Working Title

University of Oslo

Andreas Godø Lefdalsnes

February 9, 2019

Abstract

Abstract goes here [e.g. 1, page 300].

Contents

Chapter 1

Introduction

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of

the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Chapter 2

Quantum Mechanics

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of

the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Chapter 3

Manybody Quantum Mechanics

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language.

There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Chapter 4

Molecular Dynamics

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be

written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Chapter 5

Machine learning

Machine learning is the study of algorithms and statistical models employed by computing systems capable of performing tasks without explicit instruction. While traditional algorithms rely on some specified input and a ruleset for determining the output, machine learning is instead concerned with a set of generic algorithms which can find patterns in a broad class of data sets. This section will give a brief overview of machine learning, and more specifically the class of algorithms known as neural networks, and will follow closely the review by [Mehta et. al.] which the reader is encouraged to seek out.

Examples of machine learning problems include identifying objects in images, transcribing text from audio and making film recommendations to viewers based on their watch history. Machine learning problems are often subdivided into estimation and prediction problems. In both cases, we choose some observable $\boldsymbol{x}$ (e.g. the period of a pendulum) related to some parameters $\boldsymbol{\theta}$ (e.g. the length and the gravitational constant) through a model $p(\boldsymbol{x}|\boldsymbol{\theta})$ that describes the probability of observing $\boldsymbol{x}$ given $\boldsymbol{\theta}$. Subsequently we perform an experiment to obtain a dataset $\boldsymbol{X}$ and use these data to fit the model. Fitting the model means finding the parameters $\hat{\boldsymbol{\theta}}$ that provide the best explanation for the data. *Estimation* problems are concerned with the accuracy of $\hat{\boldsymbol{\theta}}$, whereas prediction problems are concerned with the ability of the model $p(\boldsymbol{x}|\boldsymbol{\theta})$ to make new predictions. Physics has traditionally been more concerned with the estimation of model parameters, while in this thesis we will be focused on the accuracy of the model.

Many problems in machine learning are defined by the same set of ingredients. The first is the dataset $\mathcal{D} = (\boldsymbol{X}, \boldsymbol{Y})$, where $\boldsymbol{X}$ is a matrix containing

observations of the independent variables $\boldsymbol{x}$, and $\boldsymbol{Y}$ is a matrix containing observations of dependent variables. Second is a model $\boldsymbol{F} : \boldsymbol{x} \to \boldsymbol{y}$ which is a function of the parameters $\boldsymbol{\theta}$. Finally we have a cost function $\mathcal{C}\left(\boldsymbol{Y}, \boldsymbol{F}\left(\boldsymbol{X}; \boldsymbol{\theta}\right)\right)$ that judges the performance of our model at generating predictions.
In the case of linear regression we consider a set of independent observations $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_N\end{bmatrix}$ related to a set of dependent observations $\boldsymbol{y} = (y_1, y_2, \dots, y_N)$ through a linear model $f(\boldsymbol{x}; \boldsymbol{\theta}) = x_1 \cdot w_1 + x_2 \cdot w_2 + \cdots + x_P \cdot w_P$, with parameters $\boldsymbol{\theta} = (w_1, w_2, \dots, w_P)$. The cost function is the well known sum of least squares $\mathcal{C}(\boldsymbol{y}, f(\boldsymbol{X}; \boldsymbol{\theta})) = \sum_i^N (y_i - f(\boldsymbol{x}_i; \boldsymbol{\theta}))^2$ and the best fit is chosen as the set of parameters which minimize this cost function: $\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \mathcal{C}(\boldsymbol{Y}, f(\boldsymbol{X}; \boldsymbol{\theta}))$.

