Deep Learning

CIVICA Data Science Summer School

Dr Christian Arnold (Instructor, Cardiff University)
Marcel Neunhoeffer (TA, LMU Munich)

July 29th, 2021

Logistics

- The repo with all material is at `https://github.com/chrisguarnold/civica_dl_class`
- Please do ask questions in the chat: Marcel will handle them.

Introduction

Motivation

My Background

This is the most exciting time

to be a social scientist.

Ever.

My Background

Academic Background

- Senior Lecturer at Cardiff University
- PhD, University of Mannheim
- Research interests using DL, e.g.
  - **Voting Fraud Detection:** Images (satellite and scanned forms) to detect voting irregularities in remote areas.
  - **Private Synthetic Data:** Generative Adversarial Networks for differentially private synthetic micro data.

Industry Background

- Data Scientist with KIANA and KPMG
- Scientific advisory committee for the GSS, Office for National Statistics
- Knowledge transfer for the National Democratic Institute

Why Deep Learning?

Deep Neural Networks

- Deep Learning as a subset of Machine Learning
- Deep Learning does not require feature definition. Instead often byproduct
- Outperforms classic ML approaches in particular for vast amounts of data

Why Deep Learning?

Deep Neural Networks

- Deep Learning as a subset of Machine Learning
- Deep Learning does not require feature definition. Instead often byproduct
- Outperforms classic ML approaches in particular for vast amounts of data

Brief History of Neural Nets

- 1980: Deep neural nets and their training via backpropagation, but training is challenging
- 1990s/early 2000s: Less attention
- Late 2000s: breakthroughs in training, e.g. ReLUs, Regularisation, but also hardware
- 2010s: Vast attention and popularity

Today's Agenda

1. Introduction
2. Training Deep Neural Nets
3. Tuning
4. Outlook

Today's Learning

Topic	Knowledge	Experience	Skill
What is DL?	X		x
How do DNN learn?	X		
How do DNN behave?		X	x
How to write code for DNN?		x	X
Typical architectures in practice	X		

Introduction

Your First Model

MNIST Data Set

- Modified National Institute of Standards and Technology database
- 60k training and 10k testing images of handwritten digits
- Black and white images from NIST normalized to fit into a 28x28 pixel box
- One of the CLASSIC machine learning data sets

Code Nr. 1

Goal

- A first impression of the performance of neural nets
- A first impression of the code

Training Deep Neural Nets

Deep Learning in a Nutshell

Learning Data Representations in Multiple Stages

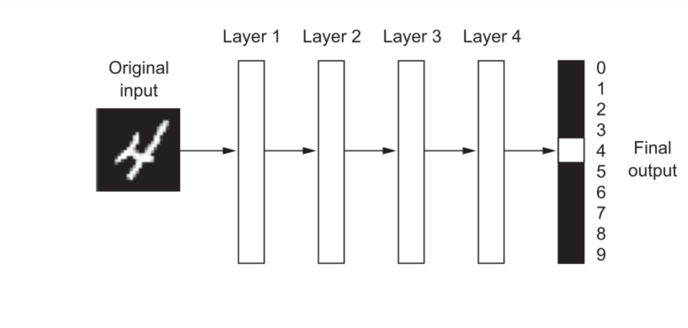


Figure 1: Source: Allaire/Chollet (2018)

What is Happening at Each Stage

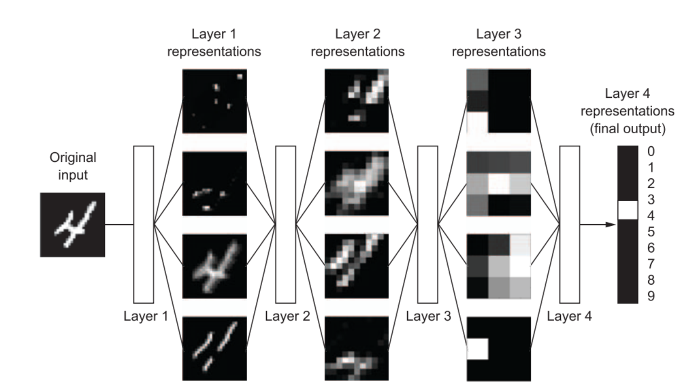


Figure 2: Source: Allaire/Chollet (2018)

How to Train Neural Nets?

