Deep Learning for Raiffeisenbank

Jiří Materna

Machine Learning College

@mlcollegecom

@mlcollegecom

#mlcollege

About me

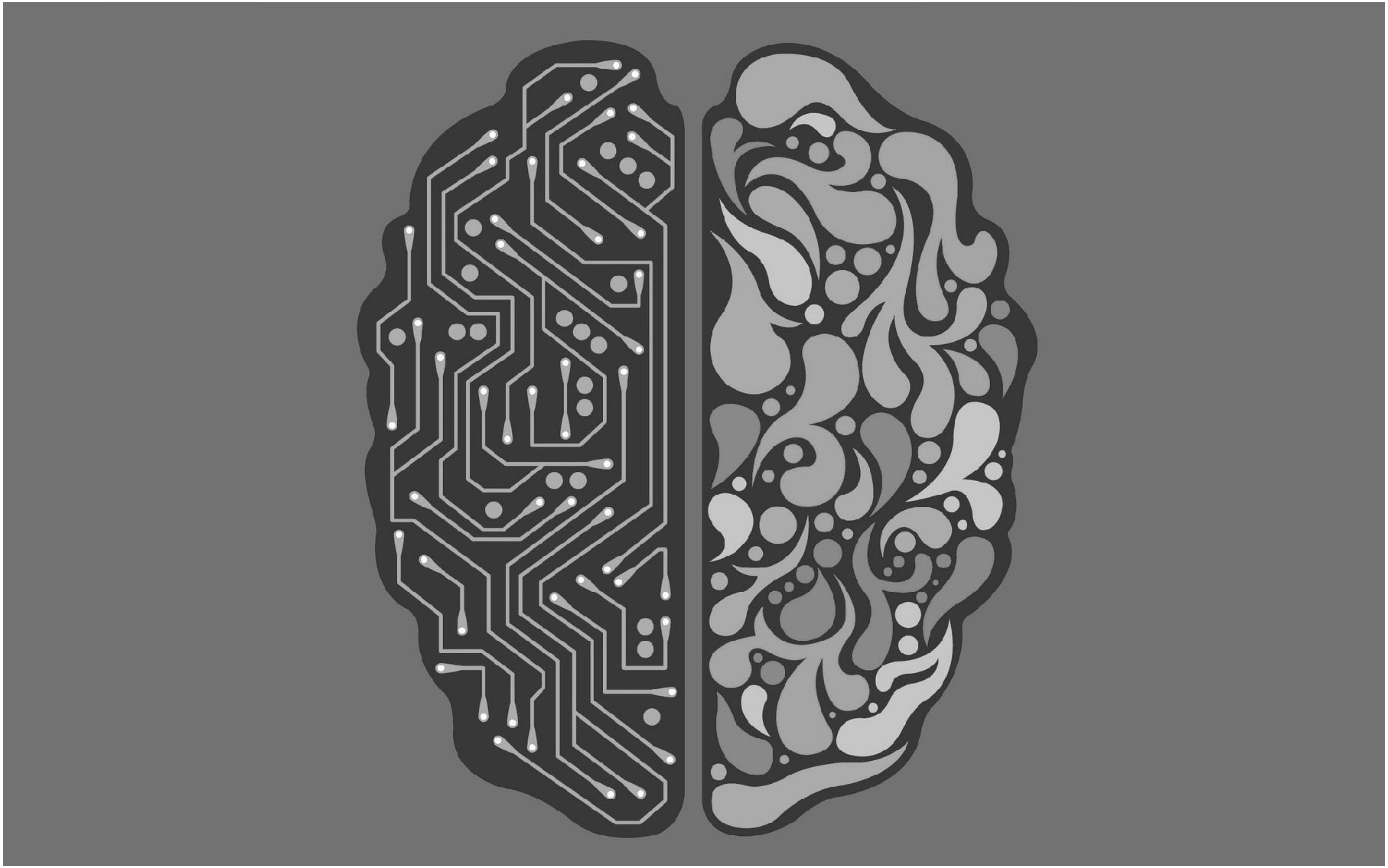
- Ph.D. in Natural Language Processing and Artificial Intelligence at Masaryk University
- 10 years at seznam.cz (last 8 years as Head Of Research)
- Founder and co-organiser of ML Prague
- Author of the ML Guru blog
- Mentor at StartupYard and Startup AI Incubator
- ML Freelacer and consultant

www.mlguru.com **www.mlprague.com** **www.mlcollege.com**

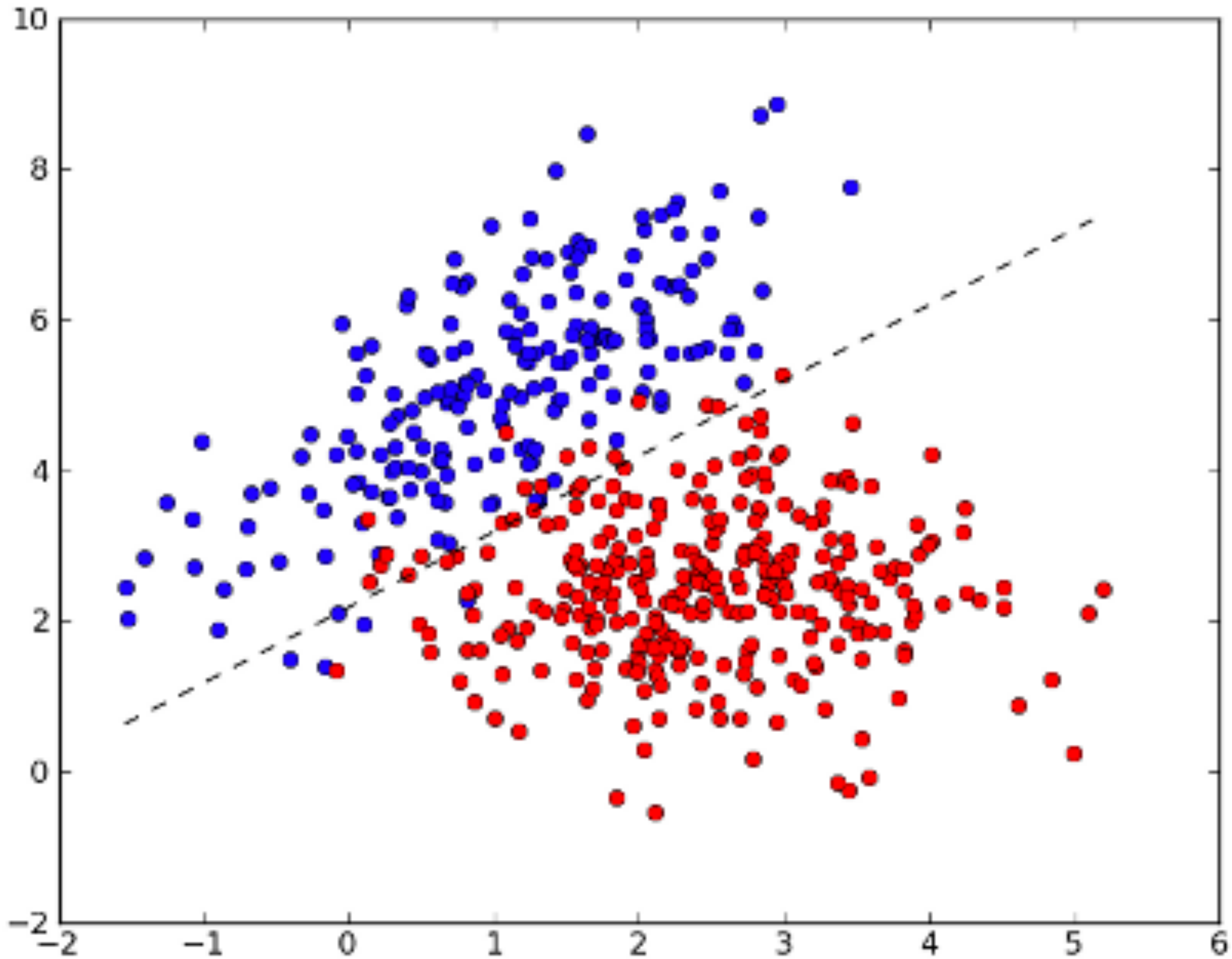
Image processing

- Introduction to neural networks
- Activation functions for neural networks
- Multilayered neural networks
- Methods for training neural networks
- Kears tutorial
- Practical classification and regression tasks solved using neural networks
- ResNet
- Transfer learning and fine-tuning
- Image classification
- Image segmentation
- GANs and superresolution

What is (not) machine learning?

