UNIVERSITÄT DES SAARLANDES
Prof. Dr. Dietrich Klakow
Lehrstuhl für Signalverarbeitung
NNTI Winter Term 2023/2024

# Exercise Sheet 4

## Machine Learning Basics

**Deadline: 30.11.2023 23:59**

**Guidelines:** You are expected to work in a group of 2-3 students. While submitting the assignments, please make sure to include the following information for all our teammates in your PDF/python script:

**Name:**

**Student ID (matriculation number):**

**Email:**

Note that the above instructions are mandatory. If you are not following them, tutors can decide not to correct your exercise.

**Exercise 4.1 - Bias and Variance** (0.5+0.5+2.5+0.5 points)

In this exercise we want to understand the relationship between Variance and Bias of an estimator and how that causes Overfitting or Underfitting. Read chapter 5.2 and 5.4 in [1] or chapter 3.2 in [2] and answer the following questions:

a) Explain the relationship between the Bias and the Variance and how they relate to the model complexity.

b) Describe Overfitting and Underfitting in terms of Variance and Bias.

c) Given the basic form of a regression model:

$$y_0 = f(x_0) + \epsilon \tag{1}$$

where $\epsilon$ represents the inexplicable noise and $f(x_0)$ being the fitting function, prove the following :

$$MSE(y_o, \hat{y_0}) = \mathrm{Var}(\hat{f}(x_0)) + \mathrm{Bias}^2(\hat{f}(x_0)) + \mathrm{Var}(\varepsilon). \tag{2}$$

Highlight every step of your proof and clearly state the formula and reasoning in each step, as well as any assumptions made.

d) Describe how the trainingset size influences the Bias of an estimator. Is the same true for the Variance? Explain your answer.

**Exercise 4.2 - Maximum Likelihood Estimate (MLE)** (1+1 points)

a) Show how a linear regression procedure with mean squared error can be justified as an MLE procedure under the assumption that the output variable $\boldsymbol{y} = y_1, \dots, y_m$ consists of $m$ i.i.d. normal variables and has likelihood

$$p(\boldsymbol{y} \mid \mathbf{X}, \boldsymbol{w}) = \prod_{m=1}^{M} \mathcal{N}(y_m \mid \boldsymbol{w}^T \boldsymbol{x}_m, \sigma^2)$$

This particular deduction is not covered in the lecture. Consult the book to gain further understanding.

b) Show how minimizing the least squares criterion with ($L_2$-)regularization, also known as ridge regression in the Statistics literature, can be justified as a MAP procedure by assuming a standard normal prior on the weights. Recall that the $L_2$-regularized least squares loss is defined as:

$$\mathcal{L}(\boldsymbol{w}) = \frac{1}{2} \sum_{n=1}^{N} (y_n - \boldsymbol{w}^T \boldsymbol{x}_n)^2 + \frac{\lambda}{2} \boldsymbol{w}^T \boldsymbol{w}.$$

Again, this derivation is not covered in the lecture, but help can be found in the book.

**Exercise 4.3 - Bias-Variance Trade-Off Exploration** (4 points)

See the accompanying jupyter notebook.

# References


[1] Ian Goodfellow and Yoshua Bengio and Aaron Courville. 2016. Deep Learning. MIT Press.

[2] Christopher M. Bishop. 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, Berlin, Heidelberg.