5.0.1 Basics of statistical learning

Statistical learning theory is a field of statistics dealing with the problem of making predictions from data. We begin with an unknown function $y = f(x)$ and our goal is to develop a function $h(x)$ such that $h \sim f$. We fix a hypothesis set $\mathcal{H}$ that the algorithm is willing to consider. The expected error of a particular h over all possible inputs x and outputs y is:

$$E[h] = \int_{X \times Y} \mathcal{C}(h(x), y) \rho(x, y) dx dy,$$

where $\mathcal{C}$ is a cost function and $\rho(x, y)$ is the joint probability distribution for x and y. This is known as the *expected error*. Since this is impossible to compute without knowledge of the probability distribution ρ, we instead turn to the *empirical error*. Given n data points the empirical error is given as:

$$E_S[h] = \frac{1}{n} \sum_i^n \mathcal{C}(h(x_i), y_i).$$

The *generalization error* is defined as the difference between the expected and empirical errors:

$$G = E[h] - E_S[h].$$

We say an algorithm is able to learn from data or *generalize* if

$$\lim_{n \to \infty} G = 0.$$

We are in general unable to compute the expected error, and therefore unable to compute the generalization error. The most common approach known as *cross-validation* is to estimate the generalization error by subdividing our dataset into a *training* set and a *test* set. The value of the cost function on the training set is called the *in-sample* error and the value of the cost function on the test set the *out-of-sample* error. Assuming the dataset is sufficiently large and representative of f, and the subsampling into train and test datasets is unbiased, the in-sample error can serve as an appropriate proxy for the generalization error.
In figure **??** we show the typical evolution of the errors as the number of data points increase. It is assumed that the function being learned is sufficiently complicated that we cannot learn it exactly, and that we have a sizeable number of data points available. The in-sample error will decrease monotonically, as our model is not able to learn the underlying data exactly. In contrast, the out-of-sample error will decrease, as the sampling noise decreases and the training data set becomes more representative of the underlying probability distribution. In the limit, these errors both approach same value, which is known the model *bias*. The bias represents the best our model could do in the infinite data limit. The out-of-sample error produced from the sampling noise is known as *variance*, and will vanish completely given an infinite representative data set.

In figure **??** we show the typical evolution of the out-of-sample error as the model *complexity* increases. Model complexity is a measure of the degrees of freedom in the model space, for example the number of coefficients in a polynomial regression. In the figure we can see that bias decreases monotonically as model complexity increases, as the model is able to fit a larger space of functions. However, the variance will also increase as the model becomes more susceptible to sampling noise. In general the lowest out-of-sample error, and therefore generalization error, is achieved at an intermediate model complexity. We also find that as model complexity increases, a larger amount of data points is required to be able to reasonably fit the true function.

5.0.2 Bias-variance decomposition

Consider a dataset $\mathcal{D}(\boldsymbol{X}, \boldsymbol{y})$ of n pairs of independent and dependent variables. Assume the true data is generated from a noisy model:

$$y = f(\boldsymbol{x}) + \epsilon,$$

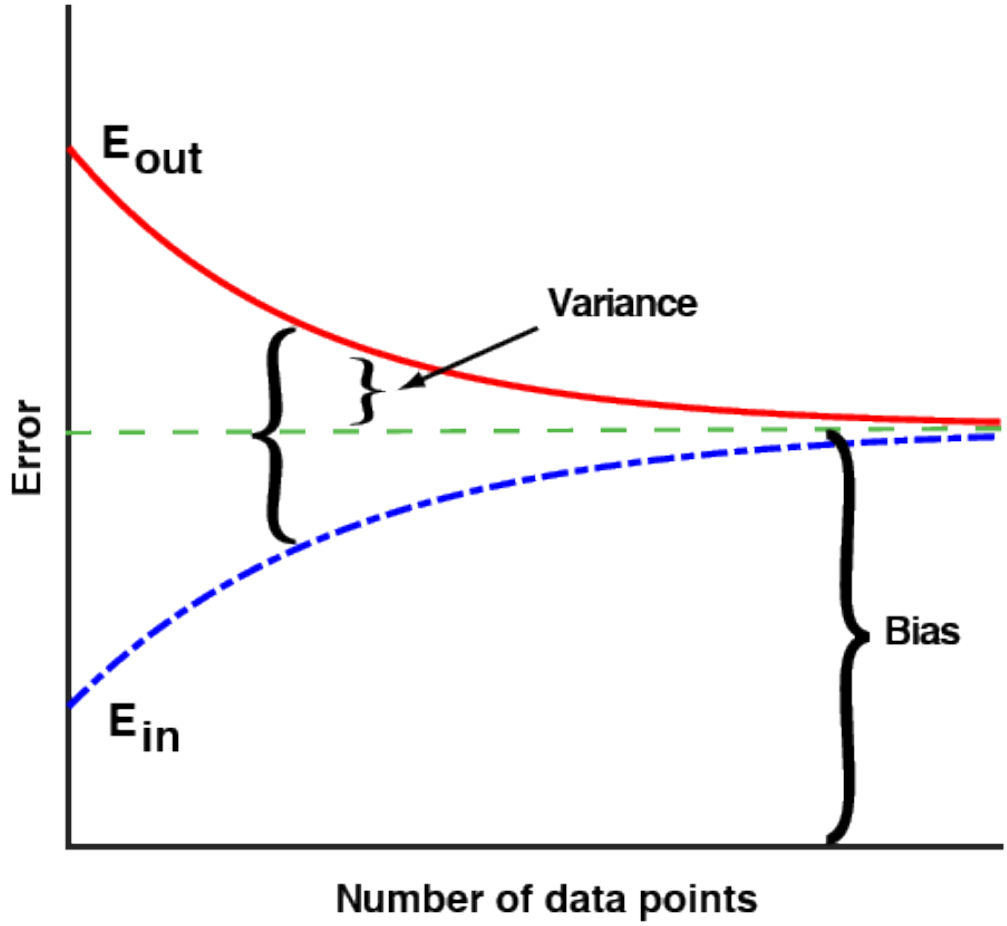

Figure 5.1: Typical in-sample and out-of-sample error as a function of the number of data points. It is assumed that the number of data points is not small, and that the true function cannot be exactly fit.

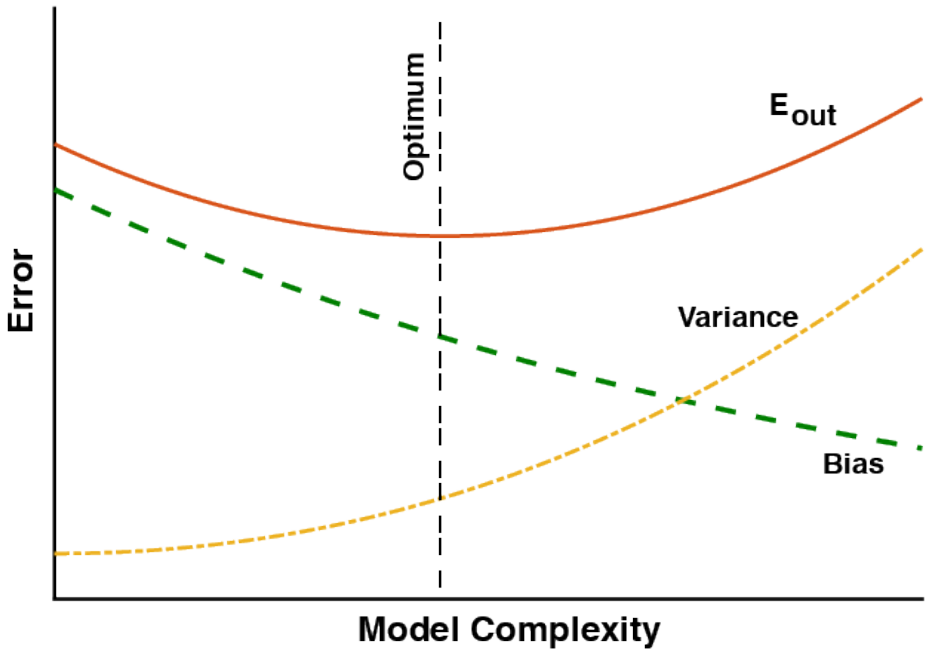

Figure 5.2: Typical out-of-sample error as a function of model complexity for a fixed dataset. Bias decreases monotonically with model complexity, while variance increases as a result of sampling noise.