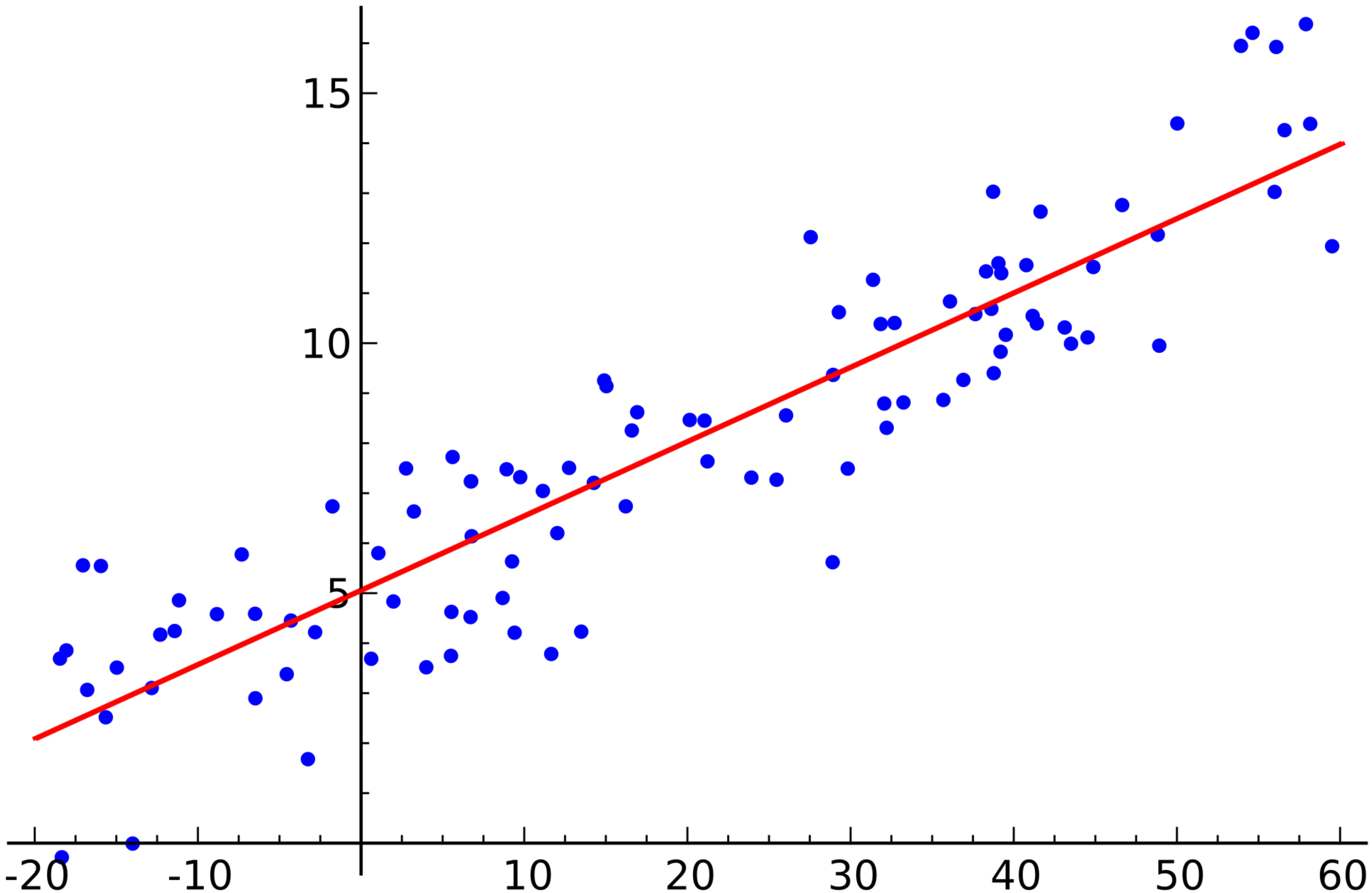


Classification



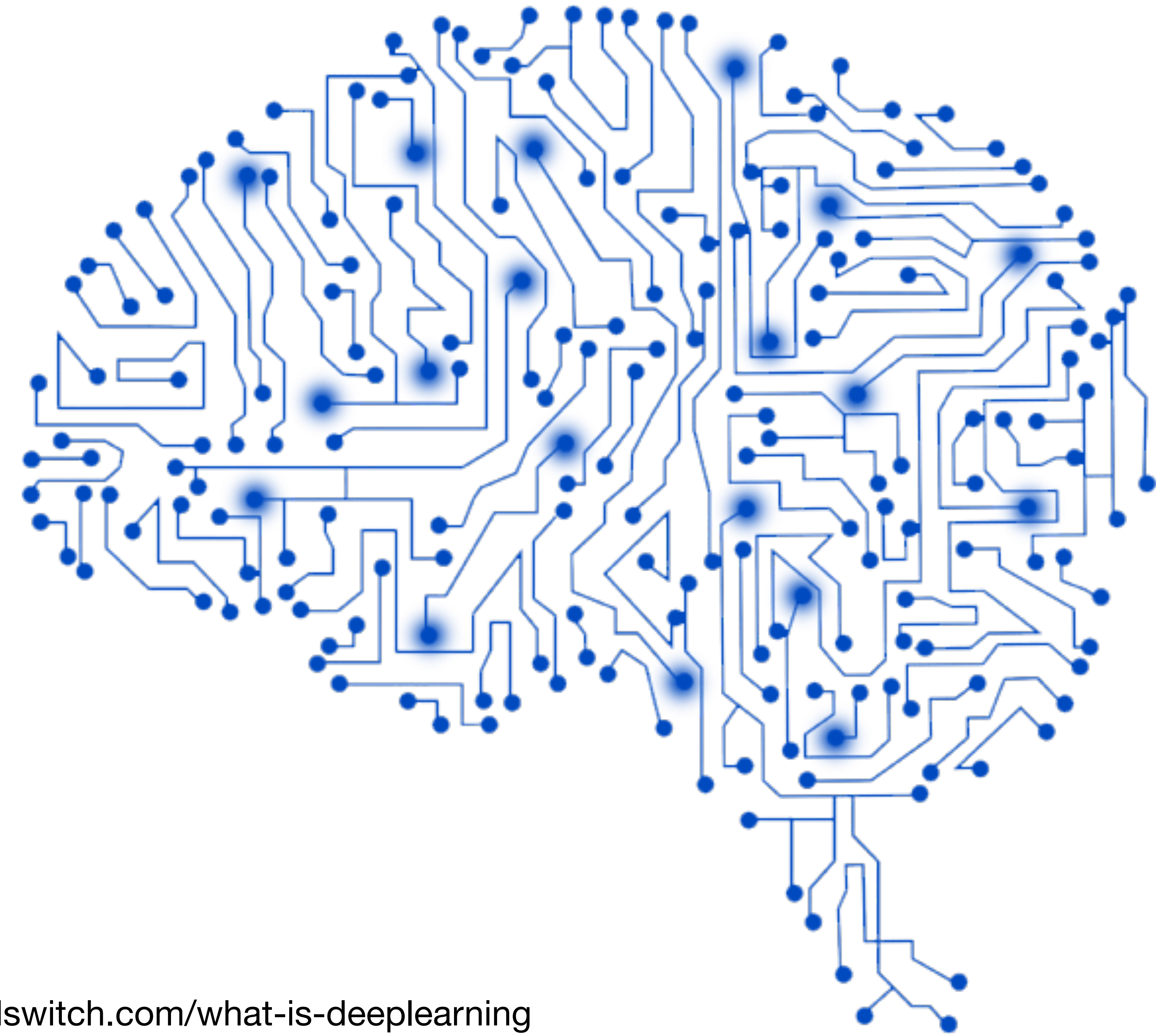
Source: http://mlpy.sourceforge.net

Regression



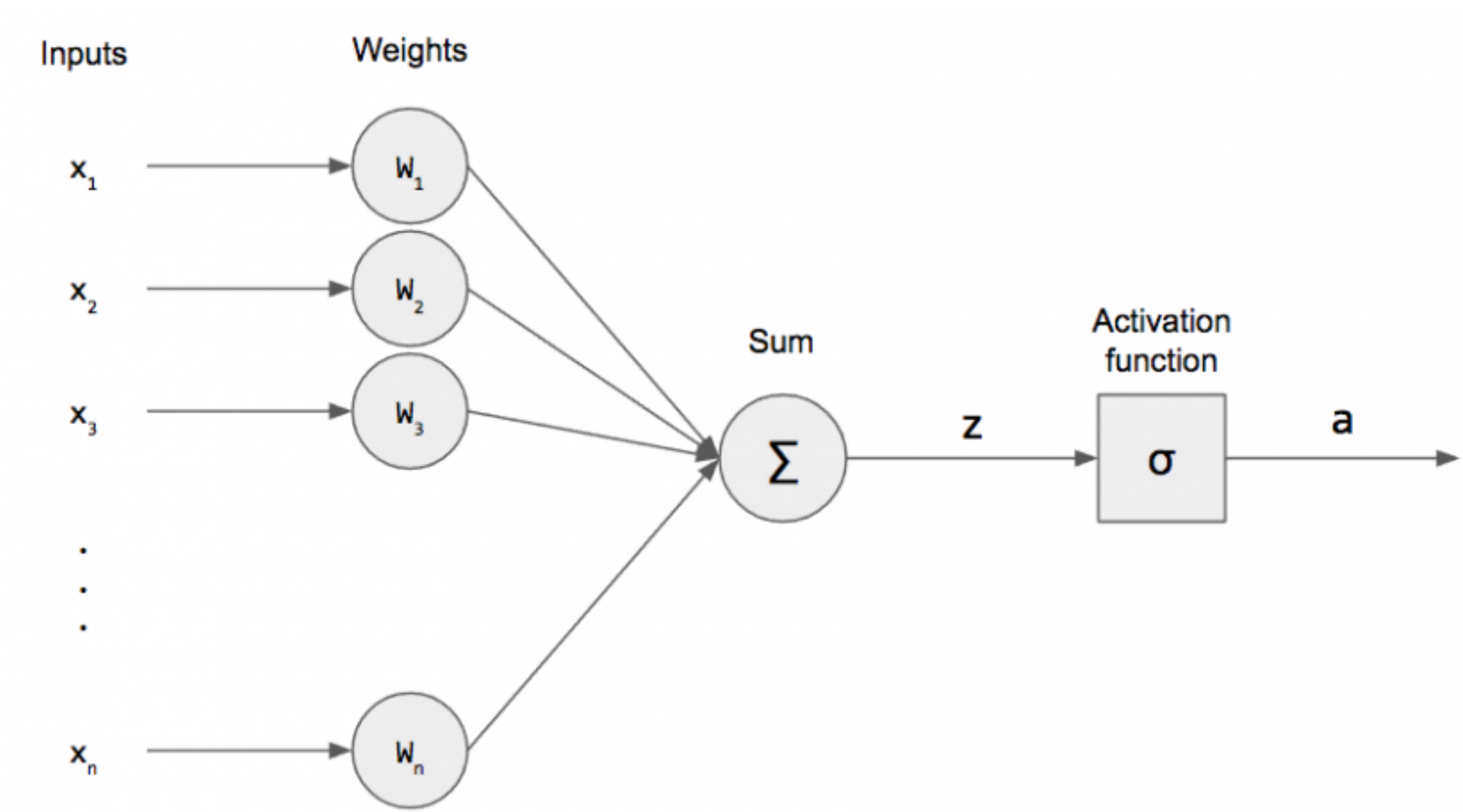
