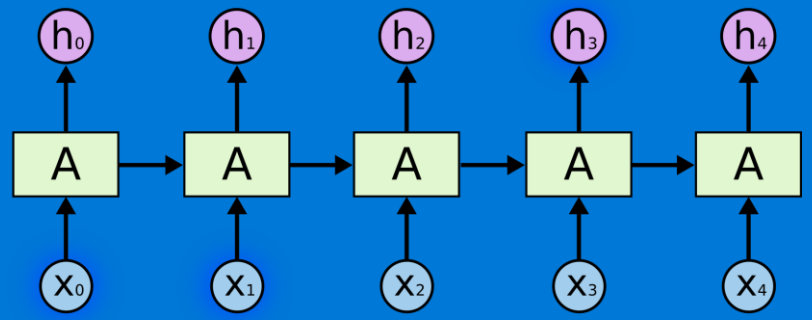
SEQUENCES + NLP

DEEP LEARNING SERIES (WWC-AI)



Tim Scarfe

Machine learning appreciator from the UK!

http://aka.ms/mdml

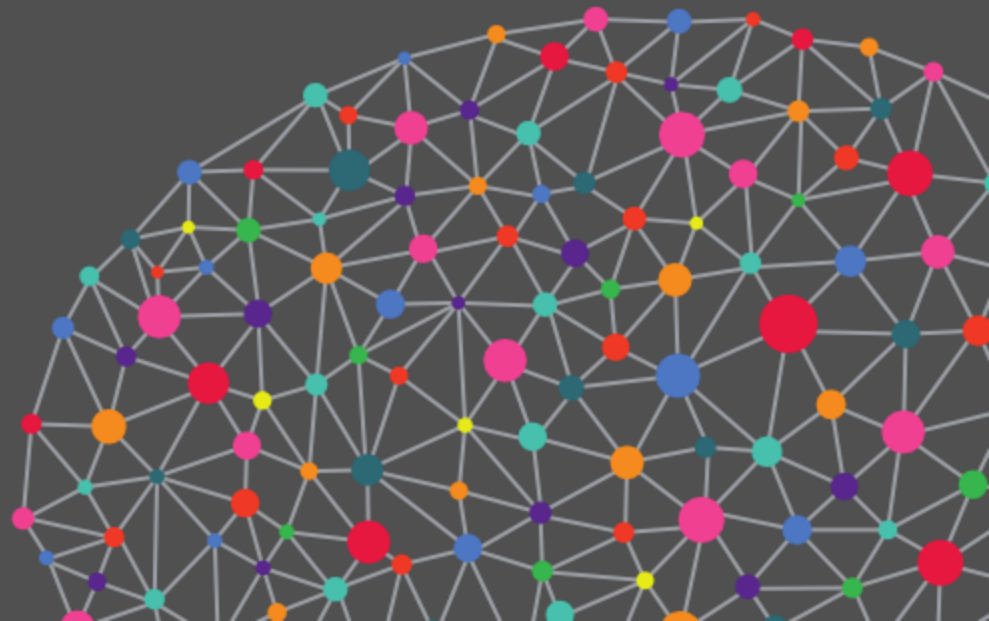
MACHINE LEARNING @ MICROSOFT

Azure

Microsoft

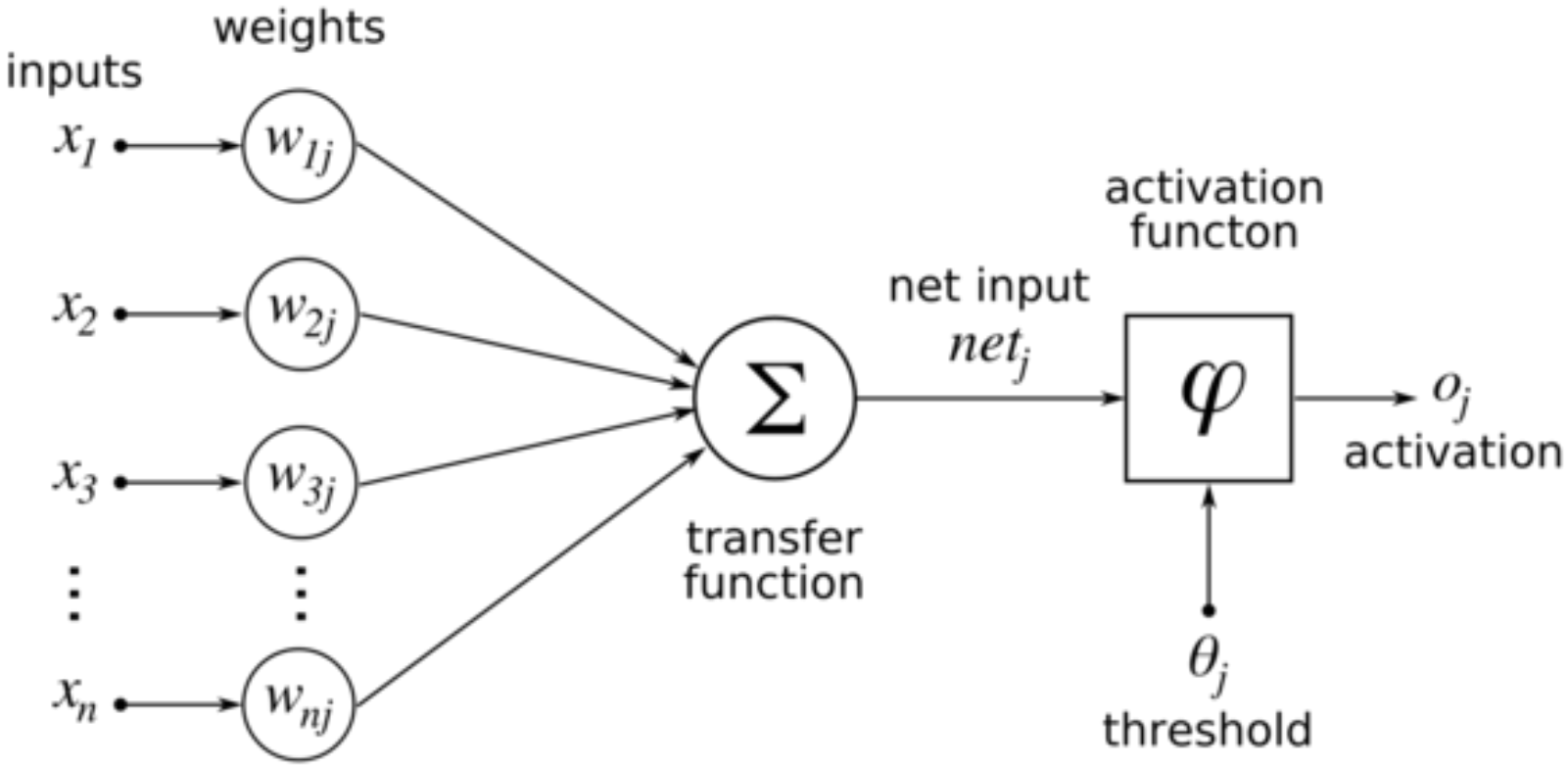
TALK OUTLINE

- Deep learning intro
- Distilled concepts of deep learning
- Why are neural networks good at sequence processing?
- What is sequence processing?
- Working with text data
- Recurrent neural networks
- 1d convolutional neural networks