The Mantra

- Predict
- Calculate how wrong the prediction is
- Propagate the information back
- Update weights

Training Deep Neural Nets

Logistic Regression

Estimating Logistic Regression with Gradient Descent

Four Steps

- Predict
- Calculate how wrong the prediction is
- Propagate the information back
- Update weights

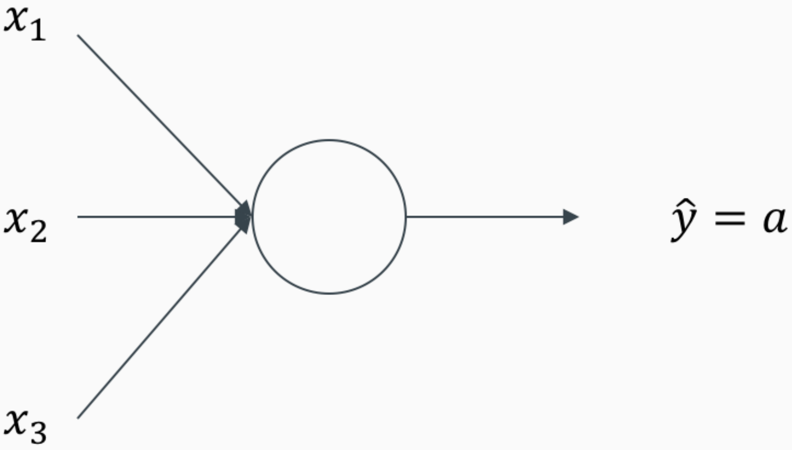
At the Whiteboard

- A refresher in logistic regression
- Gradient descent
- Logistic regression via backpropagation of errors

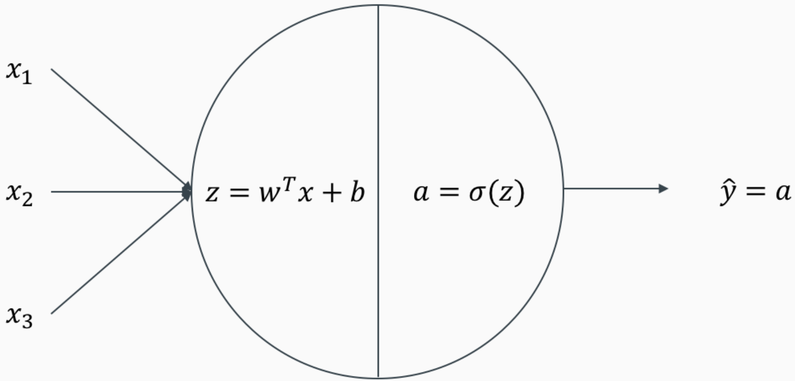
Training Deep Neural Nets

Shallow Neural Nets

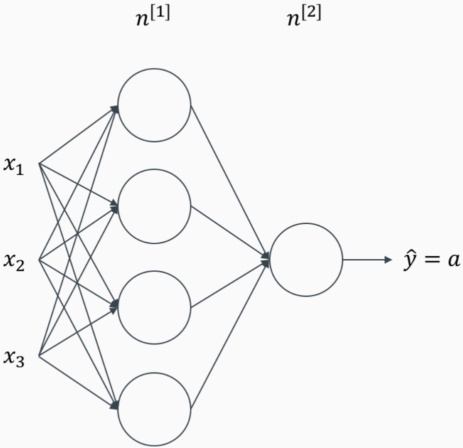
Logistic Regression



One Neuron



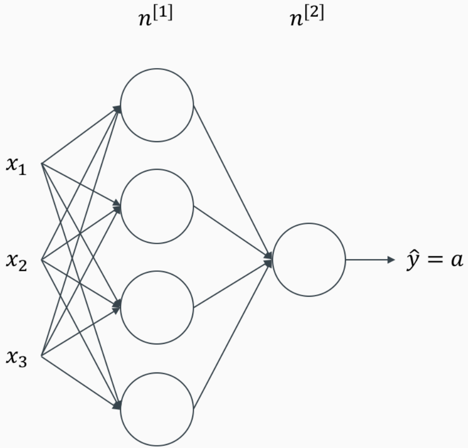
(Shallow) Neural Net



The Different Layers

- Input layer
- Hidden layer
- Output layer

(Shallow) Neural Net



Four Steps to Repeat

- Predict
- Calculate how wrong the prediction is
- Propagate the information back
- Update weights