Source: www.wikipedia.org

Neural networks and deep learning



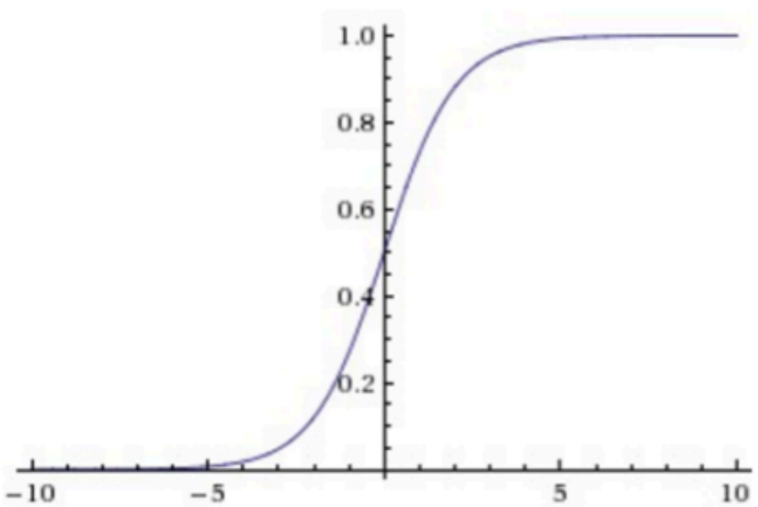
Source: http://www.rapidswitch.com/what-is-deeplearning

