Deep Learning

COMP2002

Lauren Ansell

Introduction

Today's topics:

- Deep neural networks
- When to use deep learning

Session learning outcomes - by the end of today's lecture you will be able to:

- Explain the structure of a deep neural network and describe its potential uses
- Explain why providing insight into the operation of a deep learning method is important and difficult
- Understand some of the difficulties involved in using deep learning

History Of Deep Learning

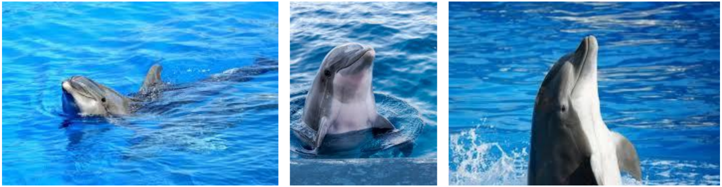
Neural networks are the cornerstone of deep learning.

Neural networks first rose to fame in the late 1980s.

They fell out of favour when SVM, boosting and random forests came along.

Neural networks resurfaced after 2010 under the new name *deep learning*.

ImageNet Dataset



Massive collection of over 14,000,000 labelled images.

Designed to overcome shortcomings with existing benchmarks – e.g. small number of categories.

Collection of nouns – each noun maps to a number of real-world images.

Based on sanitised Google Image and Flickr searches by Amazon Mechanical Turk workers

ImageNet Competition

Computer vision competition runs since 2010.

Classification of ImageNet images – 1.2 million training and 150,000 test images belonging to 1,000 classes.

Early competitions dominated by SVMs.

2012 – AlexNet achieved error rate of 15%.
By 2017 most of teams scored > 95% accuracy (> human accuracy on this task).

Top-5 Accuracy

- Model produces five class predictions
- Model is correct if one of the five is correct
- Guess “elephant”, “dog”, “dolphin”, “shark”, “bus” is correct
- Guess “elephant”, “dog”, “shark”, “bus”, “helicopter” is incorrect

Feature Extraction

In order
to effectively classify or regress raw inputs
must be extracted into high level features.

When using “ordinary ML” the
first step is feature extraction – e.g. PCA.

Deep learning is an example
of representation learning – raw data
is fed in and identifies representations
needed for accurate predictions.

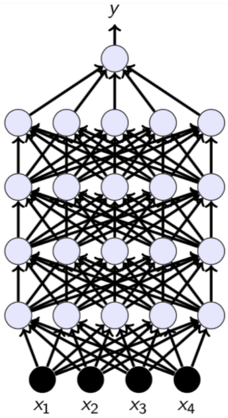
Feature extraction is done by the
algorithm not relying on human expertise.

141	145	146	131
142	142	141	103
142	151	109	108
141	143	107	105

Deep Learning

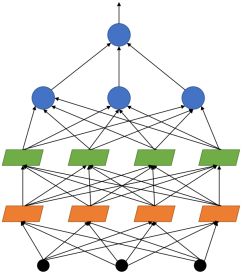
Neural networks with lots of layers

- Layers act as feature extractors
- Features
  are passed onto the next layer
- Raw
  input data is repeatedly processed
  before it reaches the output layer
- Result is a black box – so
  part of the drive behind explainable AI
- ANNs are prone to overfitting
  – so are deep learning models



Convolutional Neural Networks

- Pioneered by Yann LeCunn in the 1980s.
- Lower layers transform or convolve the inputs and pool features.
- Higher levels act as an ordinary neural network using the outputs of the lower layers as inputs.
- AlexNet was the winner of the 2012 ImageNet challenge



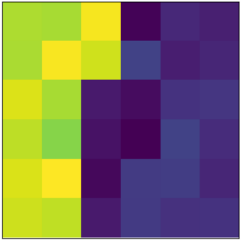
Convolutional And Pooling Maps

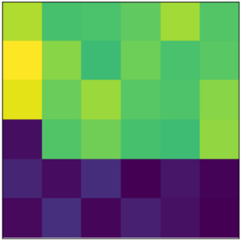
Convolutional layers

Units are organized into feature maps
Identify
local features (e.g. edges) – features
that can occur anywhere in the image

Pooling layers

Summarise
the features identified in the convolutional
layers without accounting for the location





Convolutional Layers

Each convolutional layer is made up of convolutional filters.