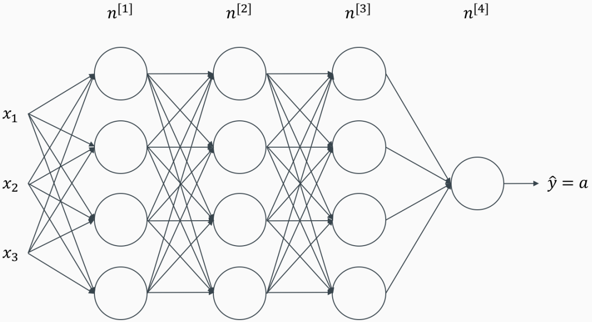
At the Whiteboard

- A (shallow) neural net
- Training neural nets via backpropagation of errors

Training Deep Neural Nets

Deep Neural Nets

Deep Neural Net



The Different Layers

- Predict
- Calculate how wrong the prediction is
- Propagate the information back
- Update weights

A Geometric Interpretation of a DNN

$$a = g(wx + b)$$

Tuning

Fitting

`https://playground.tensorflow.org/`

Goal

- Get to know the Tensorflow Playground
- Recognise key parameters and hyperparameters

Parameters and Hyperparameters

Parameters

Parameters to be optimized during training

$$W^{[1]}, b^{[1]}, W^{[2]}, b^{[2]}, W^{[3]}, b^{[3]}, W^{[4]}, b^{[4]}$$

Parameters and Hyperparameters

Parameters

Parameters to be optimized during training

$$W^{[1]}, b^{[1]}, W^{[2]}, b^{[2]}, W^{[3]}, b^{[3]}, W^{[4]}, b^{[4]}$$

Hyper Parameters

Parameters specified before training

- # hidden layers l and # hidden units $n^{[1]}, n^{[2]}, \ldots$
- Choice of activation function
- Optimizer
- Learning rate α
- Batch size

Tuning Is an Optimization Problem at Two Levels

- **Learning:** Optimise parameters $\rightarrow$ optimisation algorithm (easy)
- **Tuning:** Optimize hyperparameters $\rightarrow$ algorithm over hyperparameter space (hard)

Tuning Your Model

Why is Tuning Hard?

- Typically not possible to calculate derivative of objective $\rightarrow$ no optimization, only evaluation
- Evaluation of learning is expensive
- Space of hyperparameters typically very large in DL

Tuning Your Model

Why is Tuning Hard?

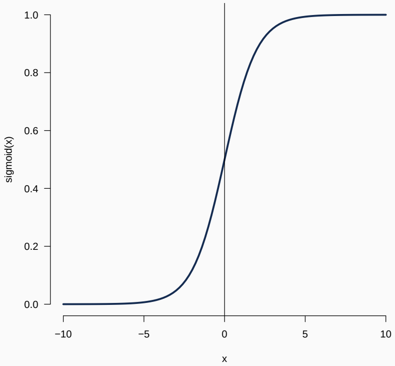
- Typically not possible to calculate derivative of objective $\rightarrow$ no optimization, only evaluation
- Evaluation of learning is expensive
- Space of hyperparameters typically very large in DL

Solution

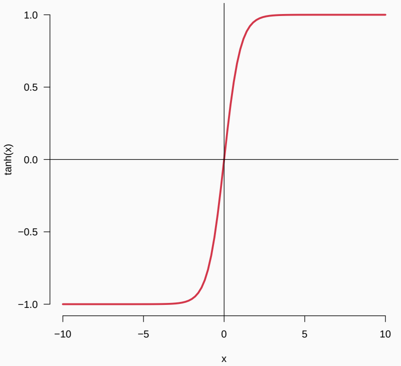
- Try different hyperparameters
- Typically grid search or random search