Perceptron

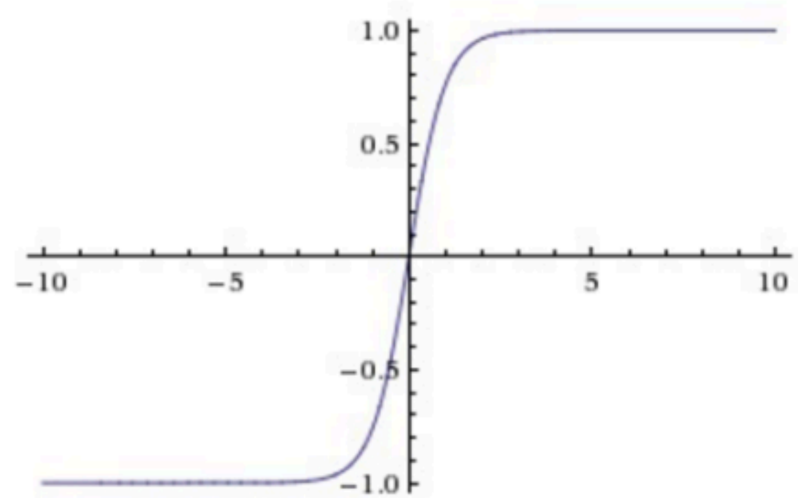


Source: https://pythonmachinelearning.pro

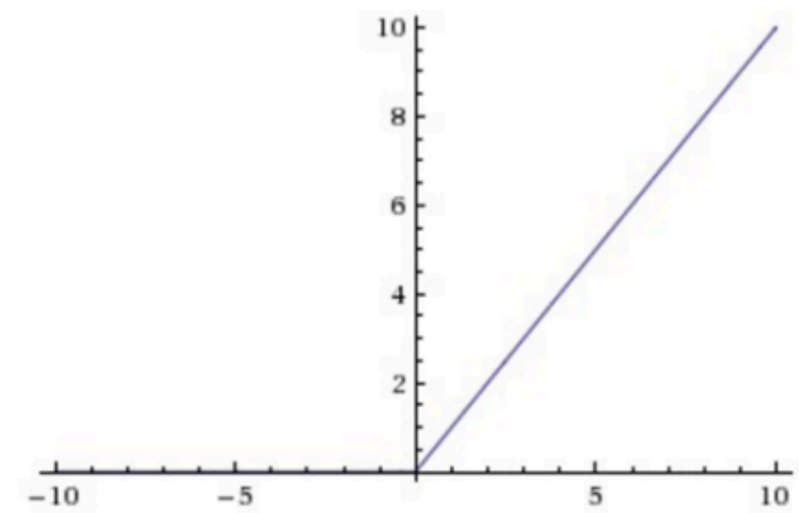
Activation functions



Sigmoid



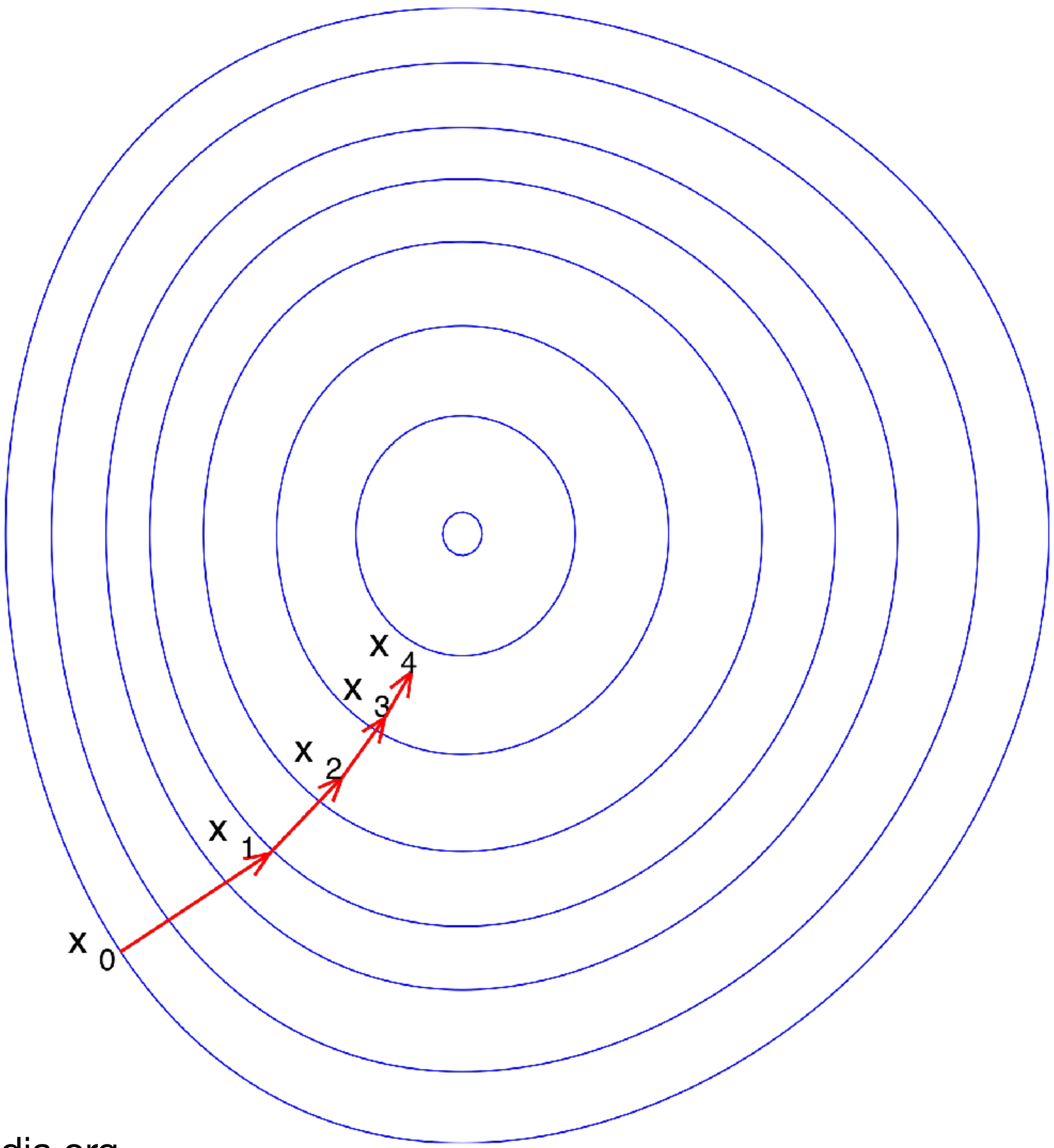
tanh



ReLU

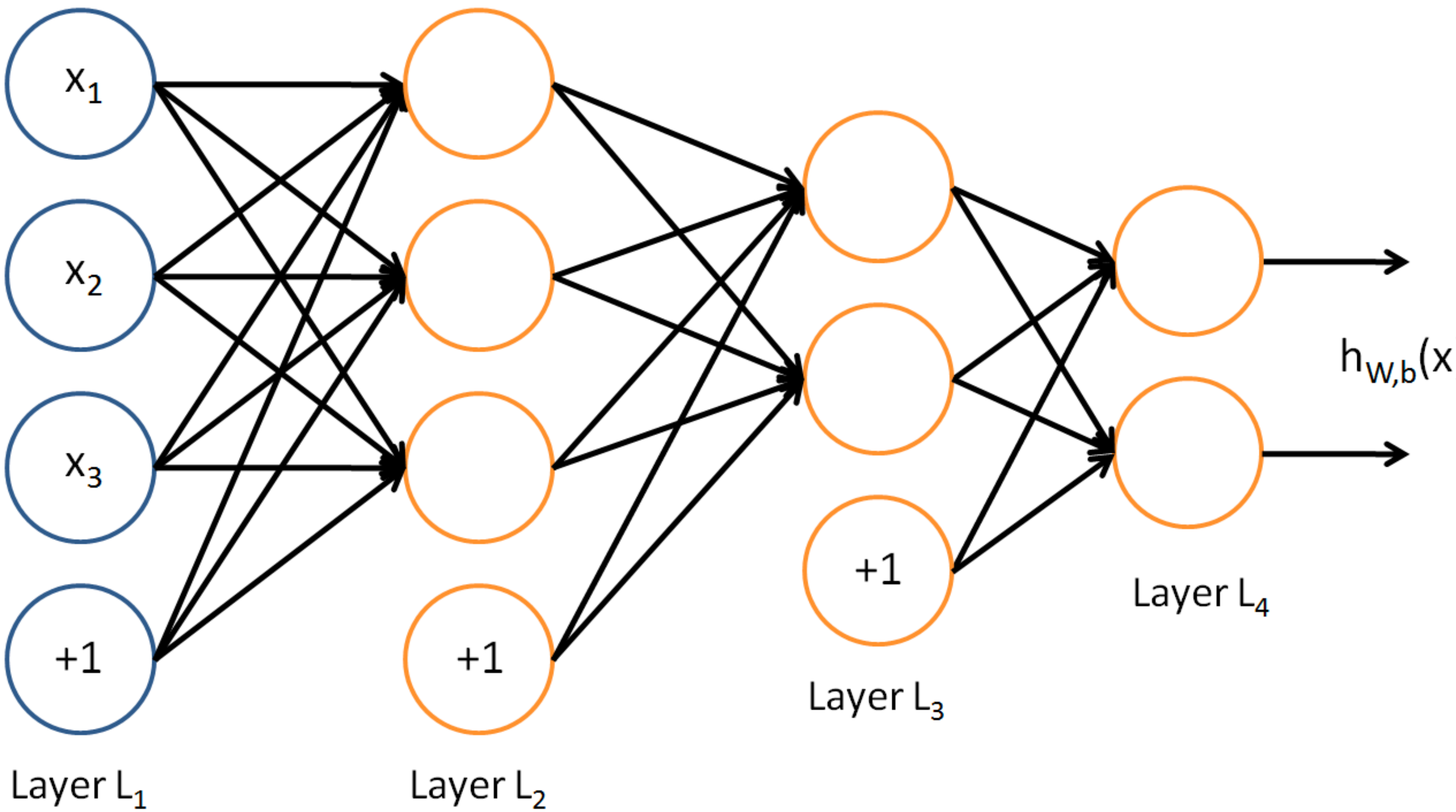
Softmax: $$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$$

