Deep Learning

Bundesbank Workshop on Deep Learning Day 1

Christian Arnold, Sebastian Sternberg & Marcel Neunhoeffer

28.03.2019

Introduction

Introduction

Motivation

This is the most exciting time to be a quantitative social scientist.

Ever.

Our Background

Present Lecturers

Your Expectations

You and DL

- Why did you enrol in this DL class?
- What do you expect to take home?
- What was the first time you came in contact with DL?

Deep Learning Workshop

Day 1: Introduction

- Motivation
- How does it work?
- Running your first models

Day 2: Advanced Applications

- Deep Learning for Sequences (Timeseries)
- Generative Deep Learning
- New Developments in Deep Learning

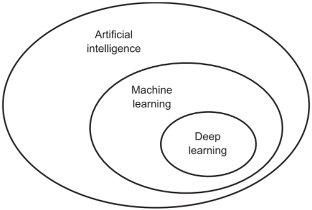
Today's Agenda

Introduction

Deep Learning: Why Now?

AI, ML & DL

- **Artificial Intelligence**: Any technique which enables computers to mimic human behavior
- **Machine Learning**: Subset of AI techniques which uses statistical methods to enable machines to improve with experience
- **Deep Learning**: Subset of ML which make the computation of multi-layer neural networks feasible.



What DL Can Do

- Digital assistants such as Google Now and Amazon Alexa
- Near-human-level autonomous driving
- Superhuman Go playing
- Mastering complex video games
- Near-human-level image classification, speech recognition, handwriting transcription.
- Etc.

$\Rightarrow$ Humanity is still exploring the full extent of what deep learning can do. And how it actually works...

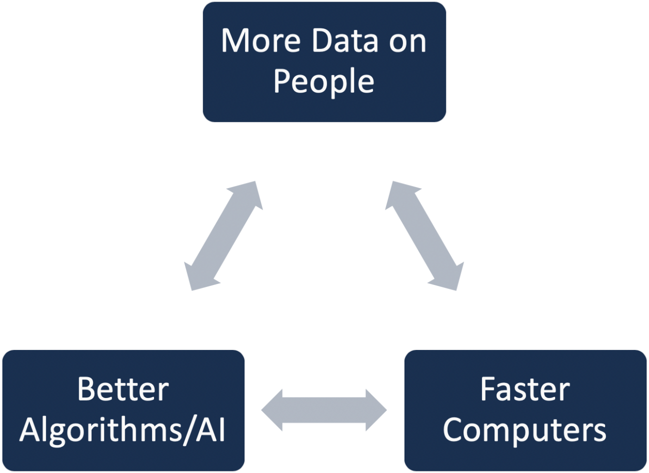
DL in Our Work

- **Sentiment Analysis**: Word Embeddings to adapt a dictionary to a certain domain
- **Binary sentence classification**: identifying vague sentences in a large corpus of court decision
- **Imputation:** Using Generative Adversarial Networks to impute missing values
- **Synthetic data:** Using a Generative Adversarial Network to create synthetic micro data (more on that tomorrow)
- **Fraud detection:** Using satellite images to detect voting fraud in remote areas

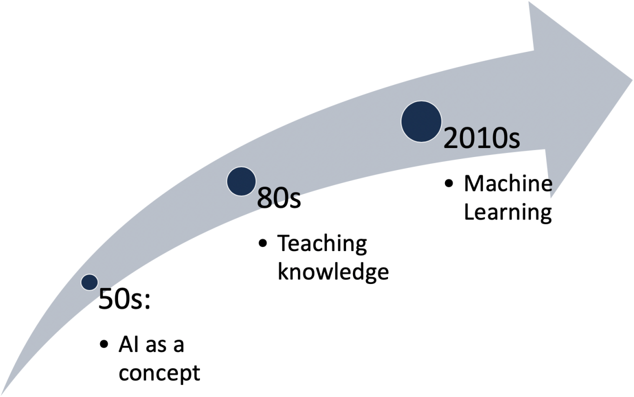
Relevant Applications for Bundesbank

- Chakraborty and Joseph (2017): Forecast UK CPI inflation using different machine learning techniques
- Rönnqvist and Sarlin (2017): Study of financial risk: predictive model that is able to detect infrequent events based on text data
- Fischer and Krauss (2018): Financial market predictions: long short-term memory networks (more on that tomorrow)
- Lecun, Bengio and Hinton (2015): Overview paper on deep learning in the “Nature”

Why DL Now?