Activation Functions

Sigmoid



Tanh



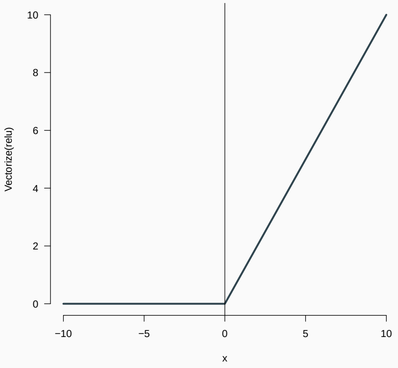
Code Nr. 2

Goal

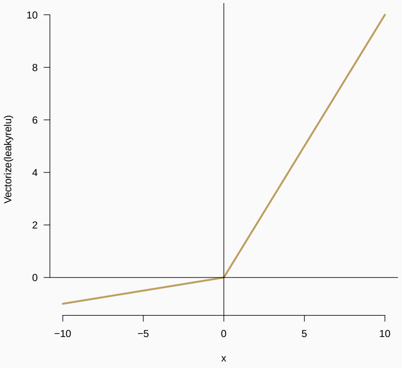
- Revisit MNIST challenge
- Understand the code
- Experience the effect from hyperparameters

Activation Functions

ReLU



Leaky ReLU



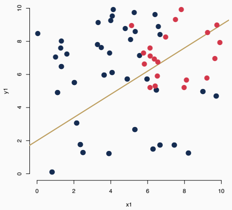
At the Whiteboard

- Why do we need an activation function?

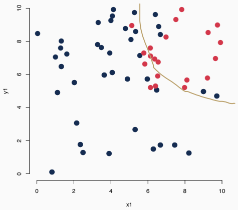
Tuning

Tackling Overfitting

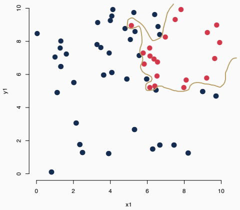
What is Overfitting?



Underfitting

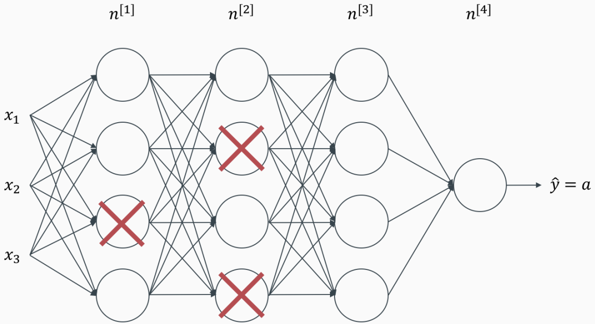


Just Right



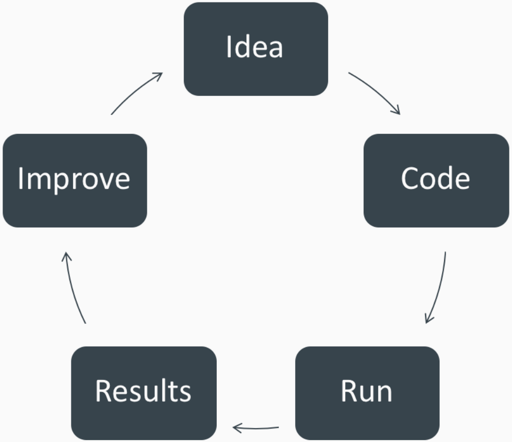
Overfitting

How to Avoid Overfitting: Dropout



- Randomly eliminate nodes in the network
- Dropout can be particularly useful for the layers with many parameters
- But: cost function J is no longer defined

Developing Your Model



- DL is experimental: Only best practice, but no proofs
- Try to use 2^x in your DNN architecture and batches
- Goal: Strike the Balance between over- and underfitting
- If cost function for training and testing are the same $\rightarrow$ bigger model
- If overfitting: regularization

`playground.tensorflow.org`

Goal

- Strike the Balance between over- and underfitting

Code Nr. 3

Goal

- Get to know a new code
- Experience the effect from hyperparameters
- Learn how to balance model performance vs. overfitting

Outlook

Is This the End, Or Is This the Beginning?