Steepest gradient descent



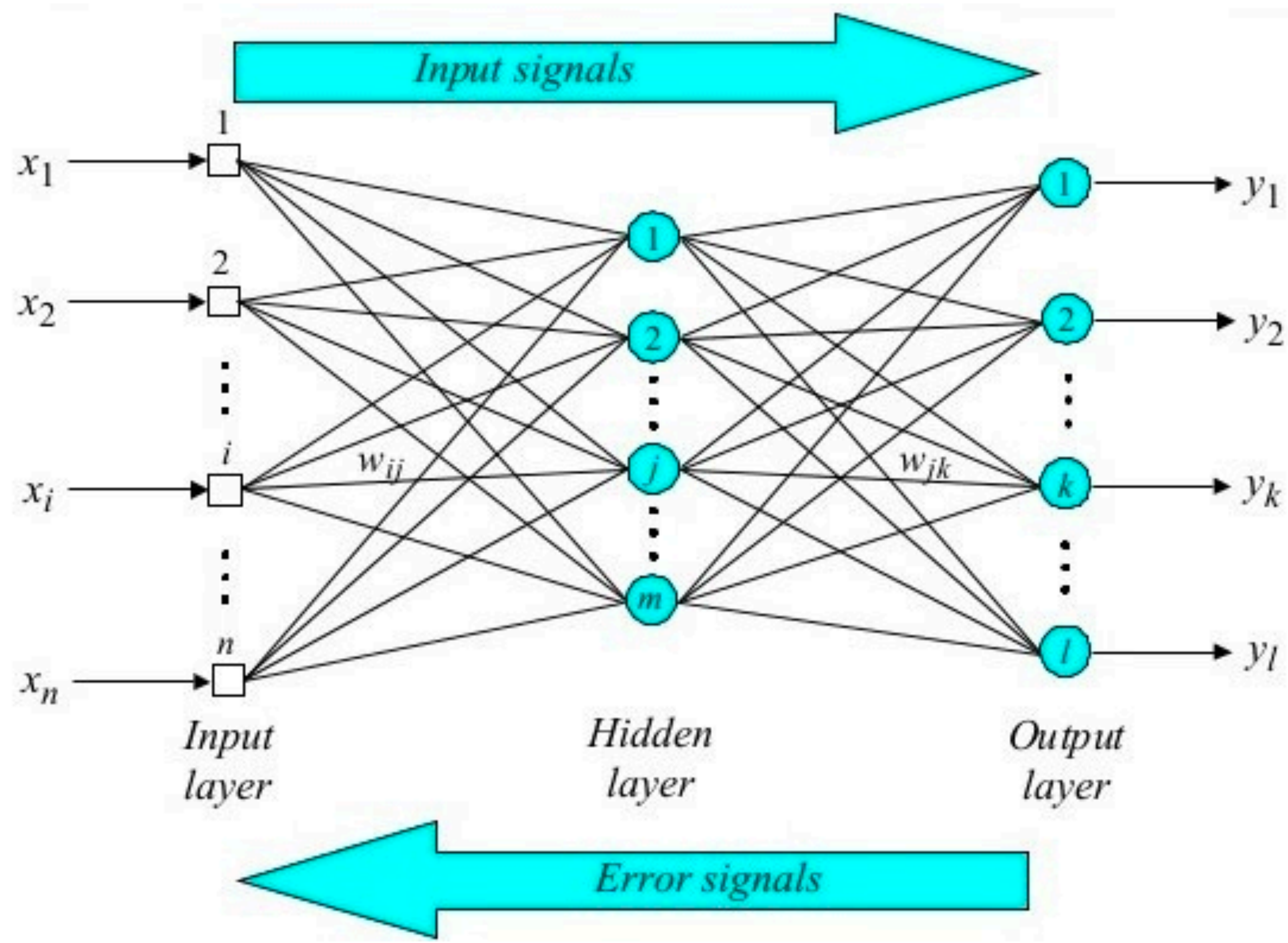
Source: https://www.wikipedia.org

Multilayer Neural Networks



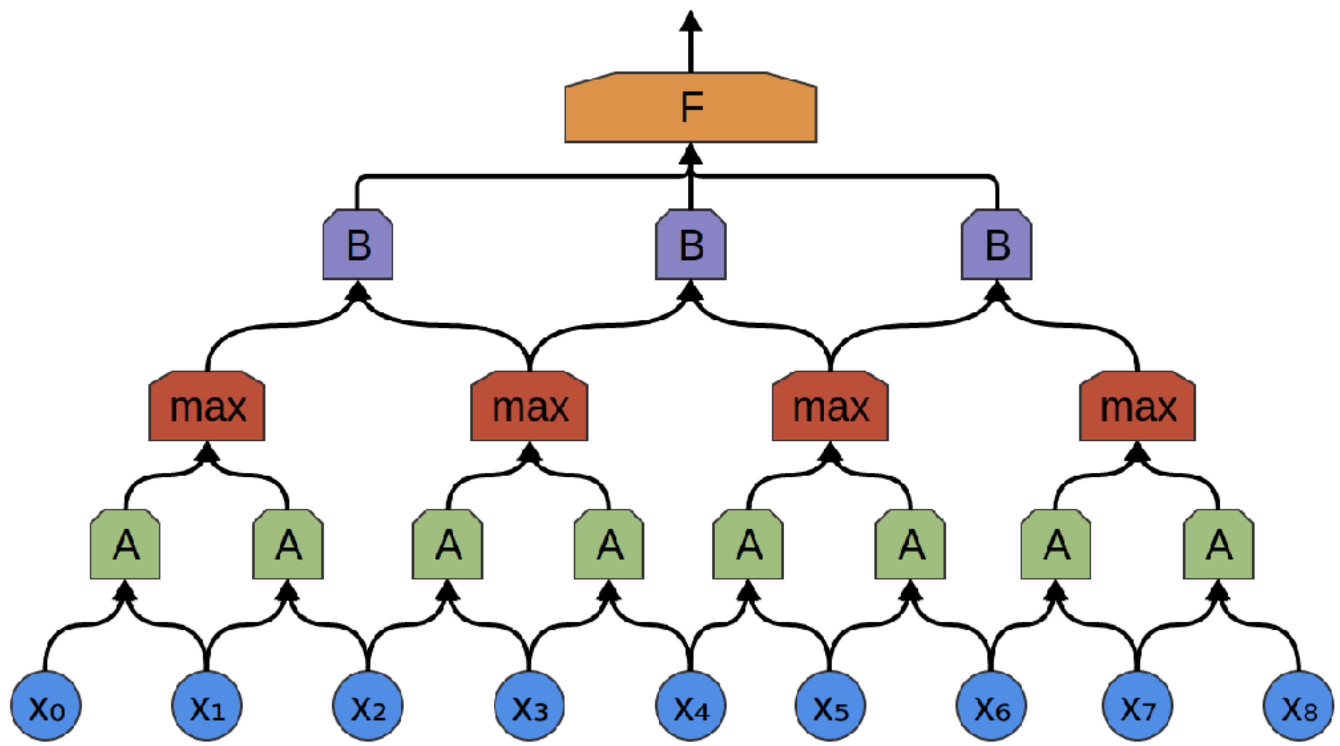
Source: http://ufldl.stanford.edu/

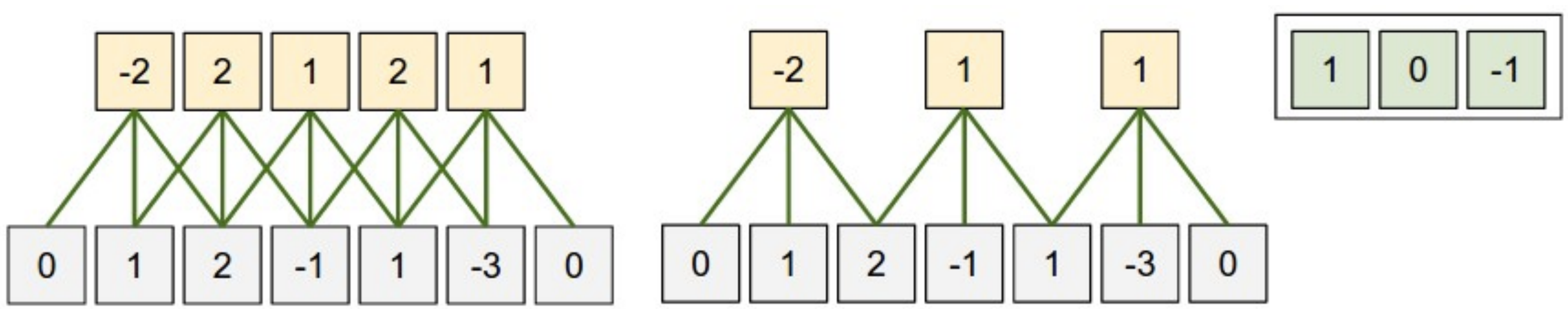
Back propagation



Source: https://www.quora.com/How-do-artificial-neural-networks-work

Convolution





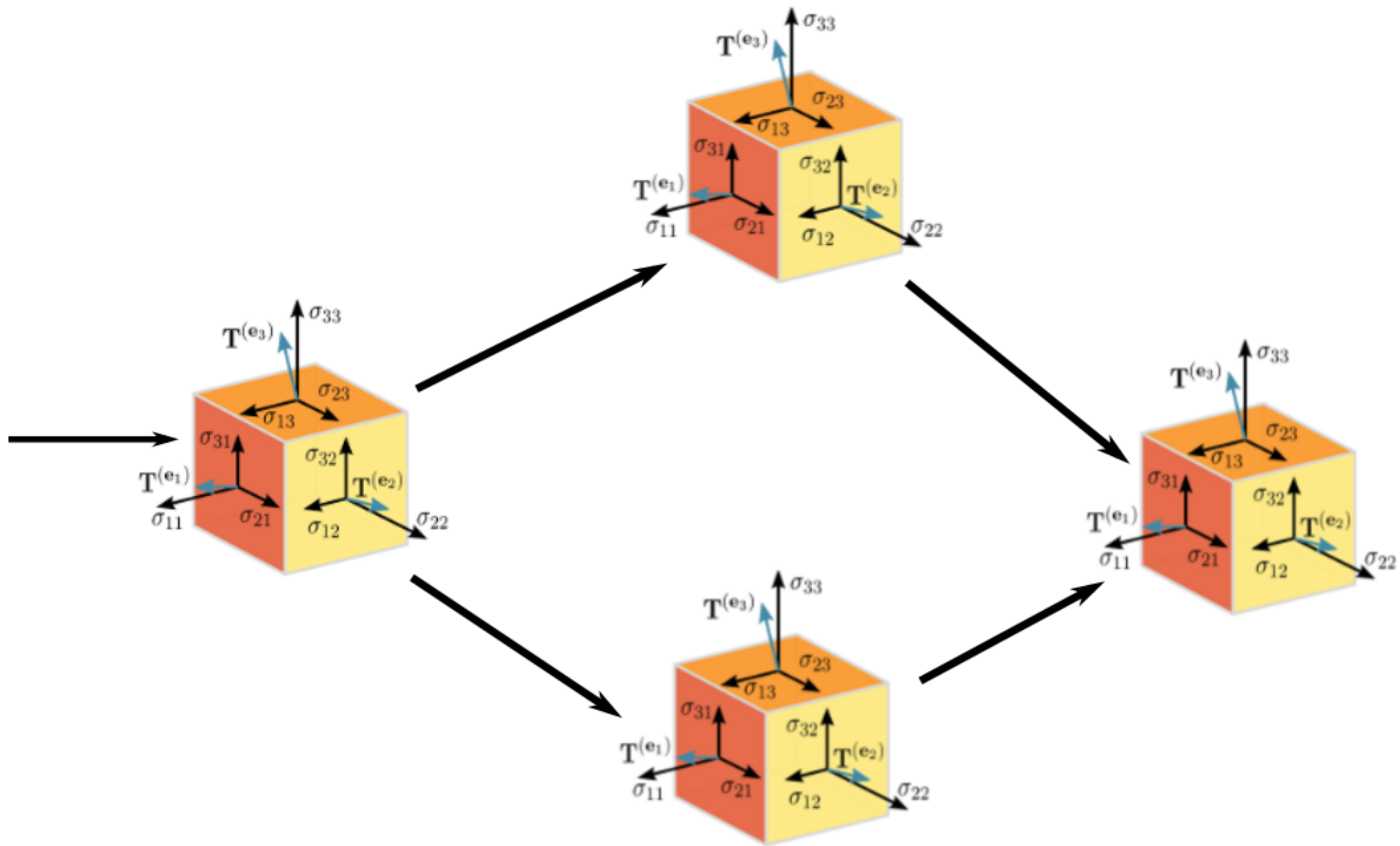
Source: https://www.tensorflow.org

Important terms

- deep learning
- stochastic gradient descent
- batch and minibatch learning
- epoch
- dropout

What is not TensorFlow

What is TensorFlow?



Keras tutorial

01-Keras-introduction.ipynb

Implementation of some classification and regression task using NN