AI Waves

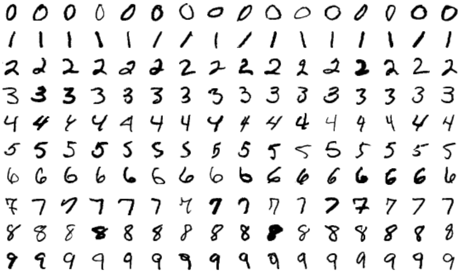


50s:
• AI as a concept
80s
• Teaching knowledge
2010s
• Machine Learning

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

The “Deep” in DL

Learning Data Representations in Multiple Stages

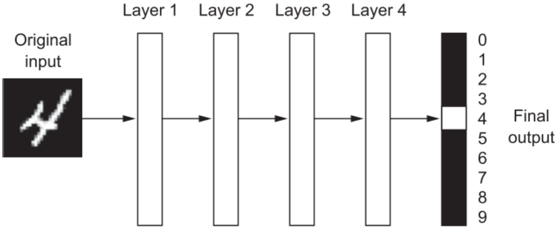


Figure 1: Source: Allaire/Chollet (2018)

What is Happening at Each Stage

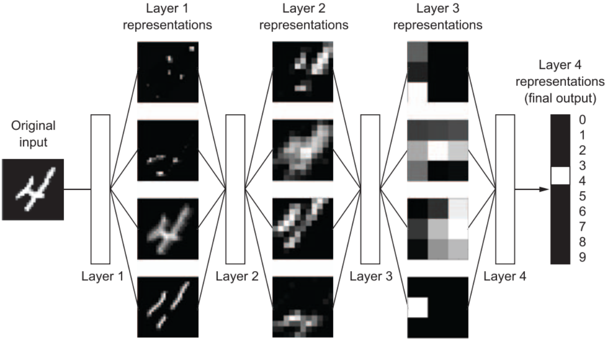


Figure 2: Source: Allaire/Chollet (2018)

Training Deep Neural Nets

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 a Logistic Regression

Four Steps

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

At the Whiteboard

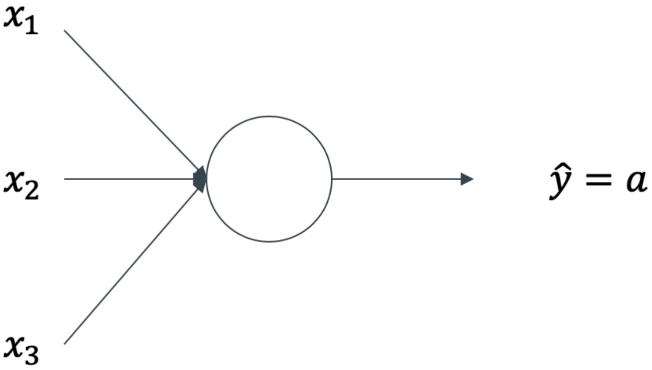
- A refresher in logistic regression
- Gradient descent
- Computation graphs
- Logistic regression with backpropagation of errors

Training Deep Neural Nets

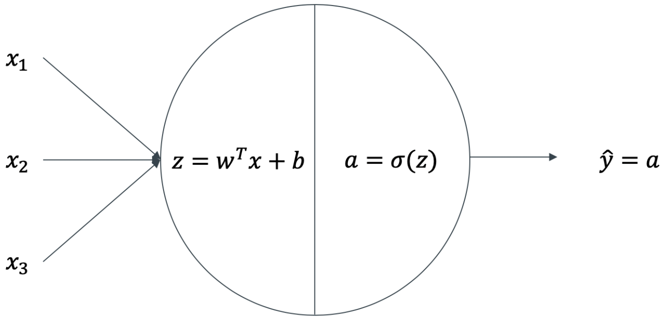
Shallow Neural Nets

What Did We Do?

