

**Deep Learning DM873**

**Exam notes**

Mark Jervelund

[Mark@jervelund.com](mailto:Mark@jervelund.com)

[Doommius.com/notes.php](http://Doommius.com/notes.php)



UNIVERSITY OF  
SOUTHERN DENMARK

IMADA

**June 26, 2019**

# Contents

<b>I</b>	<b>Formal course description</b>	<b>1</b>
	Aim . . . . .	2
	Statement of aims . . . . .	2
	Pensum . . . . .	2
<b>II</b>	<b>Exam topics</b>	<b>3</b>
	Exam Form . . . . .	4
	Feed-Forward Networks . . . . .	5
	introduction . . . . .	5
	Function Principle . . . . .	5
	Output Units . . . . .	5
	Hidden Units . . . . .	5
	Architecture design . . . . .	6
	Backpropagation . . . . .	7
	introduction . . . . .	7
	Function Principle . . . . .	7
	Computational Graphs . . . . .	7
	Mini batches . . . . .	7
	Regularization . . . . .	8
	introduction . . . . .	8
	Over/Underfitting & Model Capacity . . . . .	8
	Data augmentation . . . . .	8
	Adversarial training . . . . .	8
	Early stopping . . . . .	8
	Bagging (Bootstrap Aggregating) . . . . .	8
	Dropout . . . . .	8
	Convolutional Neural Networks . . . . .	9
	introduction . . . . .	9
	Function Principle . . . . .	9
	Pooling . . . . .	9
	Initialization of the kernels . . . . .	9
	Recurrent Neural Networks . . . . .	10
	introduction . . . . .	10
	Function Principle . . . . .	10
	Problems with long term memory . . . . .	10
	Long Short Term Memory . . . . .	10
	Optimization for Neural Networks . . . . .	11
	introduction . . . . .	11
	Parameter Initialization . . . . .	11
	Adaptive Learning . . . . .	11
	Batch Normalization . . . . .	11
	Pre-training . . . . .	11
	Local minima . . . . .	11
	Autoencoders and GANs . . . . .	12
	introduction . . . . .	12
	Autoencoders . . . . .	12

Variational Autoencoders . . . . .	12
GANs . . . . .	12

## Part I

# Formal course description

## Aim

Machine learning has become a part in our everyday life, from simple product recommendations to personal electronic assistant to self-driving cars. More recently, through the advent of potent hardware and cheap computational power, “Deep Learning” has become a popular and powerful tool for learning from complex, large-scale data. In this course, we will discuss the fundamentals of deep learning and its application to various different fields. We will learn about the power but also the limitations of these deep neural networks. At the end of the course, the students will have significant familiarity with the subject and will be able to apply the learned techniques to a broad range of different fields.

The course builds partly on the knowledge acquired in the course DM555 but can be taken by any Computer Science or Computational BioMedicine Master student.

In relation to the competence profile of the degree it is the explicit focus of the course to:

- giving the competence to plan and execute a deep learning task by means of deep neural networks.
- providing knowledge on the different types of deep learning approaches including their advantages and disadvantages.
- transfer learned methods to new fields of applications.
- challenges the student with real-life datasets and problem-solving skills

## Statement of aims

- The learning objectives of the course is that the student demonstrates the ability to:
- Describe the principles of deep neural networks in a scientific and precise language and notation
- Analyze the various types of neural networks, the different layers and their interplay
- Describe the feasibility of deep learning approaches to concrete problems
- Understand the theoretical mathematical foundations of the field
- Apply deep learning frameworks for solving concrete problems

## Pensum

- All lecture slides are relevant for the exams.
- All readings noted in the lecture list are relevant for the exam.
- Ian Goodfellow, Yoshua Bengio, Aaron Courville - The Deep Learning Book
- Gareth James, Daniela Witten, Trevor Hastie - Robert Tibshirani An Introduction to Statistical Learning (ISL)

# Part II

## Exam topics

## Exam Form

The exam will last about 15-20 minutes. At the beginning, one topic from the list below will be drawn randomly. For each topic the examinee should be prepared to make a short presentation of 5 minutes. It is allowed to bring one page of hand-written notes (DIN A4 or US-Letter, one-sided) for each of the topics. The examinee will have 2 minutes to briefly study the notes for the drawn topic before the presentation. The notes may be consulted during the presentation if needed but it will negatively influence the evaluation of the examinee's performance. During the presentation, only the blackboard can be used (you cannot use overhead transparencies, for instance).