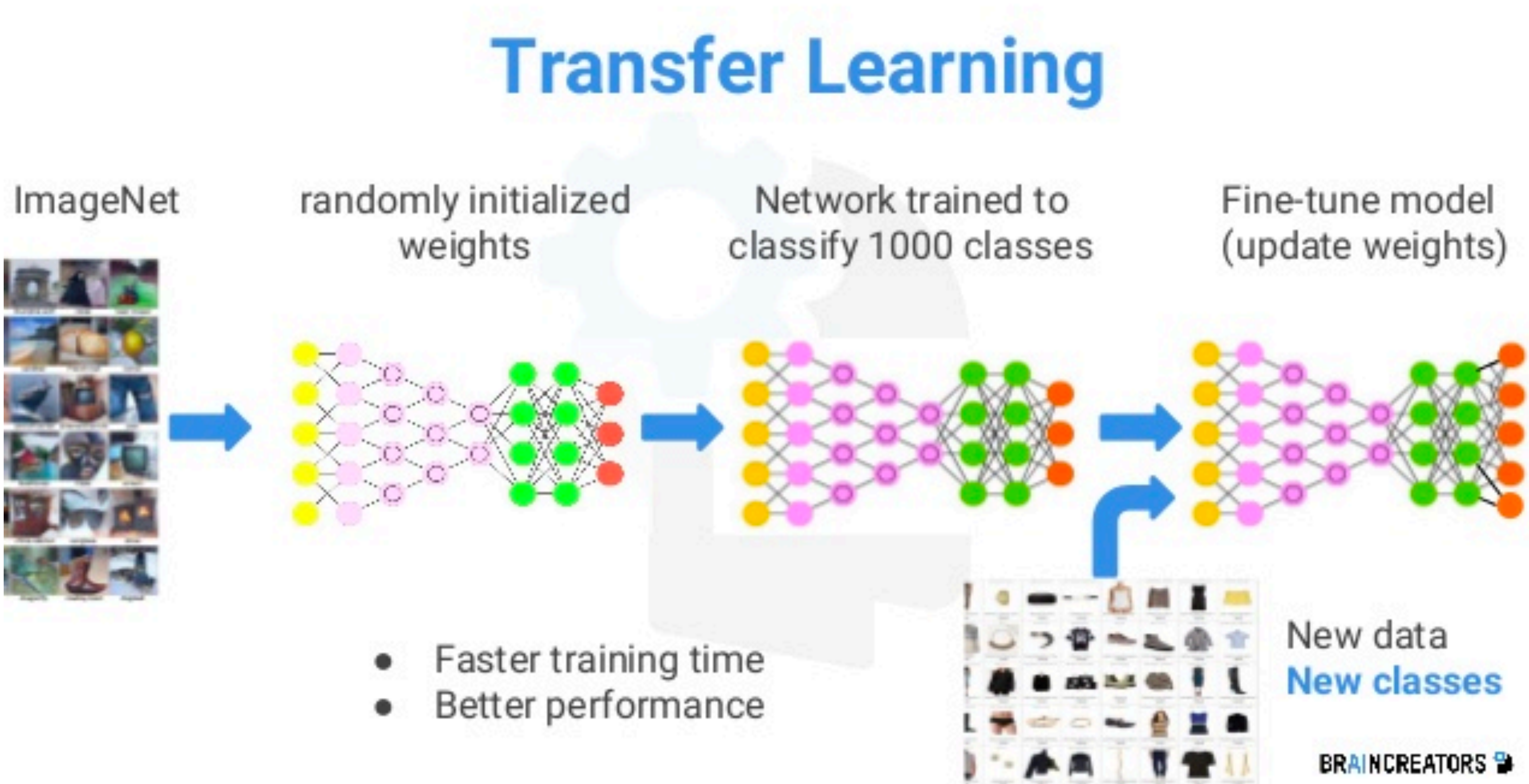
02-Classification-nn-assignment.ipynb

03-Regression-nn-assignment.ipynb

ResNet

residual connection
x
weight layer
$\mathcal{F}(\mathbf{x})$
relu
weight layer
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu
Residual Network
Plain Network
image
7x7 conv, 64, /2
pool, /2
3x3 conv, 64
3x3 conv, 128, /2
3x3 conv, 128
3x3 conv, 256, /2
3x3 conv, 256
3x3 conv, 512, /2
3x3 conv, 512
avg pool
fc 1000

Finetuning


Transfer Learning
ImageNet
randomly initialized weights
Network trained to classify 1000 classes
Fine-tune model (update weights)
Faster training time
Better performance
New data
New classes
BRAINCREATORS