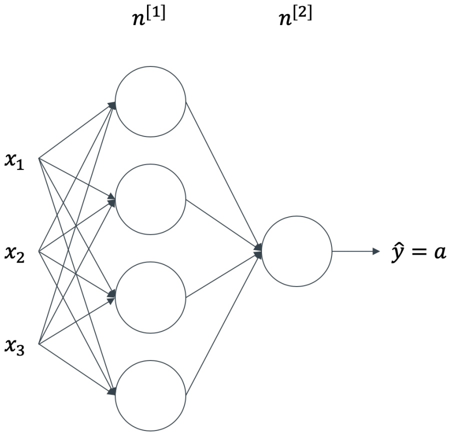
Logistic Regression



Zooming In On One Neuron

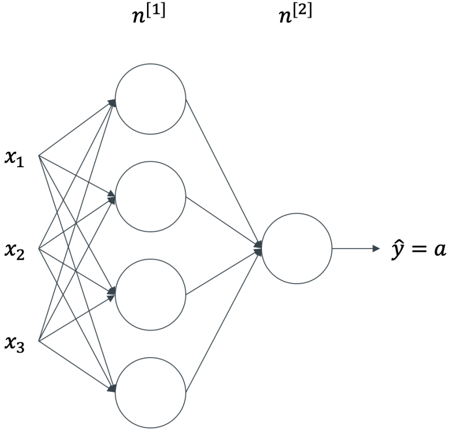


(Shallow) Neural Net



The Different Layers

- Input layer
- Hidden layer
- Output layer



Four Steps

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

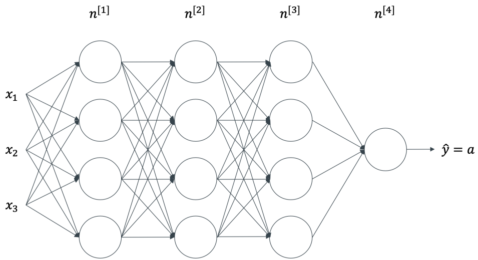
At the Whiteboard

- 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

Code Nr. 1

A Geometric Interpretation of a DNN

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

Code Nr. 2

Tuning

Parameters and Hyperparameters

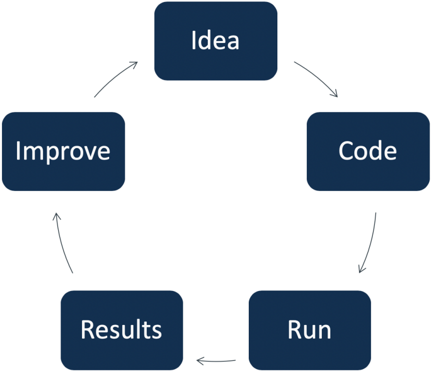
Parameters

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

Hyper Parameters

- Learning rate α
- # iterations
- # hidden layers l
- # hidden units $n^{[1]}, n^{[2]}, \ldots$
- choice of activation function

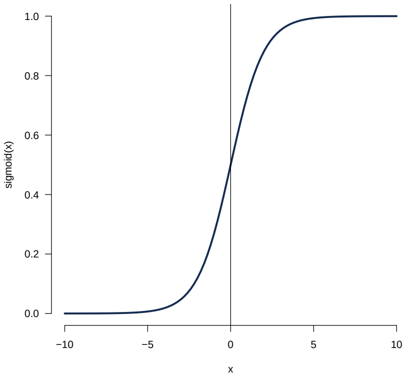
Tuning Your Model



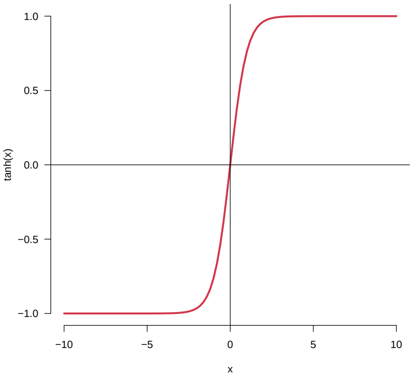
- DL is experimental
- HP depend on model and data
- Indicators: development of costfunction with iterations