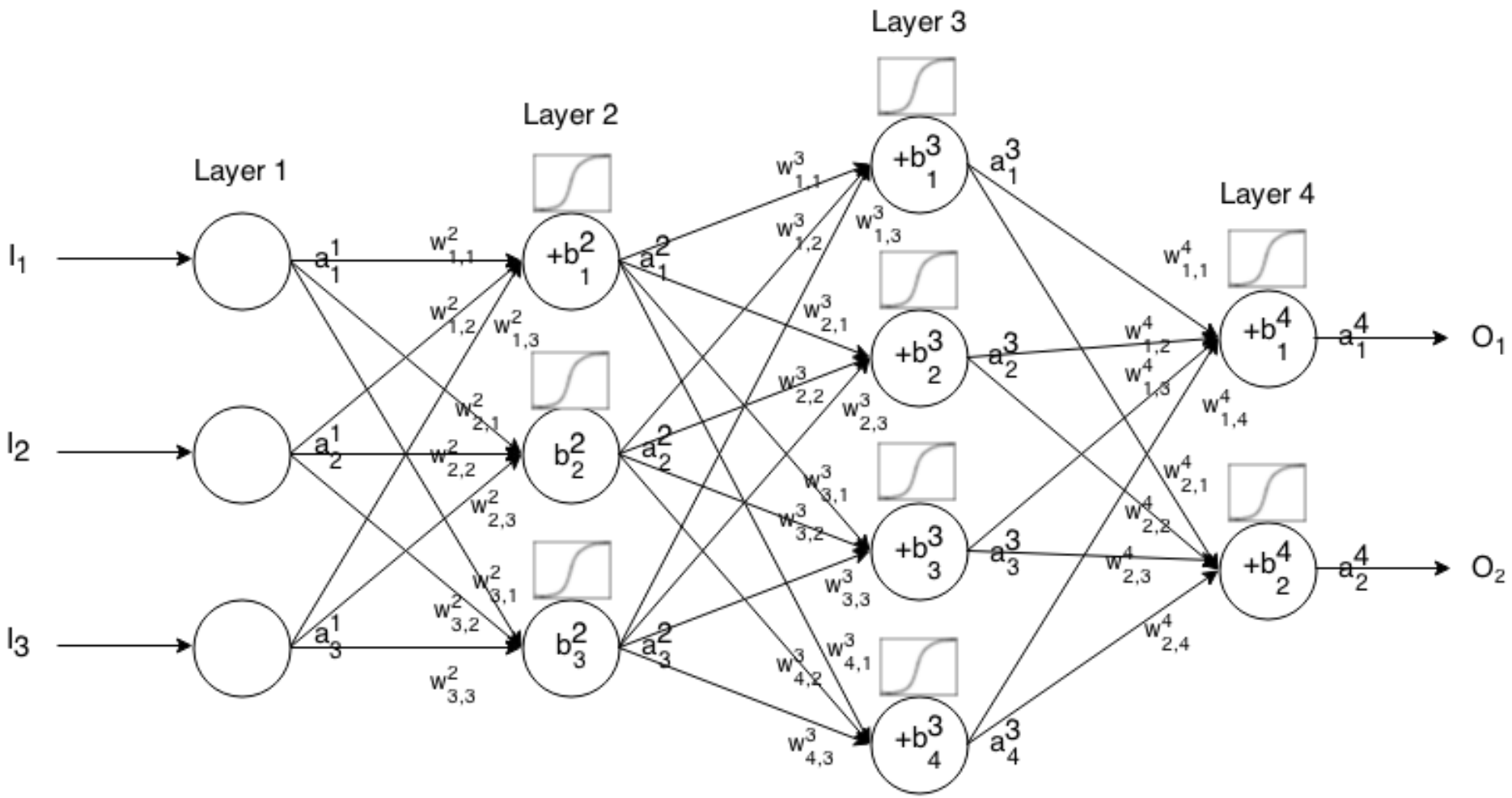
Microsoft

WHAT IS A NEURAL NETWORK?

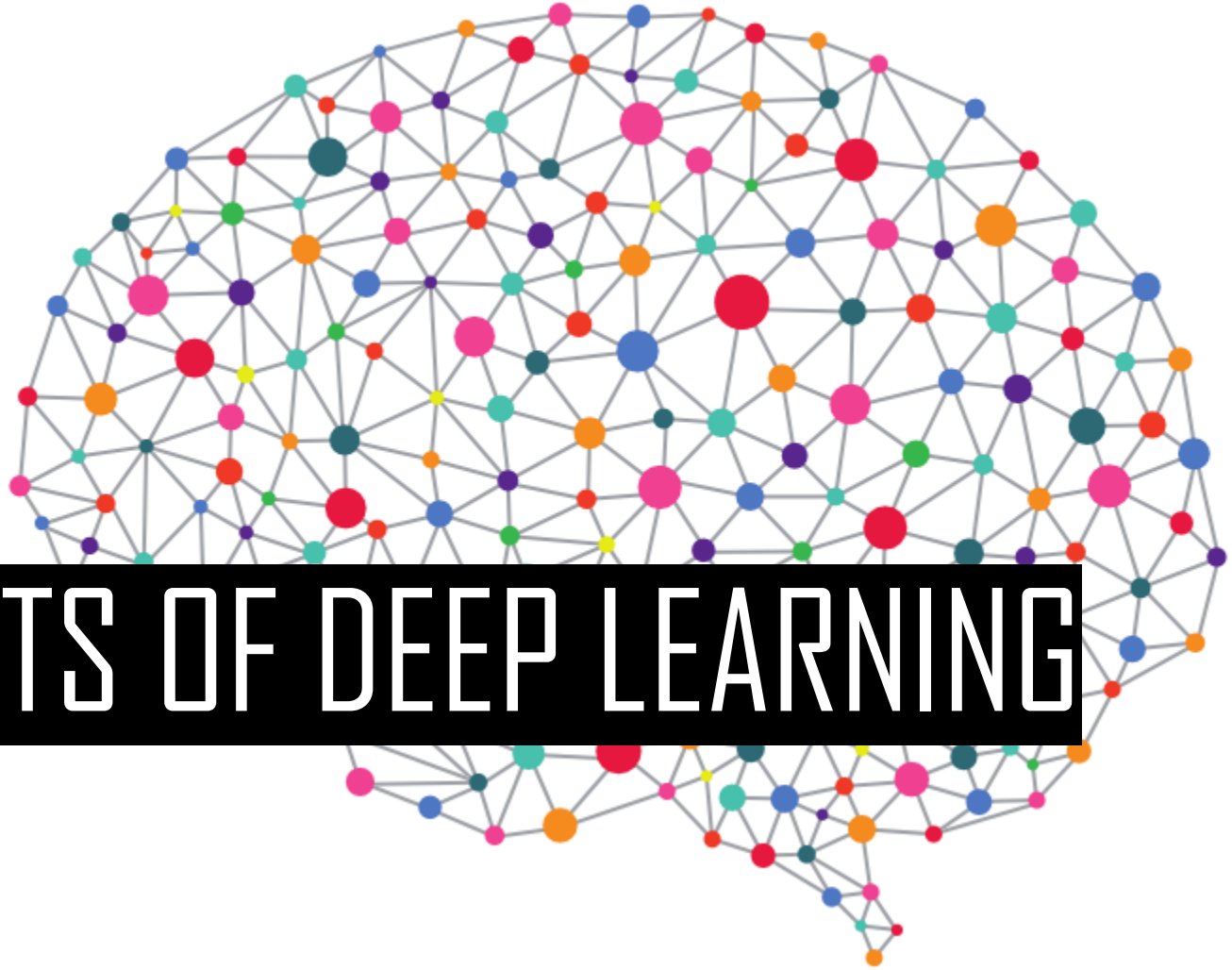


Microsoft

WHAT IS A DEEP NEURAL NETWORK?



Microsoft



DISTILLED CONCEPTS OF DEEP LEARNING

Microsoft

- The networks have **many levels** of depth
- Machine learns a hierarchy of representations
- **No feature extraction** required

Traditional ML

Hand crafted features

Feature Extractor → Trainable Classifier

Deep Learning

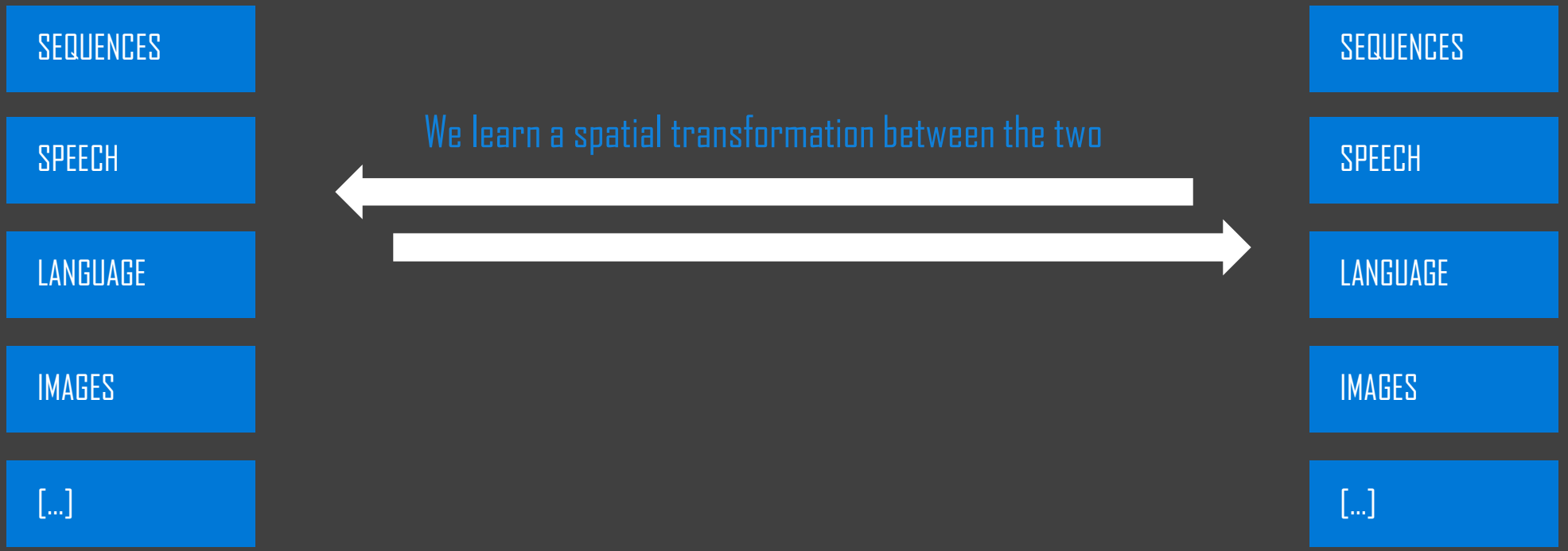
Representations are hierarchical and trained automatically

Low Level Features → Mid Level Features → High Level Features → Trainable Classifier