Transfer learning example

04-Transfer_learning.ipynb

Adversarial Patch

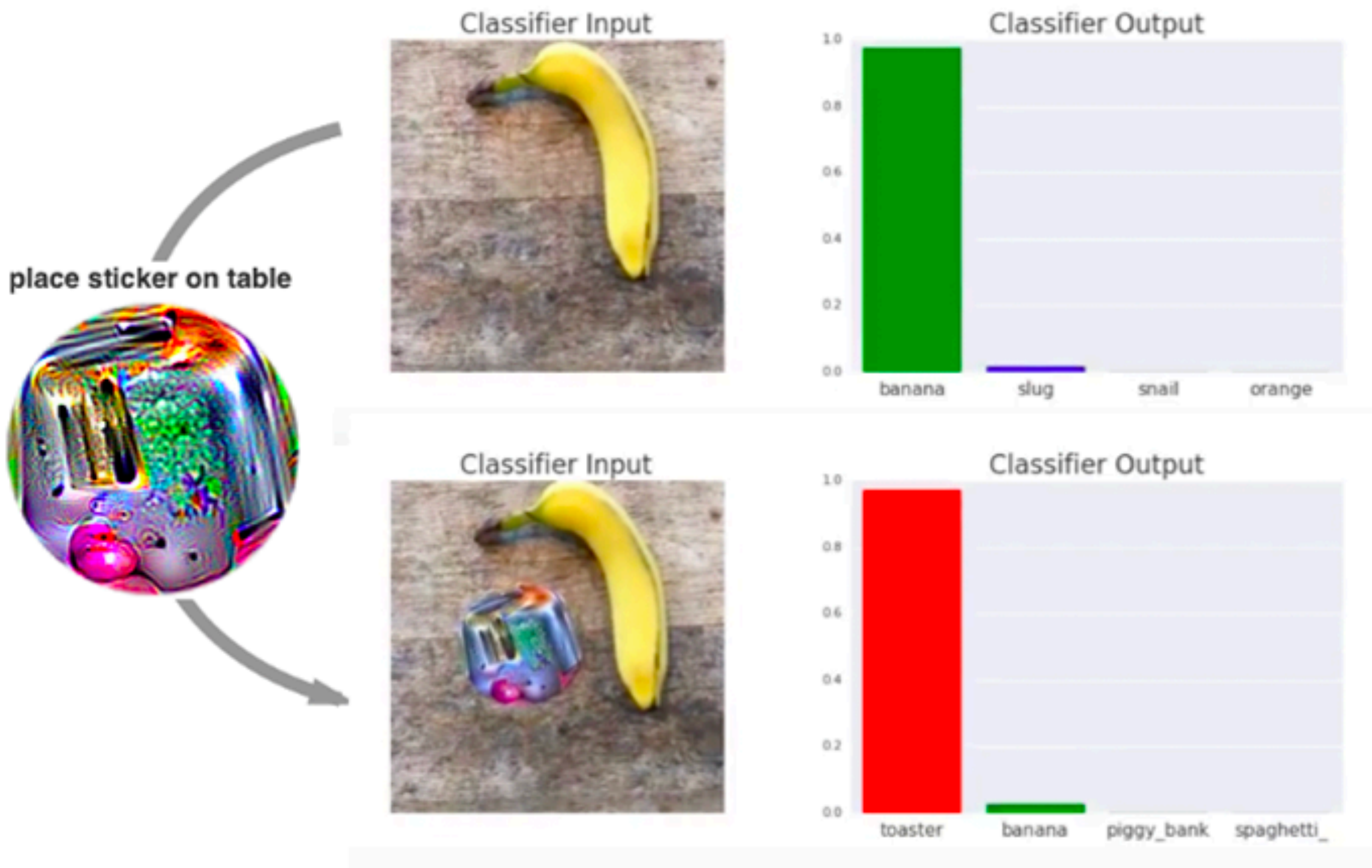
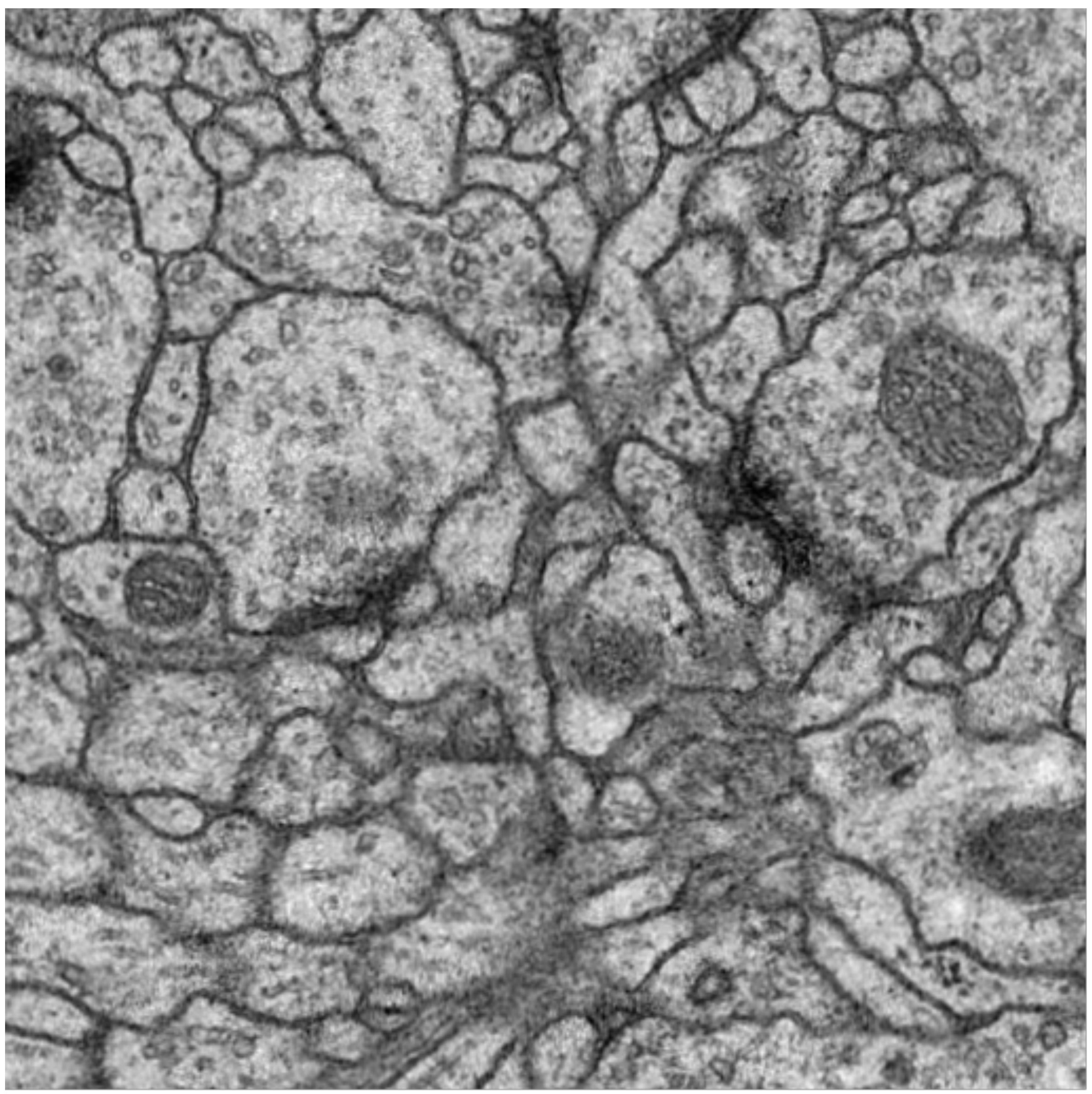
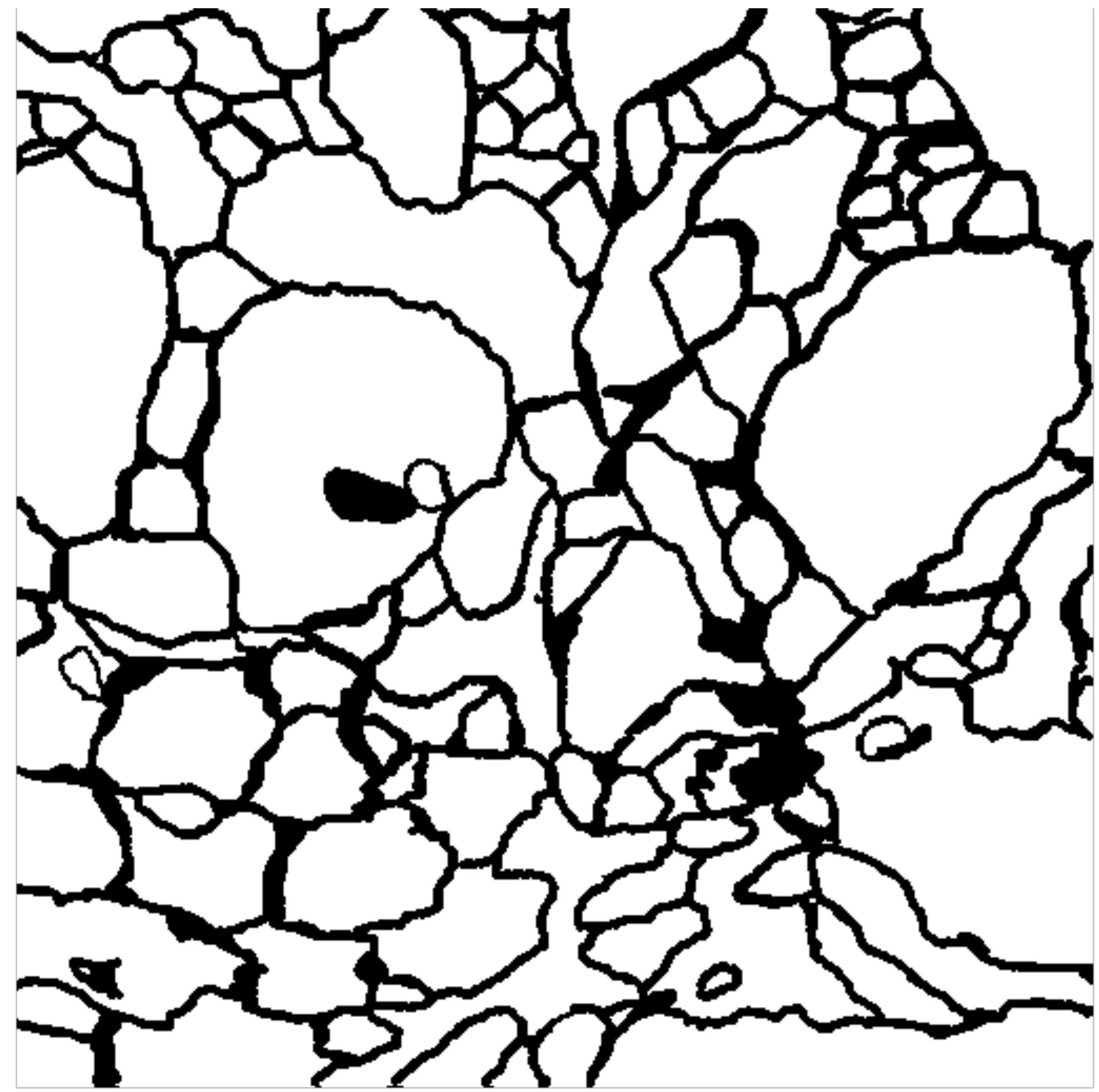