How convolution filters work

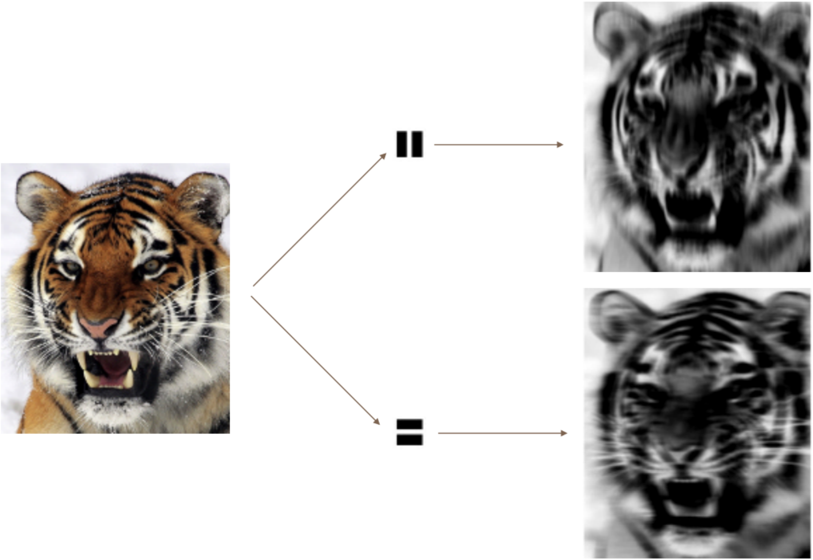
$$\text{Original Image} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \\ j & k & l \end{bmatrix} \quad \text{Convolution Filter} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$$

$$\text{Convolved Image} = \begin{bmatrix} a\alpha + b\beta + d\gamma + e\delta & b\alpha + c\beta + e\gamma + f\delta \\ d\alpha + e\beta + g\gamma + h\delta & e\alpha + f\beta + h\gamma + i\delta \\ g\alpha + h\beta + j\gamma + k\delta & h\alpha + i\beta + k\gamma + l\delta \end{bmatrix}$$

If the $n \times n$ submatrix of the original matrix is similar to the convolution filter, then it will have a large value in the convolved image.

This means that the convolved image highlights regions of the original image that resemble the convolution filter.

In general convolution filters are small ($\ell_1 \times \ell_2$ arrays).



In image processing, standard practice is to draw from filters which have been predefined.

With CNNs the filters are learned for the specific classification task.

- As the input is colour, there are three channels represented by a three-dimensional feature map.
- If we use K different convolution filters at the first hidden layer, we get K two dimensional output feature maps.
- We typically apply the ReLU activation function to the convolved image.

Pooling Layers

This layer provides a way to condense a large image into a smaller summary image.

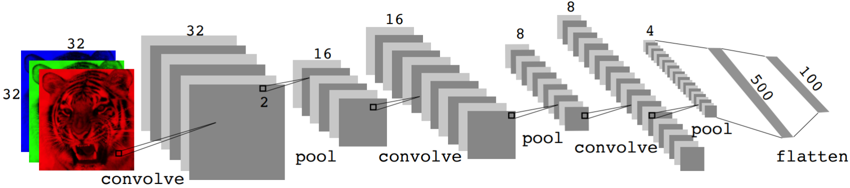
The max pooling operation summarises each non-overlapping 2×2 block of pixels in an image using the maximum value in the block.

Provides some location invariance.

$$\text{Max pool} \begin{bmatrix} 1 & 2 & 5 & 3 \\ 3 & 0 & 1 & 2 \\ 2 & 1 & 3 & 4 \\ 1 & 1 & 2 & 0 \end{bmatrix} \to \begin{bmatrix} 3 & 5 \\ 2 & 4 \end{bmatrix}$$

Architecture Of A CNN

The number of convolution filters in a convolution layer can be equated to the number of units at a particular hidden layer of a fully connected neural network.



Each subsequent convolve layer is similar to the first.

Since the channel feature maps are reduced in size after each pool layer, the number of filters in the next convolve layer is increased.

Occasionally several convolve layers are repeated before a pool layer.

Once each channel feature map has been reduced to the size of a few pixels in each dimension, we say they have been *flattened*.

There are many tuning parameters that need to be selected during construction of the network.

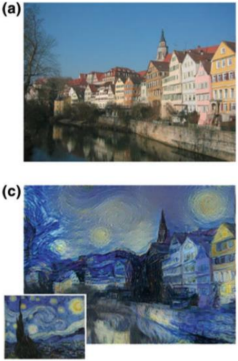
Creativity With Deep Learning

“Creating”
paintings with deep learning

Model takes two images
and recreate one in the style of the other

Deep learning – 21 layers

CNN – assigns importance
to features identified in the image



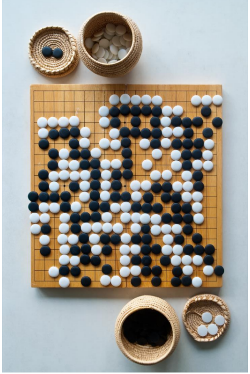
AlphaGo

AlphaGo combines a CNN with
a Monte Carlo Tree and is trained using
reinforcement learning while playing itself.

AlphaGo
also uses experience from human players.

AlphaGo
Zero doesn't use human experience.

CNN analyses the board to compute the
probability of the current player winning
given the current state of the board and
the best move from the current position

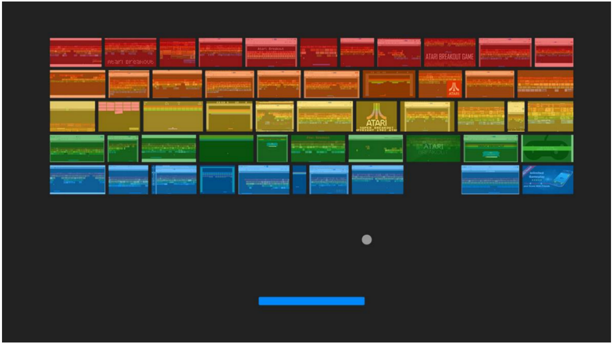


Breakout

Play Atari game from 1970s using deep reinforcement learning.