ENTIRE MACHINE IS TRAINABLE

Microsoft

- Unlike other shallow ML algorithms; you can map *between data domains*



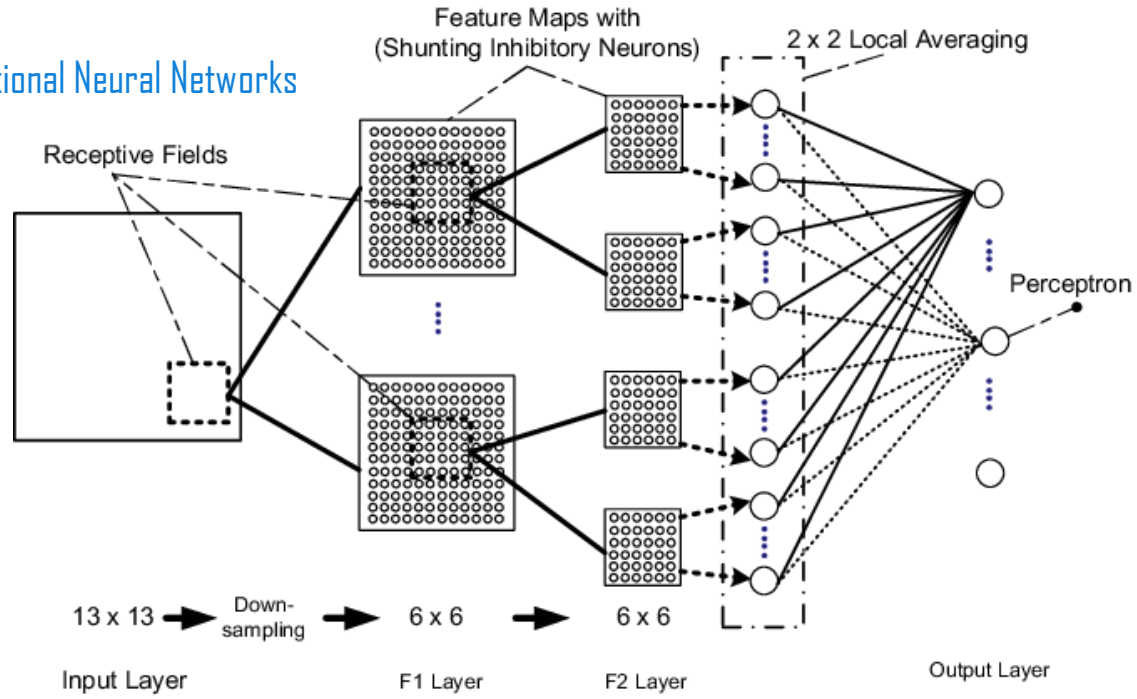
UNIVERSAL FUNCTION APPROXIMATORS

Microsoft

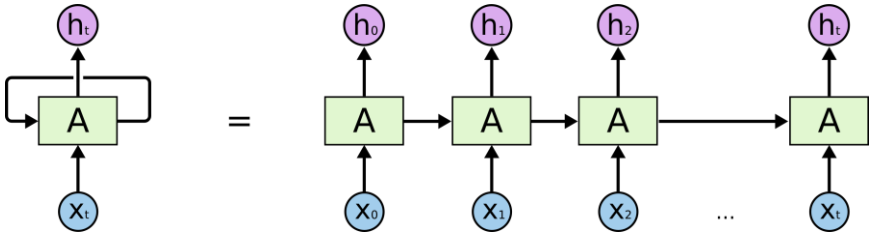
Unlike other algorithms, NNs can natively encode useful and obvious relationships in the data domain

- Local spatial dependencies (vision)
- Time dependencies (language, speech)

Convolutional Neural Networks



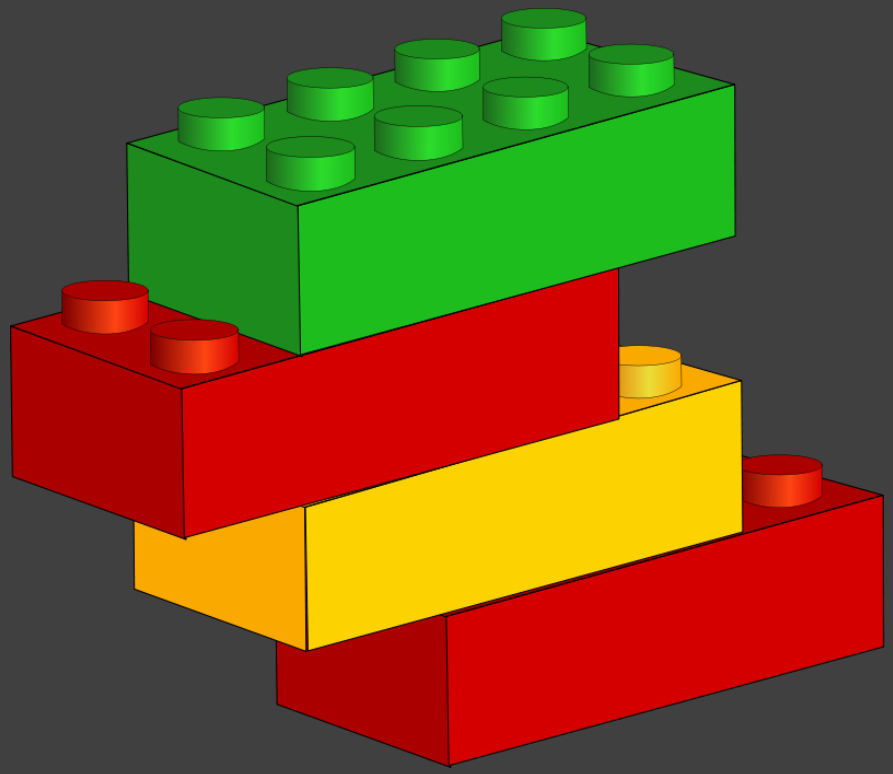
Recurrent Neural Networks



NATIVE DATA-DOMAIN FEATURES

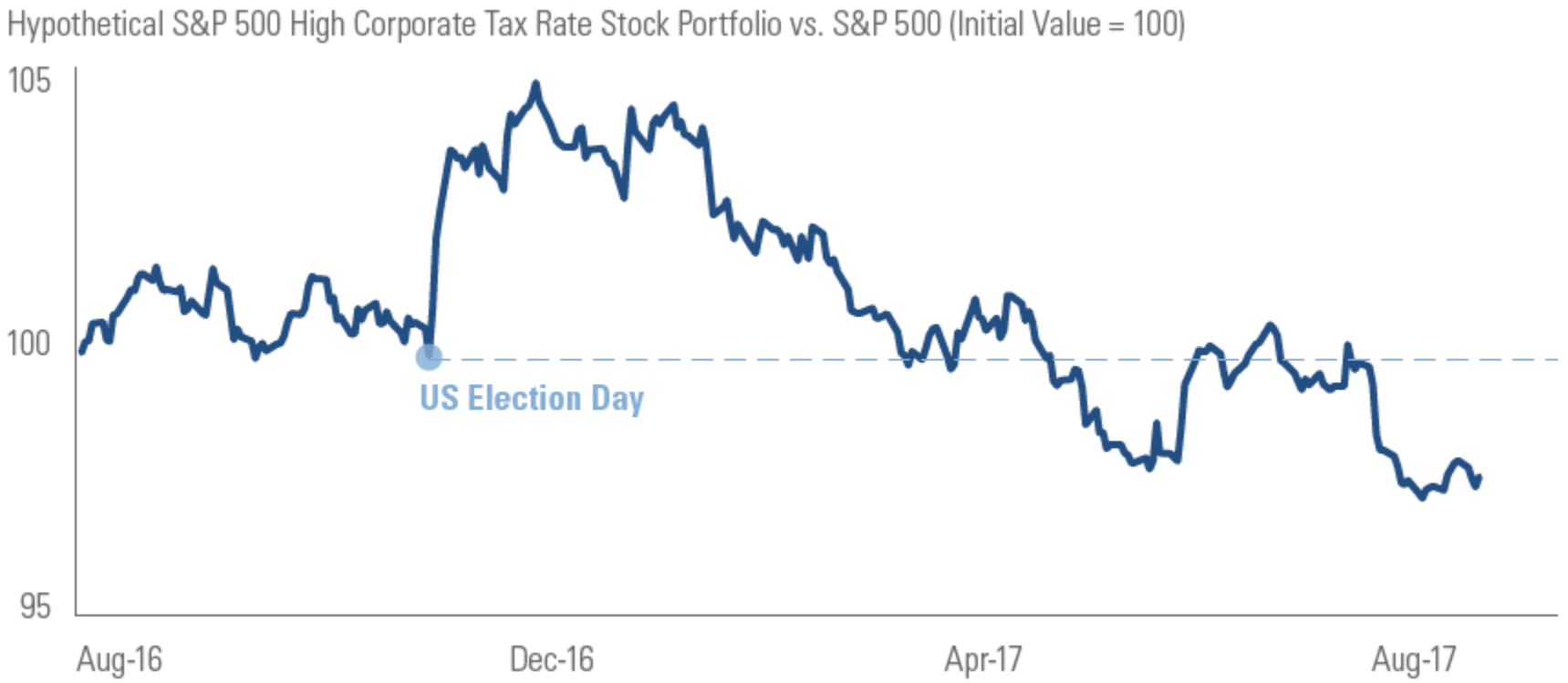
Microsoft

- Composability
- Deep Learning research is very applied
- Accessibility
- Software analogy