where ϵ is normally distributed with mean μ and standard deviation σ. Assume that we have an estimator $h(\boldsymbol{x};\boldsymbol{\theta})$ trained by minimizing a cost function $\mathcal{C}(\boldsymbol{y}, h(\boldsymbol{x}))$ which we take to be the sum of squared errors:

$$\mathcal{C}(\boldsymbol{y}, h(\boldsymbol{x})) = \sum_i^n (y_i - h(\boldsymbol{x}_i;\boldsymbol{\theta}))^2.$$

Our best estimate for the model parameters:

$$\boldsymbol{\theta}_{\mathcal{D}} = \arg\min_{\boldsymbol{\theta}} \mathcal{C}(\boldsymbol{y}, h(\boldsymbol{x};\boldsymbol{\theta})),$$

is a function of the dataset $\mathcal{D}$. If we imagine we have a set of datasets $\mathcal{D}_j = (\boldsymbol{y}_j, \boldsymbol{X}_j)$, each with n samples, we would like to calculate the expectation value of the cost function over all these datasets $E_{\mathcal{D},\epsilon}$. We would also like to calculate the expectation value over different instances of the noise ϵ. The expected generalization error can be decomposed as:

$$\begin{aligned} E_{\mathcal{D},\epsilon}[\mathcal{C}(\boldsymbol{y}, h(\boldsymbol{X};\boldsymbol{\theta}_{\mathcal{D}}))] &= E\left[\sum_i (y_i - h(\boldsymbol{x}_i;\boldsymbol{\theta}_{\mathcal{D}}))^2\right] \\ &= \sum_i \sigma_\epsilon^2 + E_{\mathcal{D}}[(f(\boldsymbol{x}_i) - f(\boldsymbol{x}_i;\boldsymbol{\theta}_{\mathcal{D}}))^2]. \end{aligned} \tag{5.1}$$

The second term can be further decomposed as

$$\begin{aligned} &E_{\mathcal{D}}[(f(\boldsymbol{x}_i) - f(\boldsymbol{x}_i;\boldsymbol{\theta}_{\mathcal{D}}))^2] \\ &= (f(\boldsymbol{x}_i) - E_{\mathcal{D}}[h(\boldsymbol{x}_i;\boldsymbol{\theta}_{\mathcal{D}})])^2 + E[(h(\boldsymbol{x}_i;\boldsymbol{\theta}_{\mathcal{D}}) - E[h(\boldsymbol{x}_i;\boldsymbol{\theta}_{\mathcal{D}}])^2] \end{aligned} \tag{5.2}$$

The first term is what we have referred to as the bias:

$$\text{Bias}^2 = \sum_i (f(\boldsymbol{x}_i) - E_{\mathcal{D}}[h(\boldsymbol{x}_i;\boldsymbol{\theta}_{\mathcal{D}})])^2. \tag{5.3}$$

The bias measures the expectation value of the deviation of our model from the true function, i.e. the best we can do in the infinite data limit.
The second term is what we have referred to as the variance:

$$\text{Var} = \sum_i E[(h(\boldsymbol{x}_i;\boldsymbol{\theta}_{\mathcal{D}}) - E[h(\boldsymbol{x}_i;\boldsymbol{\theta}_{\mathcal{D}}])^2] \tag{5.4}$$

The variance measures the deviation of our model due to finite-sampling effects. Combining these effects we can decompose the out-of-sample error into:

$$E_{\text{out}} = \text{Bias}^2 + \text{Var} + \text{Noise}, \tag{5.5}$$

with Noise $= \sum_i \sigma_\epsilon^2$.
In general it can be much more difficult to obtain sufficient good data than to train a very complex model. Therefore it is often useful in practice to use a less complex model with higher bias, because it is less susceptible to finite-sampling effects.