Q-learning – learn a state transition policy to determine the next move (a useful move leads to a reward).

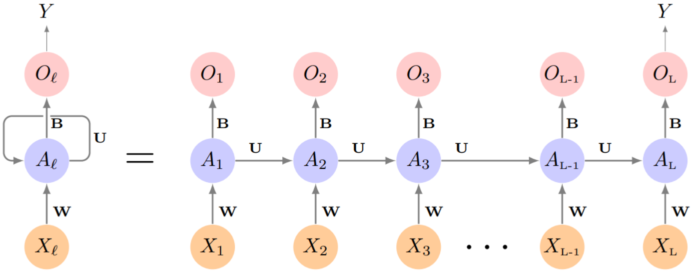
CNN learns the Q table – inputs are pixel positions the most recent few frames.



Recurrent Neural Networks

RNNs allow previous outputs to be used as inputs so that history can be incorporated.

Useful in natural language processing and sentiment analysis.

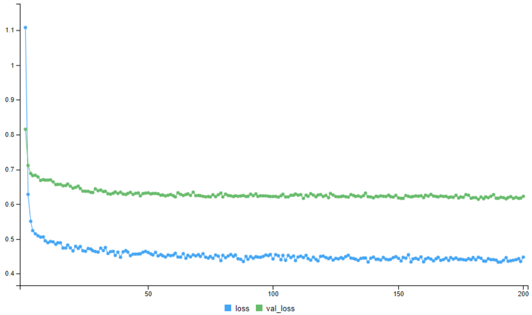


RNN Example

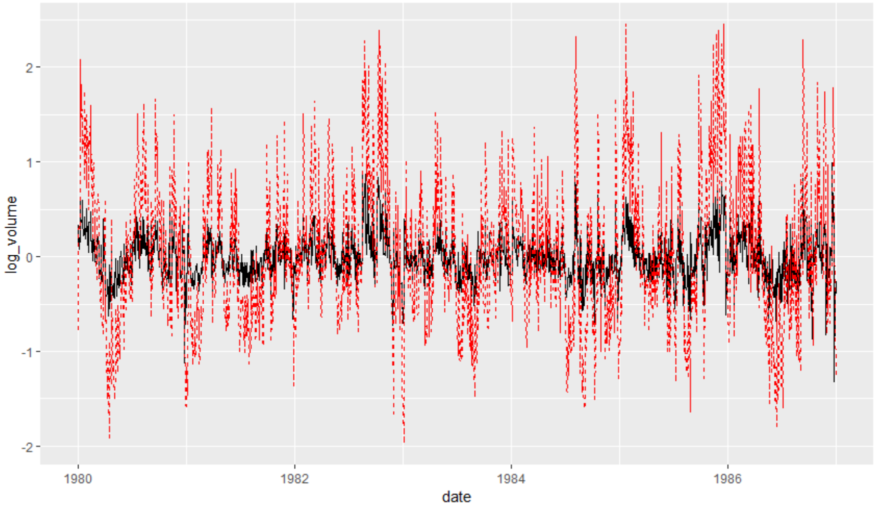
Use a RNN for time series prediction.

New York Stock Exchange data from December 3^{rd} 1962 to December 31^{st} 1986.

The model is fitted with 12 hidden units.



Results



Generative Adversarial Networks

Given a dataset generative models are unsupervised techniques for producing new examples that could belong to the dataset.

GANs enable supervised learning – train two sub-models:

- One to generate new examples
- One to classify examples as either real or fake

First proposed in 2014.

Deep Convolutional Generative Adversarial Networks proposed in 2015 and are the basis for most modern GANs.

Deepfakes

Explainable Deep Learning

Deep learning models are generally considered black boxes – it's not possible to see their inner workings.

Many application areas cannot have this lack of transparency.

Explainable AI is intended to address this by exposing how the models have generated a result.

ImageNet challenges have considered this – with a localization sub-challenge.

Difficulties With Deep Learning

Lots of data is required – this is becoming less of a challenge but there are still difficulties (e.g. getting labelled data).

Training deep learning networks requires substantial computational power because of massive number of networks weights to calibrate – GPUs often used.

Deep neural networks are generally for supervised learning – if we want to achieve general AI we will need unsupervised learning models.

“Deep” refers to the network architecture rather than understanding

When To Use Deep Learning

There are numerous applications for deep learning across multiple different sectors.

Comparison of neural networks and other methods

Model	# Parameters	MAE	Test set R^2
Linear Regression	20	254.7	0.56
Lasso	12	252.3	0.51
Neural Network	1409	257.4	0.54

We need to make a choice between performance and complexity.

Summary

Deep neural networks

- Lots of hidden layers
- Layers have specific roles – e.g. convolution and pooling
- RNNs for temporal encoding

ImageNet

- Massive dataset for object recognition and challenge 2010-2017