COMPOSABILITY

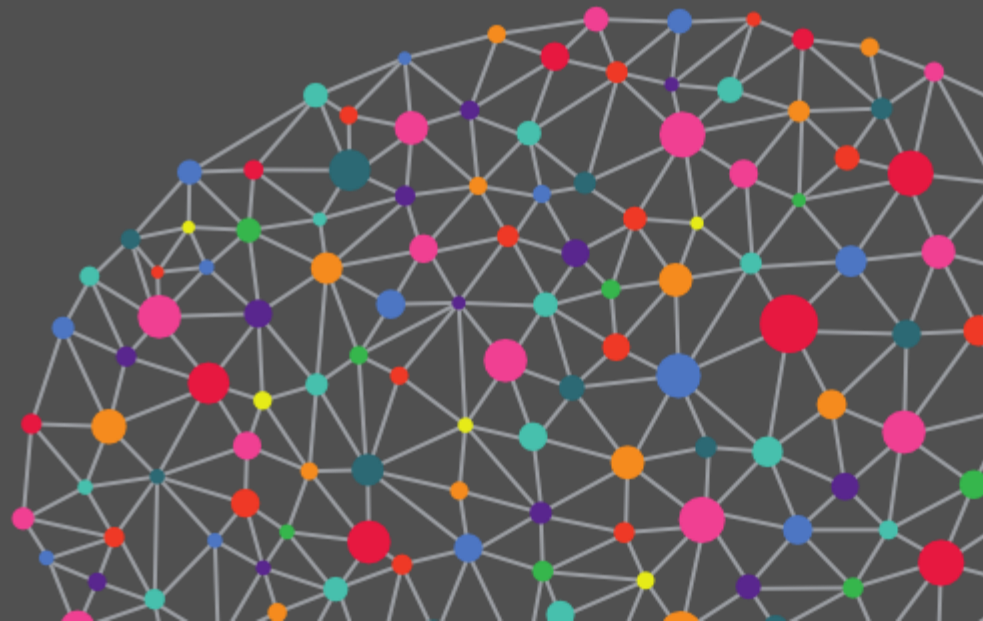
WHAT IS A SEQUENCE?



Microsoft

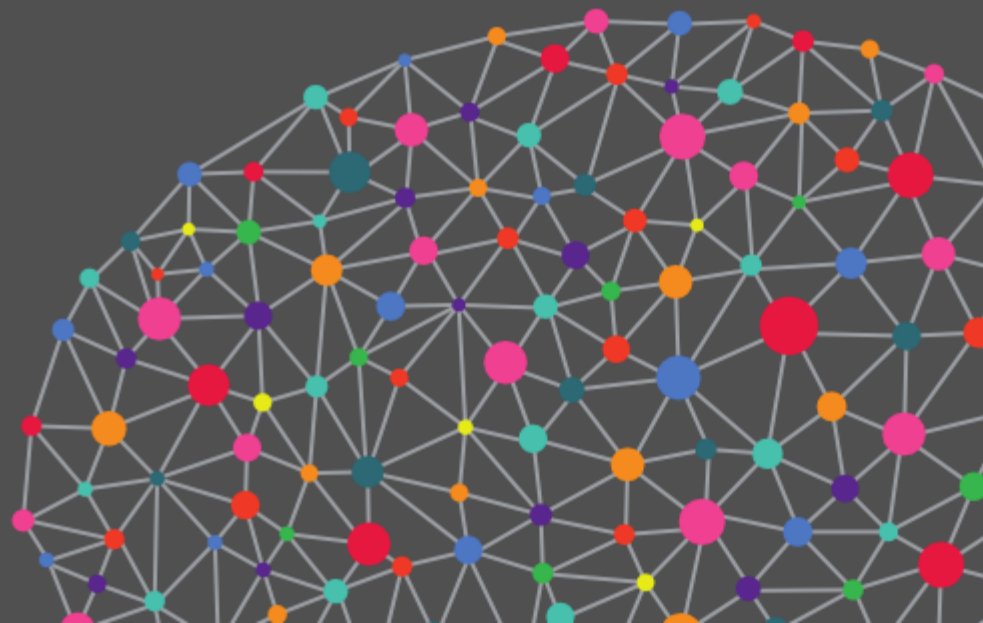
WHAT IS SEQUENCE PROCESSING

- RNNs
  - Timeseries classification
  - Anomaly detection in timeseries
  - Entity recognition
  - Revenue forecasting
  - Question + Answer
- 1d Convnets
  - Spelling correction
  - Document classification
  - Machine translation



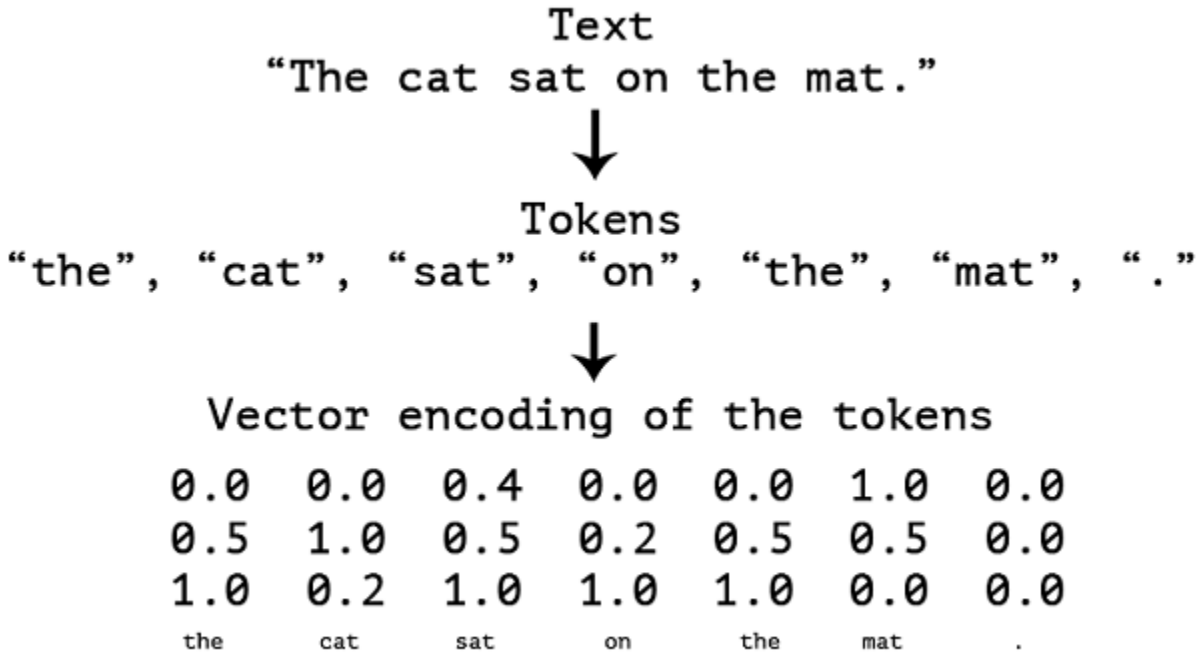
WHAT IS SEQUENCE PROCESSING

- RNNs
  - When global order matters
- 1d Convnets
  - Speed
  - Local temporal dependencies
- You can stack them!



TOKENIZATION

- Words
- Characters
- N-grams

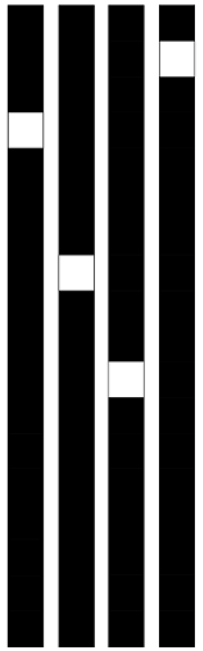


N-Grams example

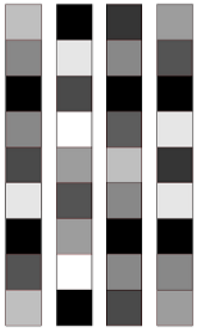
```
{"The", "The cat", "cat", "cat sat", "The cat sat",
  "sat", "sat on", "on", "cat sat on", "on the", "the",
  "sat on the", "the mat", "mat", "on the mat"}
```

Microsoft

WORD VECTORS VS WORD EMBEDDINGS?