A Look at Different NN architectures

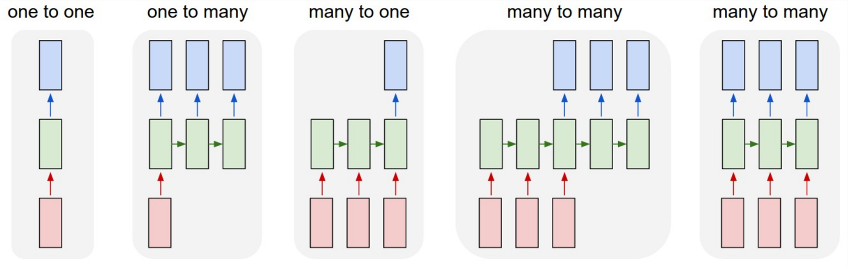


Figure 3: Different NN architectures. Source: Andrej Karpathy

- How could we process sequences with that architecture?

A Simple RNN

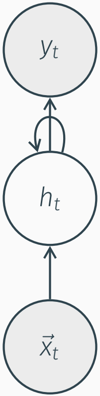


Figure 4: A simple RNN loops the output of h_t as an additional input to h_{t+1}.

Unrolling the RNN

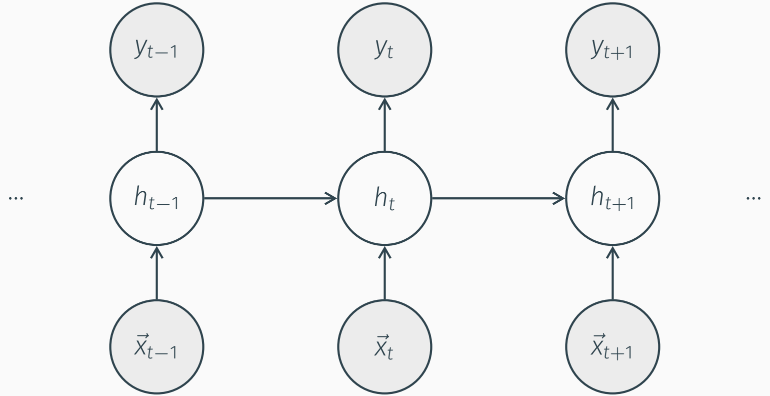


Figure 5: Unrolling the loop shows what happens in the RNN.

Visual Data Analysis with Convolutional Neural Nets (CNN)

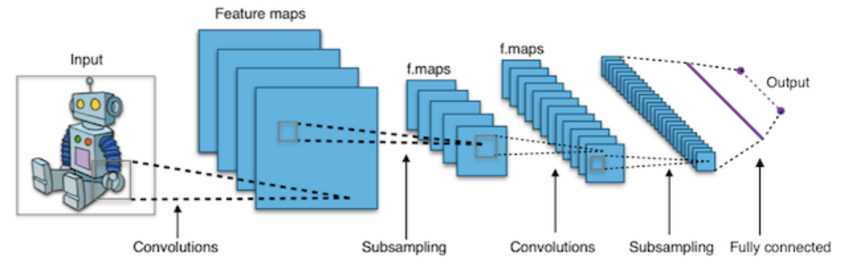


Figure 6: Typical CNN Architecture. Source: Keras Tutorial

What Does A Schematic GAN Look Like?

