\tilde{y}$  around its mean:

$$\text{Var}(\tilde{y}) = \frac{1}{n} \sum_{i=0}^{n-1} (\tilde{y}_i - \mathbb{E}[\tilde{y}])^2$$

$\Rightarrow$  In order to minimize the MSE, we need to find a method which allows to obtain low bias and low variance at the same time.

By requiring a low variance, we require that the estimate of our model doesn't vary too much between training sets (if the variance is high, a small change in the data causes a large change in the model)

$\swarrow$  low complexity: high bias and low variance  
 $\searrow$  high complexity: high variance and low bias

Perform then a bias-variance analysis of a simple one-dimensional (or other models of your choice) function by studying the MSE value as function of the complexity of your model. Use ordinary least squares only.

Discuss the bias and variance trade-off as function of your model complexity (the degree of the polynomial) and the number of data points, and possibly also your training and test data using the **bootstrap** resampling method. You can follow the code example in the jupyter-book at

[https://compphysics.github.io/MachineLearning/doc/LectureNotes/\\_build/html/chapter3.html#the-bias-variance-tradeoff](https://compphysics.github.io/MachineLearning/doc/LectureNotes/_build/html/chapter3.html#the-bias-variance-tradeoff).

See also the whiteboard notes from week 37 at

<https://github.com/CompPhysics/MachineLearning/blob/master/doc/HandWrittenNotes/2023/NotesSep14.pdf>

---

By Morten Hjorth-Jensen

© Copyright 2021.