One-hot word vectors:
- Sparse
- High-dimensional
- Hard-coded



Word embeddings:
- Dense
- Lower-dimensional
- Learned from data

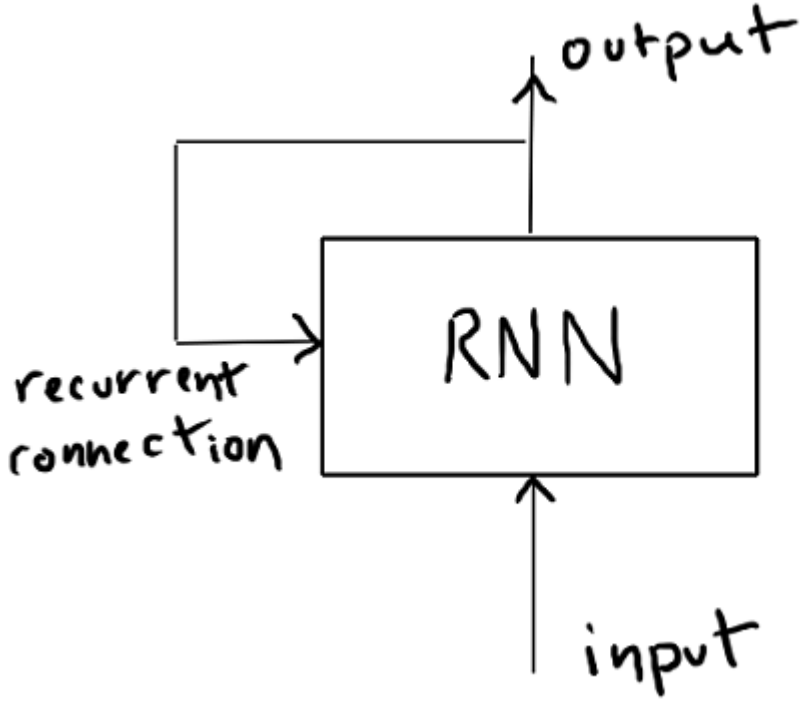
Microsoft

WORD EMBEDDINGS

- Word2Vec
- GloVe

Microsoft

RECURRENT NEURAL NETWORKS



RECURRENT NEURAL NETWORKS

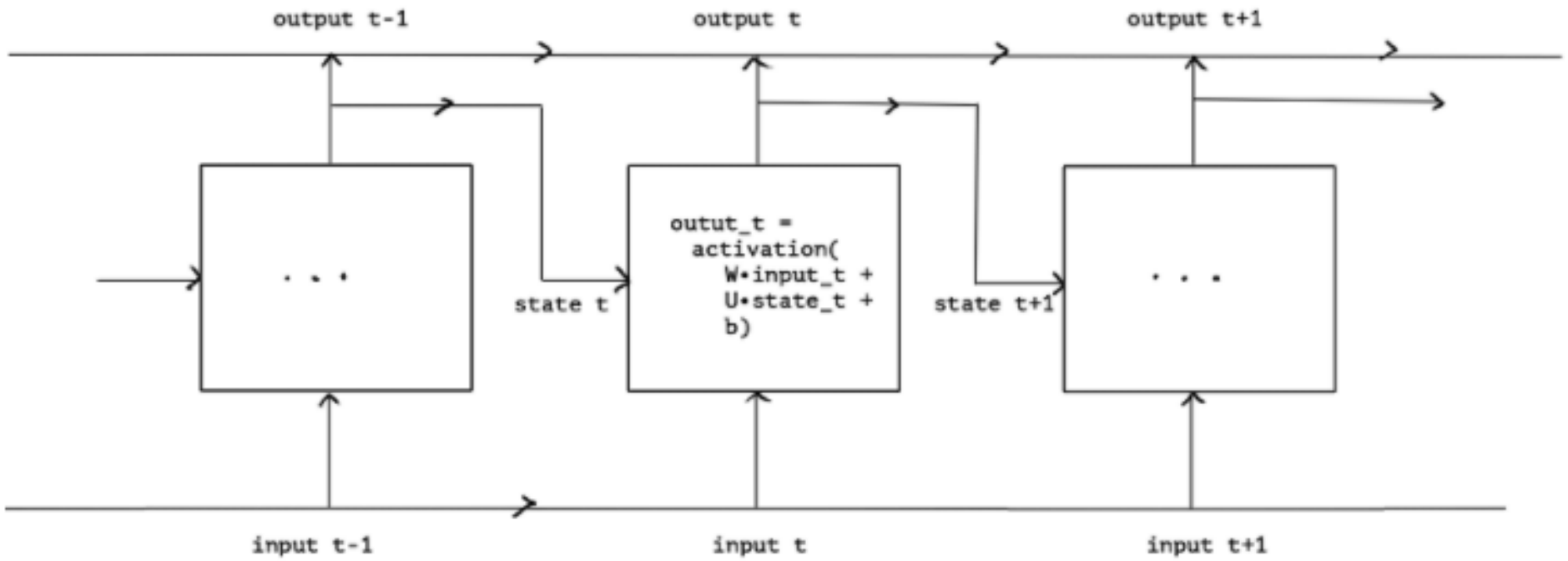


Image © Francois Chollet

LSTMS(1)

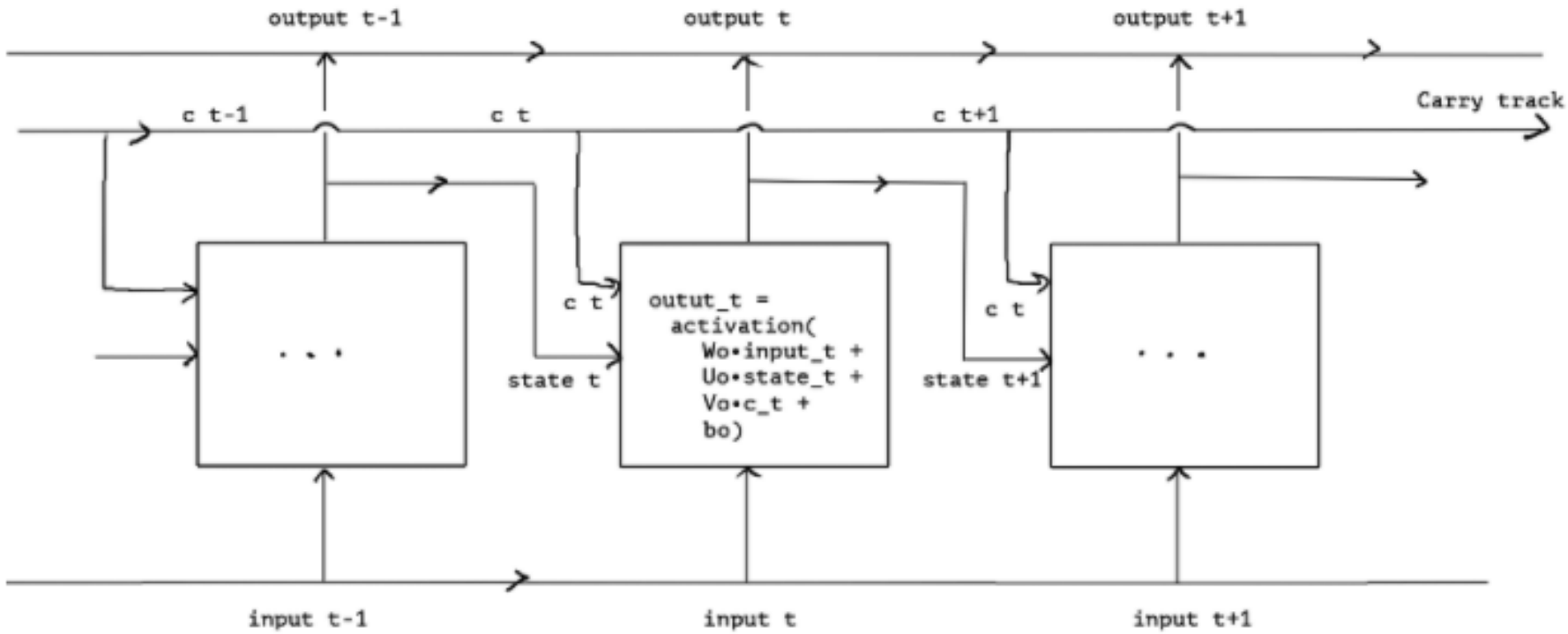
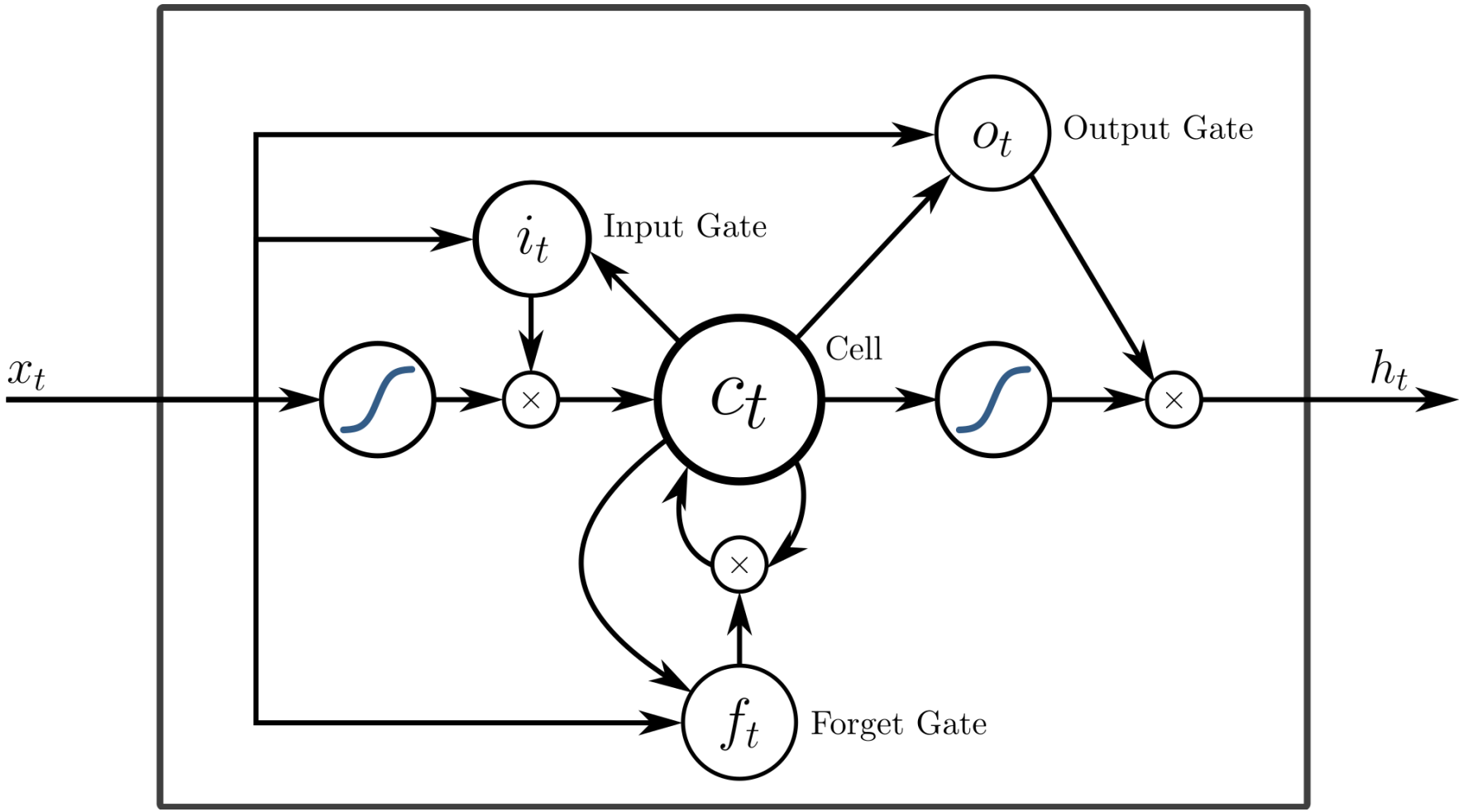