5.0.3 Neural networks

Artificial Neural Networks (ANN) or Deep Neural Networks (DNN) are supervised learning models vaguely inspired by biological neural networks.

5.0.4 Optimization

More text.

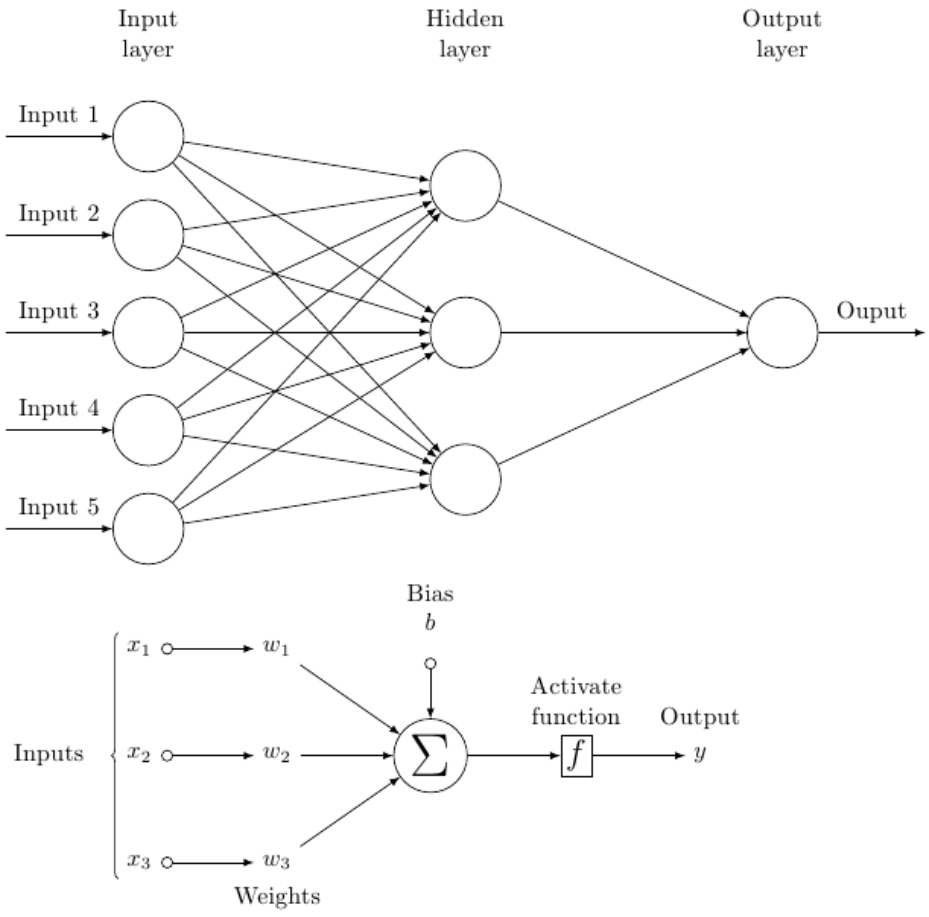

Figure 5.3: Text.

Chapter 6

Conclusion

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of

the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Appendix A

Appendix Title

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of

the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Bibliography

[1] Albert Einstein. “Zur Elektrodynamik bewegter Körper. (German) [On the electrodynamics of moving bodies]”. In: *Annalen der Physik* 322.10 (1905), pp. 891–921. DOI: http://dx.doi.org/10.1002/andp.19053221004.