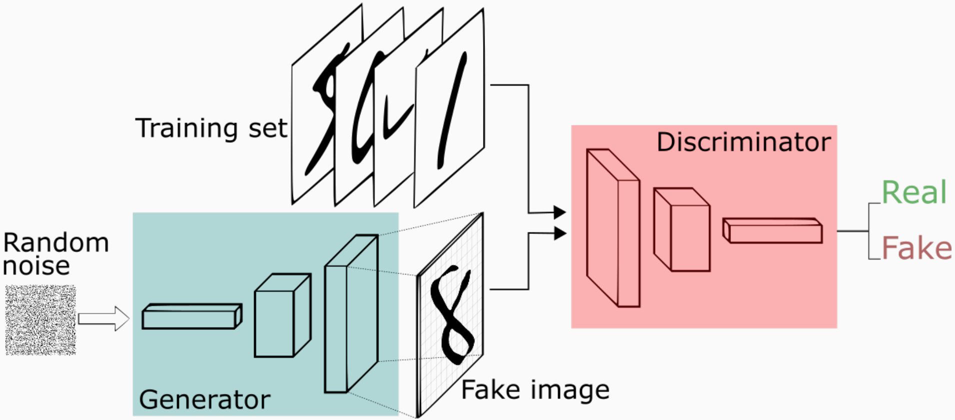


Figure 7: The architecture of a GAN. Source: Freecodecamp.org, Thalles Silva

What Can GANs Do?

Figure 8: Faces from a GAN by NVIDIA. Source: Karras, Laine and Aila 2018.

Sources

Books

Chollet, François and J.J. Allaire. 2018. *Deep Learning with R*. Manning Publications.

Goodfellow, Ian and Yoshua Bengio and Aaron Courville. 2016. *Deep Learning*. MIT Press.

Internet Resources

Ng, Andrew. *Deep Learning Specialization*. `coursera.org`.

Deep Learning Papers Reading Roadmap
`https://github.com/floodsung/Deep-Learning-Papers-Reading-Roadmap`.

Introduction to Machine Learning
`https://introduction-to-machine-learning.netlify.app/`.

Working with Deep Learning? Get in touch!

Dr Christian Arnold
Cardiff University
@chrisguarnold

Marcel Neunhoeffer
LMU Munich
@mneunho

Appendix

Ethics of Doing Social Science in Times of Big Data

The Power of AI Systems

- Face2Face
- Adobe VoCo
- Google Duplex

Digitalisation and Internet as Data Drivers

The Hunger for Data

- Digitalisation leads to collect tremendous amounts of data
- IBM: In the last 2 years, humanity has collected more data than between ever and 2 years ago
- How are *you* generating collecting data every day?
- Can you use that data?

Who Owns the Technology?

- Who is working on AI?

Who Owns the Technology?

- Who is working on AI?
- What is the role of universities?
- What is the role of companies?

Who Owns the Data?

- What is the role of data in deep learning?

Who Owns the Data?

- What is the role of data in deep learning?
- Who owns the data we produce?

Who Owns the Data?

- What is the role of data in deep learning?
- Who owns the data we produce?
- How can researchers develop algorithms?

Appendix

Further Material

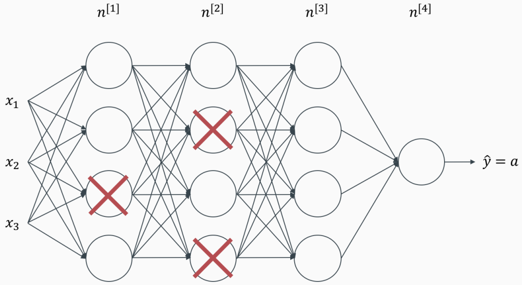
Overfitting: L2 Regularisation

At the Whiteboard

- How does L2 regularisation work for deep neural nets?

Overfitting: L2 Regularisation Intuition

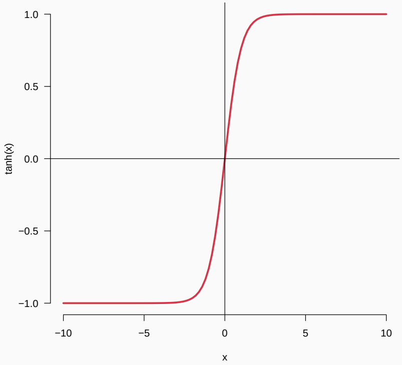
Intuition 1



- High λ forces weights to be close to 0
- Some nodes almost cancel out
- Net simplifies
- Tackles overfitting on demand

Overfitting: L2 Regularisation Intuition

Intuition 2



- As long as z is close to 0, no problem
- But if z large, L_2 Regularisation penalises to become close
- Node becomes almost linear
- No non-linearity possible