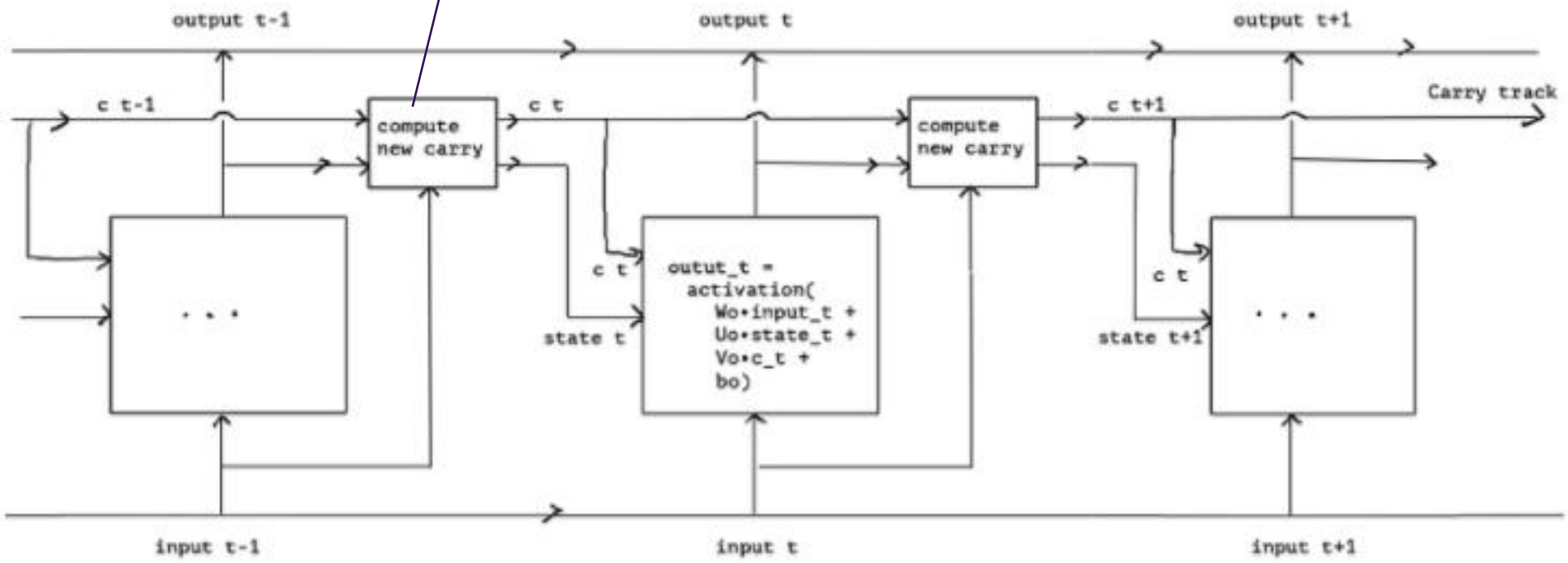
Image © Francois Chollet

LSTMS(2)



LSTMS(3)

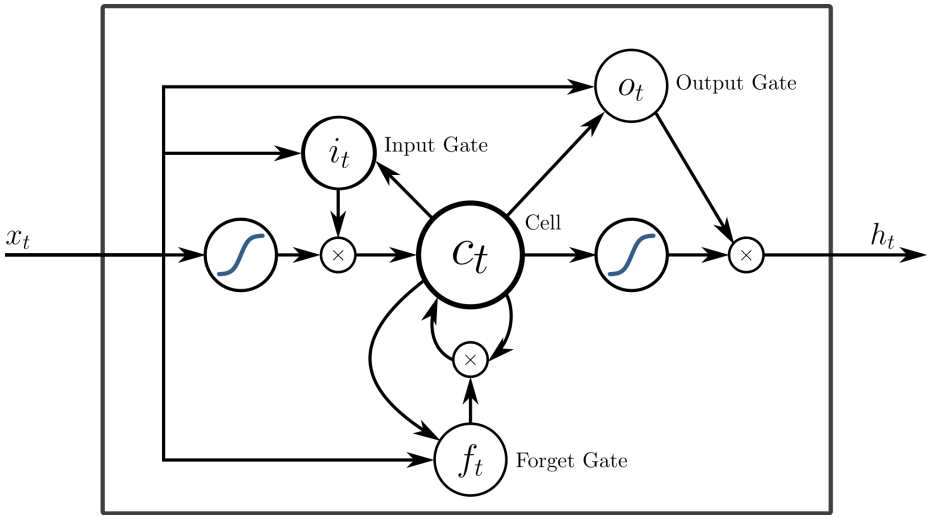
```
c_t+1 = i_t * k_t + c_t * f_t
```



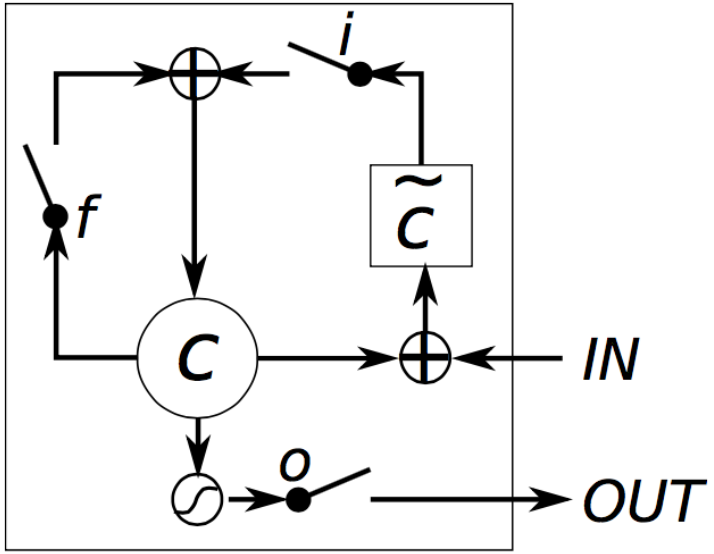

```
i_t = activation(dot(state_t, Ui) + dot(input_t, Wi) + bi)
f_t = activation(dot(state_t, Uf) + dot(input_t, Wf) + bf)
k_t = activation(dot(state_t, Uk) + dot(input_t, Wk) + bk)
```

