Data Science Campus
Coffee and Coding Session: Advanced Level

Alex Noyvirt
Claus Sthamer
Data Science Campus, ONS

26 January 2022
Deep learning & Neural networks (NN)

- Deep learning – successive layers of representations stacked on top of each other
- The model used for this is called a neural network
How a neural network (NN) works?
Loss and Loss Functions

• Cross Entropy Loss/Negative Log Likelihood:

\[ L(\theta) = -\sum_{i=0}^{N} \hat{y}_i \cdot \log(y_i) \]

\( \hat{y}_i \) : entries in the prediction vector \( \hat{y} \)

\( y_i \) : entries in the ground truth label \( y \)

• Mean Squared Error Loss Function:

\[ L(\theta) = \frac{1}{N} \sum_{i=0}^{N} (y_i - \hat{y}_i)^2 \]

\( \hat{y}_i \) : entries in the prediction vector \( \hat{y} \)

\( y_i \) : entries in the ground truth label \( y \)

• Mean Absolute Percentage Error:

\[ MAPE = \frac{100\%}{N} \sum_{i=0}^{N} \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right| \]
Introduction to Tensors

A tensor is an N-dimensional array of data

Operations with tensors – Addition, Dot product, broadcast etc.

Frameworks for deep learning

Keras

TensorFlow

PyTorch
GPU Accelerators and Cloud Services

Libraries for accelerated data science – DL frameworks, Rapids, etc.
A NN classification example

An example of simple fully connected NN classifying human activities from a accelerometers (pyTorch version):
https://rb.gy/6bnqnu
The importance of ETL and feature engineering

• Are there gaps in the data?
• Is the dataset imbalanced?
• Are you feeding the right data types to the model?
• Are you using the right range for the input variables?
• Is it possible to extract features that will help the model?
• How are you planning to evaluate the performance of the model?
The importance of model performance metrics

• Accuracy – not entirely appropriate for unbalanced data and not handy if we are interested in a single class \( \frac{TP+TN}{P+N} \)

• Precision: \( \frac{TP}{TP+FP} \)

• Recall: \( \frac{TP}{TP+FN} \)
Side by side comparison of several ML methods

Please follow: rb.gy/6bmqnu
Deep learning methods need more training data

Image from Andrew Ng's talk at Bay Area Deep Learning School (25th - 26th September, 2016)
Is your NN overfitting or underfitting?

How to prevent overfitting in neural network?

- Provide more training data / use data augmentation
- Use regularization (L1, L2)
- Use dropout
- Simplify the model
- Early stopping
Teaser question

• How to know that we have a sufficiently complex model?
Many different type of neural networks

- CNN
- UNETs
- Resnet
- Autoencoders/ VAE
- RNN/LSTM/GRU
- GANs
- Transformers (self-attention)
Generating synthetic UK Census data with GAN

What is a Generative Adversarial network?

Please follow:
https://rb.gy/zwbqyr
Recommended further reading/watching: Recent advances in Deep Learning

Please follow:
rb.gy/gtsgwp
Questions?