After the short presentation, additional question about the presentation's topic but also about other topics in the curriculum will be asked.

Below is the list of possible topics and some suggested content. The listed content are only suggestions and is not necessarily complete nor must everything be covered in the short presentation. It is the responsibility of the examinee to gather and select among all relevant information for each topic from the course material. On the course website you can find suggested readings for each of these topics.

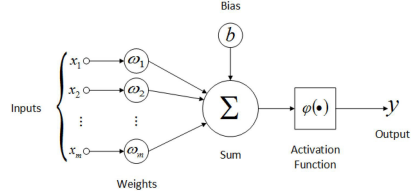
# Feed-Forward Networks

## introduction

A feed forward is the oldest and simplest network we have, it only supports feeding information in one direction (forward) eg we can't have loops.

## Function Principle

The Artificial Neuron



Network build up Layers of neurons,

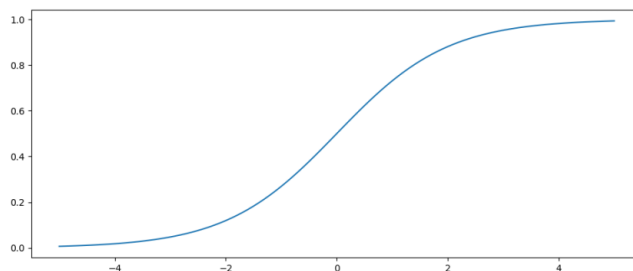
## Output Units

### Sigmoid

The sigmoid function is a binary type of output unit, which also means it can only produce a binary variable, and can only predict the Bernoulli distribution. It is of the form

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \quad (1)$$

The activation function for the sigmoid looks like the following.



### Softmax

The softmax function is a more powerful output that is able to handle Multinoulli distribution/categorical results.

The softmax consists of two layers, first in linear layer followed by the softmax function.

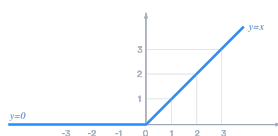
$$\text{Softmax}(z)_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (2)$$

The output vector of the softmax layer contains values in the interval [0,1] which represent the probability, the sum of all values in the output vector must always equal 1.

## Hidden Units

### ReLU - Rectified linear unit

$$f(x) = \max(0, x) \quad (3)$$





## **Architecture design**

Size and dept.

# Backpropagation

## introduction

This is the way the network calculates the gradient, This is then used to adjust the weights so the network can reach a minima in the learning function.

## Function Principle

Make example

## Computational Graphs

### Mini batches

We split the dataset into smaller batches and run one iteration of the Backpropagation function once per mini batch.

# Regularization

## introduction

Our problem is that our network needs to perform good on not just our training data but also new data.

## Over/Underfitting & Model Capacity

Overfitting got this.

## Data augmentation

We can augment data by injection noise, rotating the images a few degrees, blur,

## Adversarial training

## Early stopping

We can obtain a model with better validation set error (and thus better test error) by returning to the parameter setting at the point of time with the lowest validation set error, and at the same time reduce the chance of over fitting.(if we are limited in data set)

## Bagging (Bootstrap Aggregating)

## Dropout

# Convolutional Neural Networks

introduction

Function Principle

Pooling

Initialization of the kernels

...

# Recurrent Neural Networks

## introduction

Recurrent Neural Networks are a family of neural networks for processing sequential data  
Speech Recognition, etc.

## Function Principle



## Problems with long term memory

It should be able to handle large gaps inbetween information but in practice this fails in practice.  
Eg it can maybe understand "i'm hungry" but it has issues understanding the connection over larger paragraphs. here LSTM offer a solution.

## Long Short Term Memory

...

# Optimization for Neural Networks

introduction

Parameter Initialization

Adaptive Learning

Batch Normalization

Pre-training

Local minima

# Autoencoders and GANs

introduction

Autoencoders

Variational Autoencoders

GANs

...