Image © Francois Chollet

LSTM VS GRU



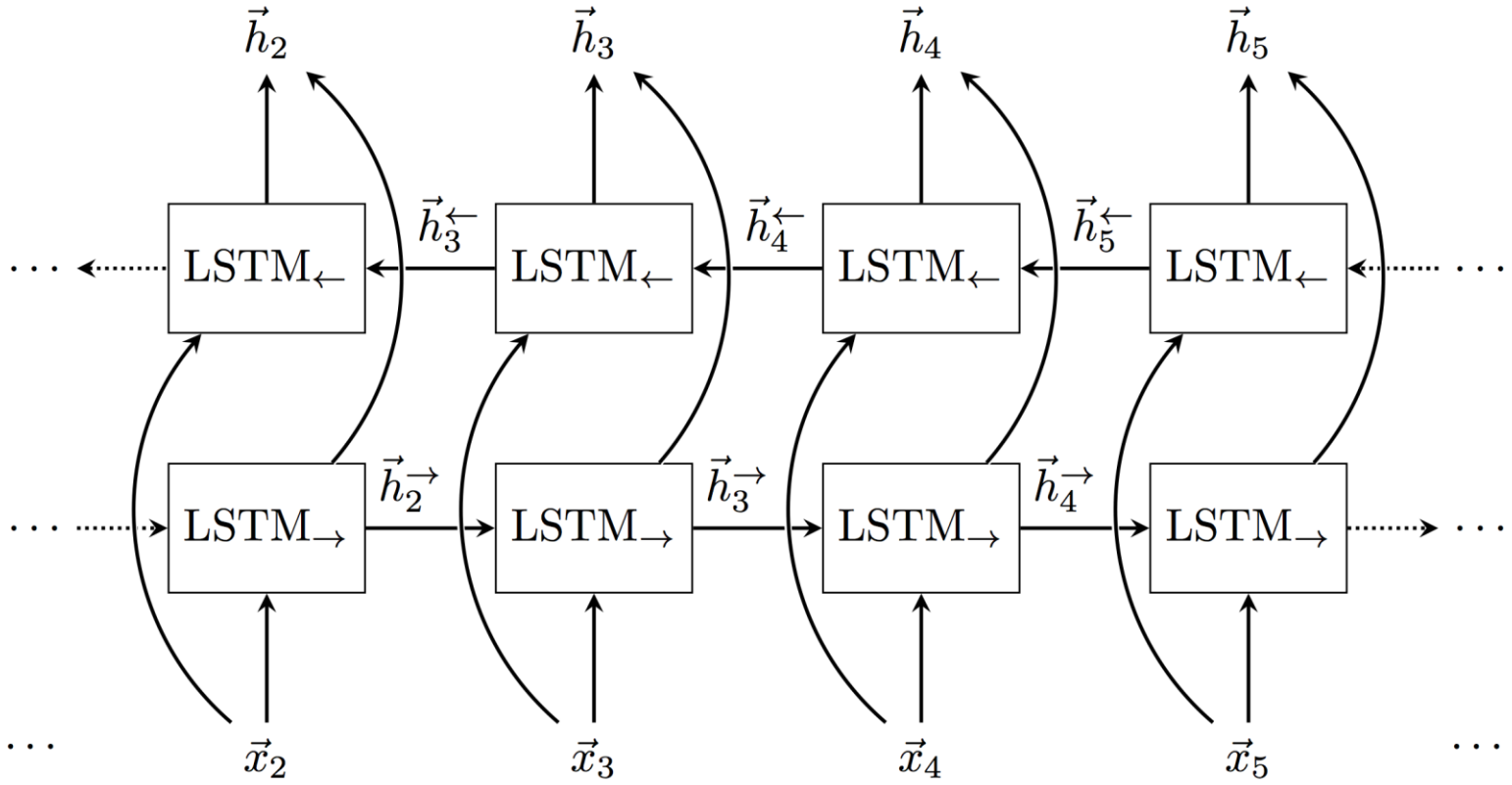
LSTM



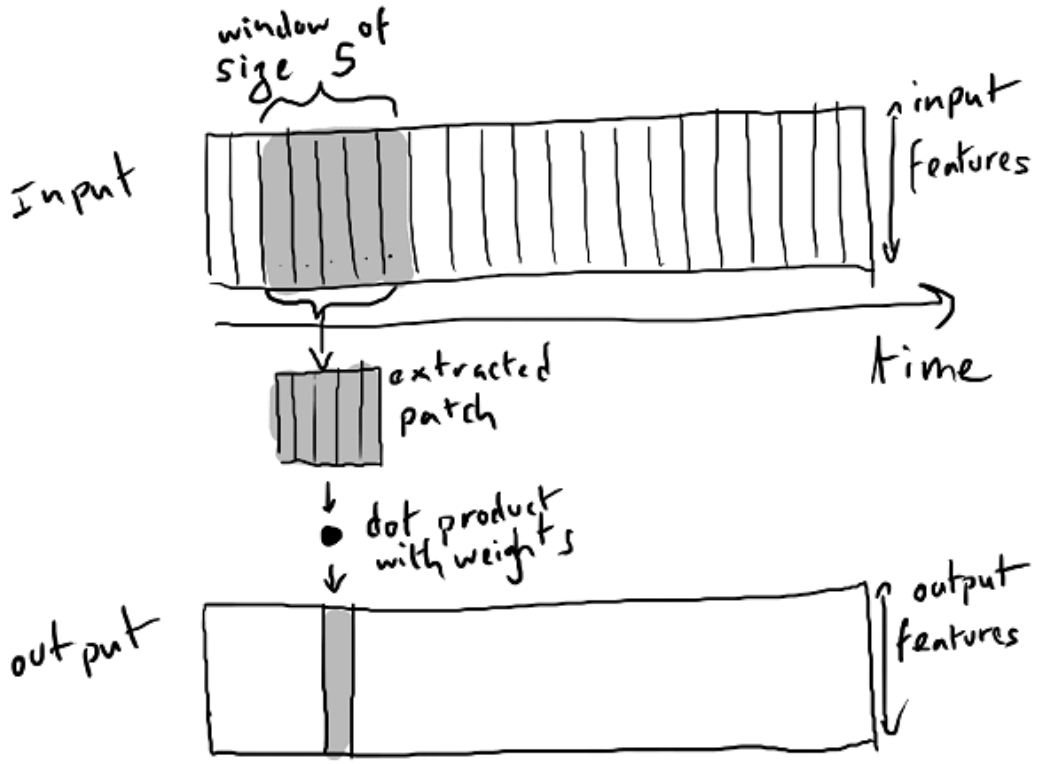
GRU

Microsoft

BI-DIRECTIONAL LSTMS

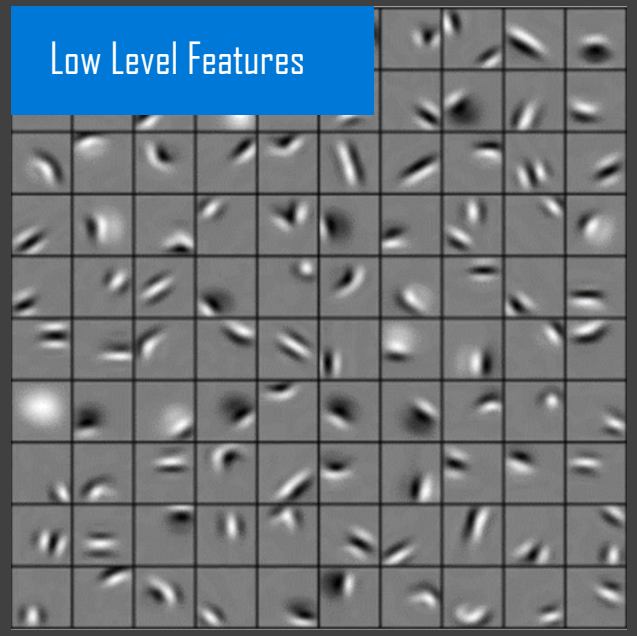


1D-CNNS



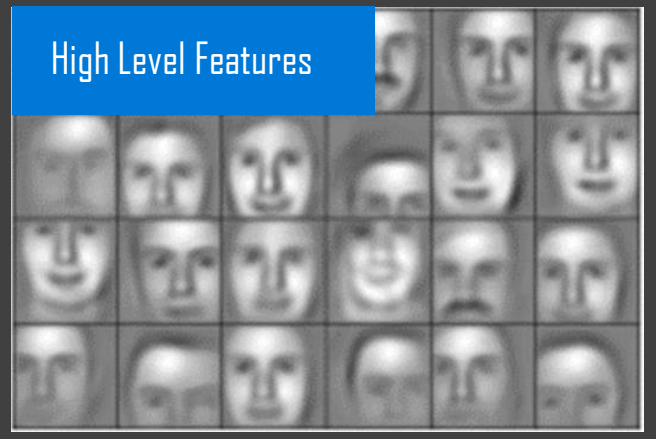
Microsoft

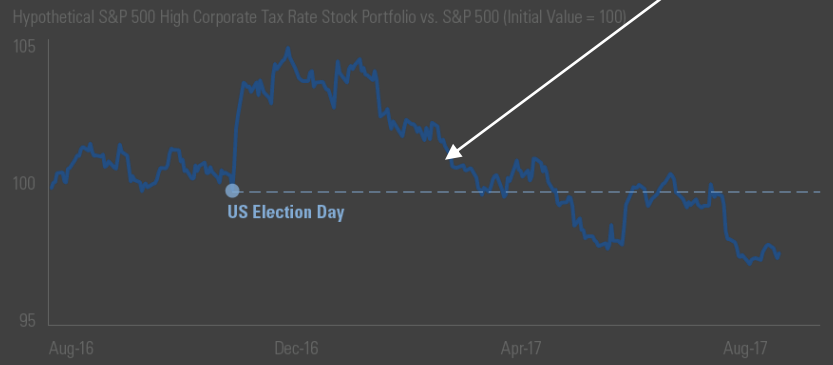
Low Level Features



Mid Level Features

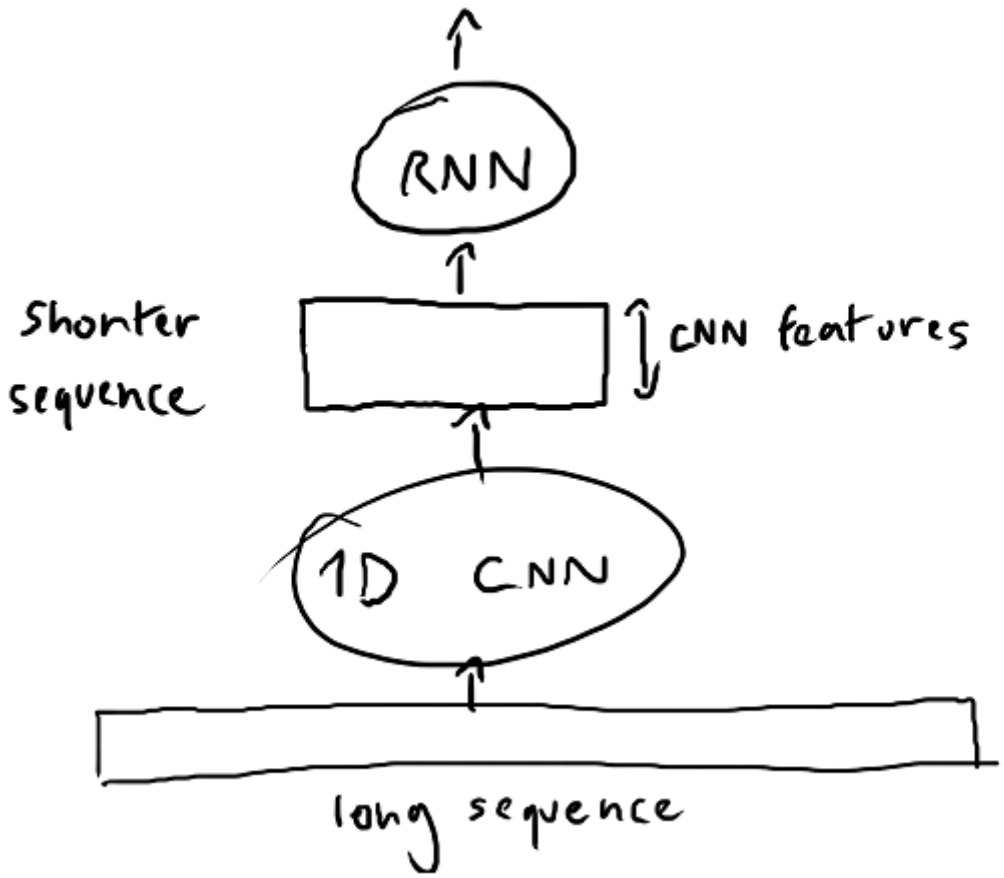
High Level Features





2DCNNS – SAME CONCEPT

Microsoft