Activation Functions

Sigmoid

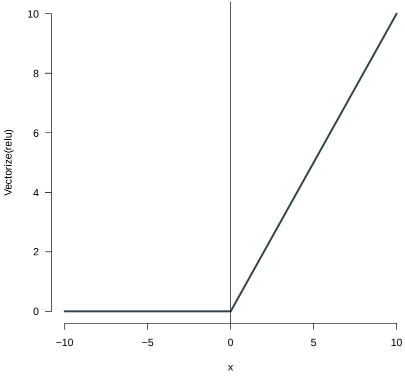


Tanh

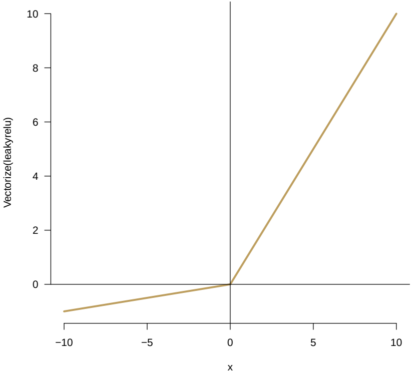


Activation Functions

ReLU



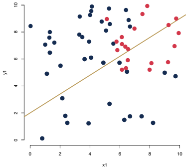
Leaky ReLU



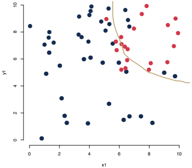
Understanding Activation Functions

At the Whiteboard

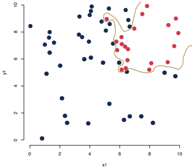
- Why do we need an activation function?



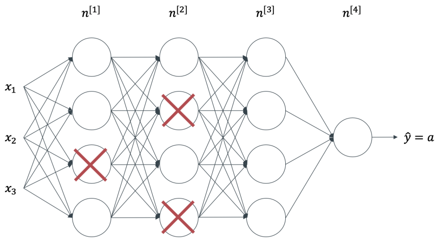
Underfitting

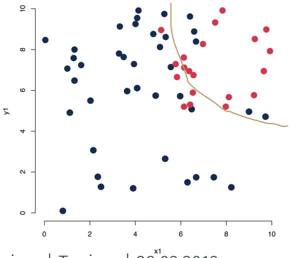


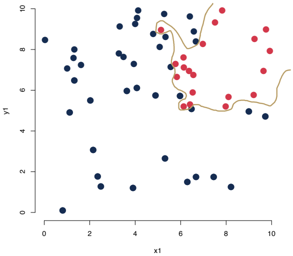
Just Right



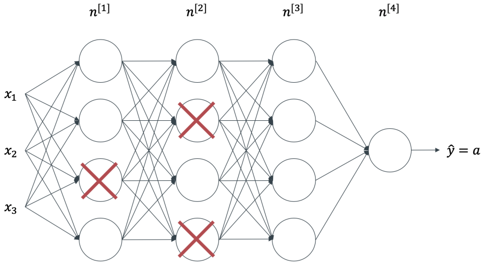
Overfitting


$n^{[1]}$
$n^{[2]}$
$n^{[3]}$
$n^{[4]}$
x_1
x_2
x_3
$\hat{y} = a$


y1
x1


y1
x1

Overfitting: Dropout



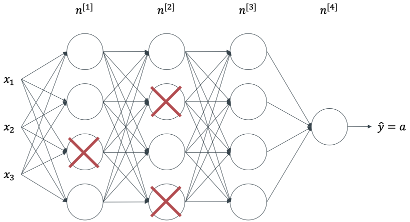
- Randomly eliminate nodes in the network
- Dropout can be particularly useful for the layers with many parameters
- But: cost function J is no longer defined

At the Whiteboard

- How does L2 regularisation work for deep neural nets?

Overfitting: L2 Regularisation Intuition

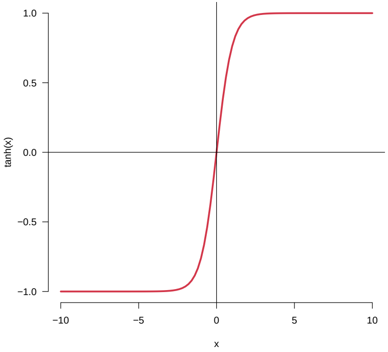
Intuition 1



- 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

Playtime

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

Code Nr. 3

Ethics of Doing Social Science in Times of Big Data

The Power of AI Systems

- Face2Face
- Adobe VoCo
- Google Duplex

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?
- 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 we produce?
- How can researchers develop algorithms?

Appendix

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. github.com.

References

Chakraborty, Chiranjit and Andreas Joseph. 2017. “Staff Working Paper No . 674: Machine learning at central banks.”.

Fischer, Thomas and Christopher Krauss. 2018. “Deep learning with long short-term memory networks for financial market predictions.” *European Journal of Operational Research* 270(2):654–669.
URL: *https://www.sciencedirect.com/science/article/abs/pii/S0377221717310652*

Lecun, Yann, Yoshua Bengio and Geoffrey Hinton. 2015. “Deep learning.”.

References II

Rönnqvist, Samuel and Peter Sarlin. 2017. "Bank distress in the news: Describing events through deep learning." *Neurocomputing* 264:57–70.