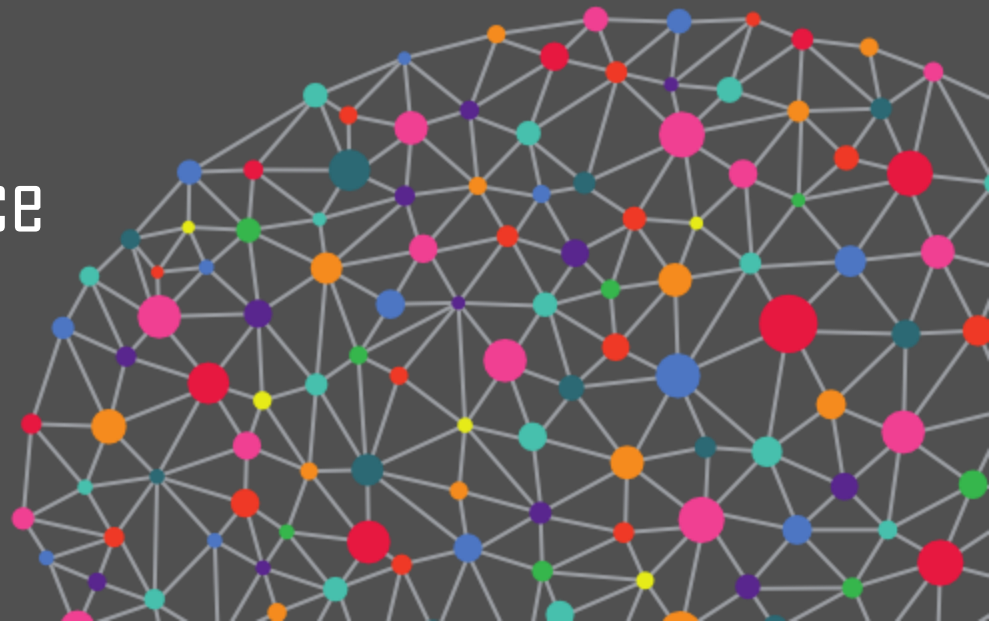


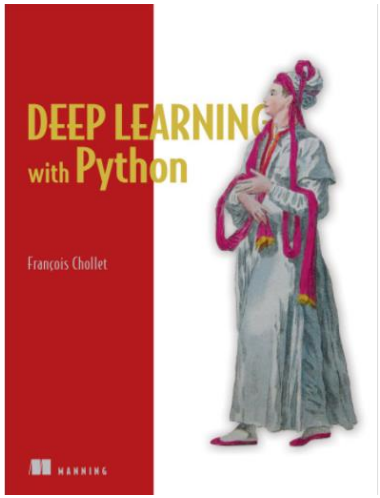
STACKING IS COOL

Microsoft

UNIVERSAL MACHINE LEARNING PROCESS

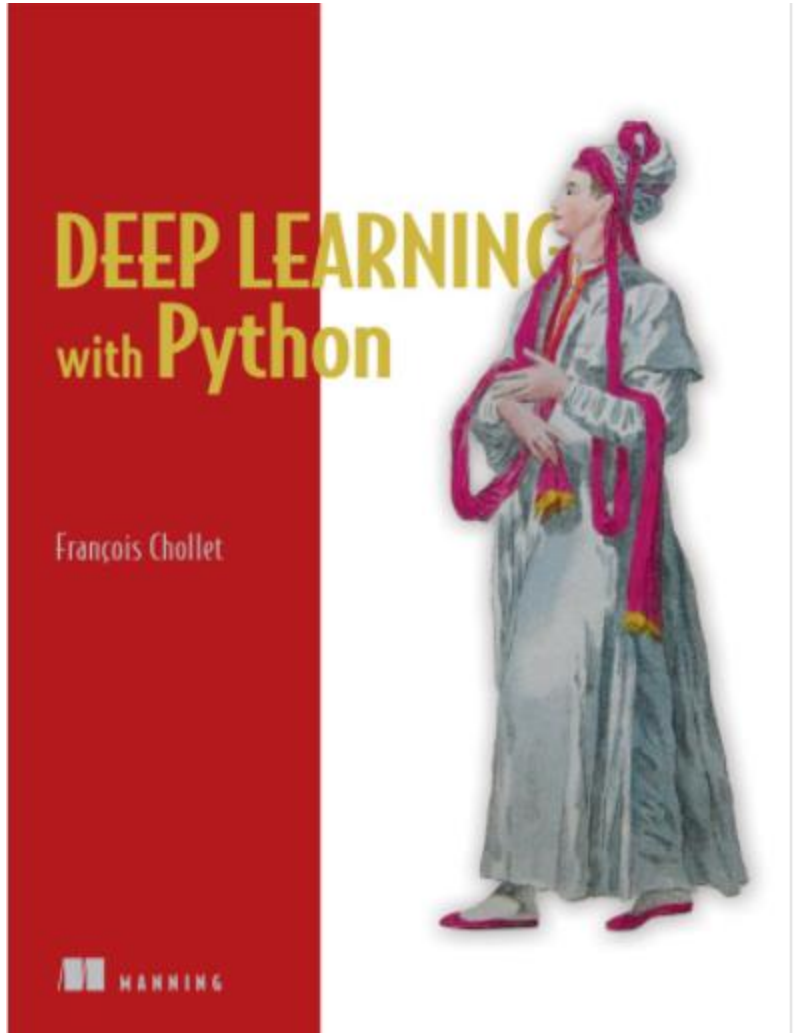
- Define problem
- Define success
- Validation process
- Vectorise/normalize data
- Develop naïve model
- Refine model based on validation performance





DEMOS OF RNNS

Microsoft



https://www.manning.com/books/deep-learning-with-python

Microsoft

THANK YOU!

Azure

Microsoft