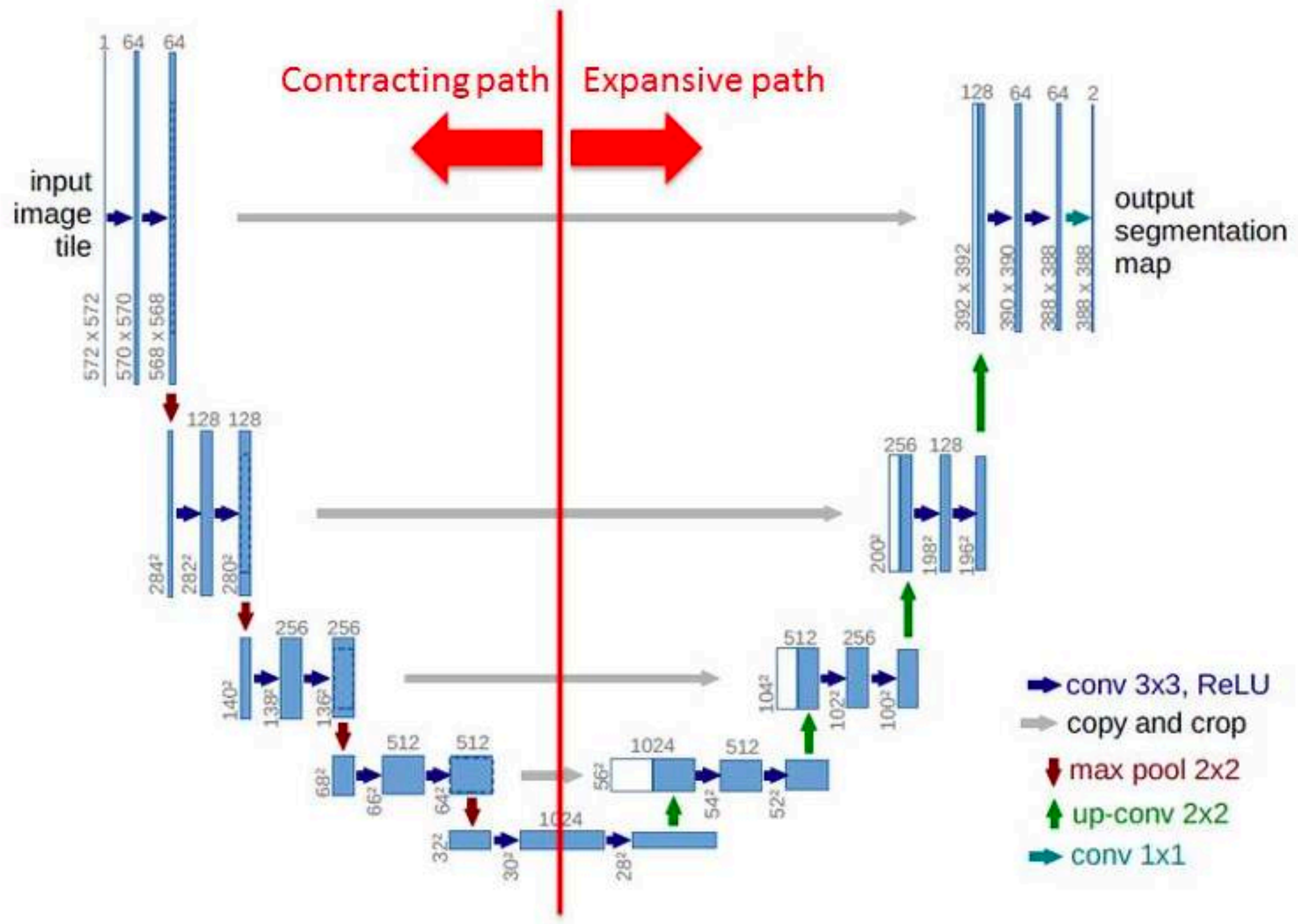
Image segmentation





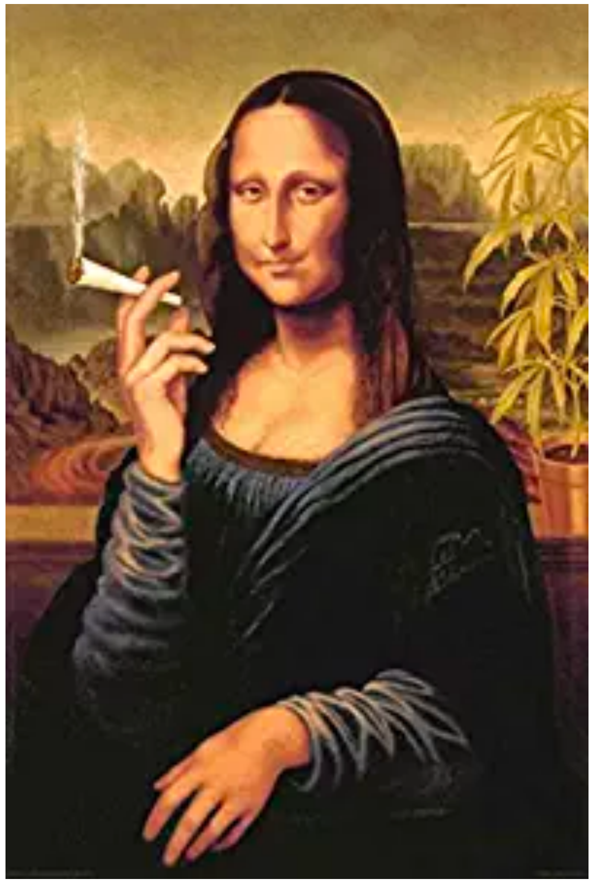
U-Net

Network Architecture

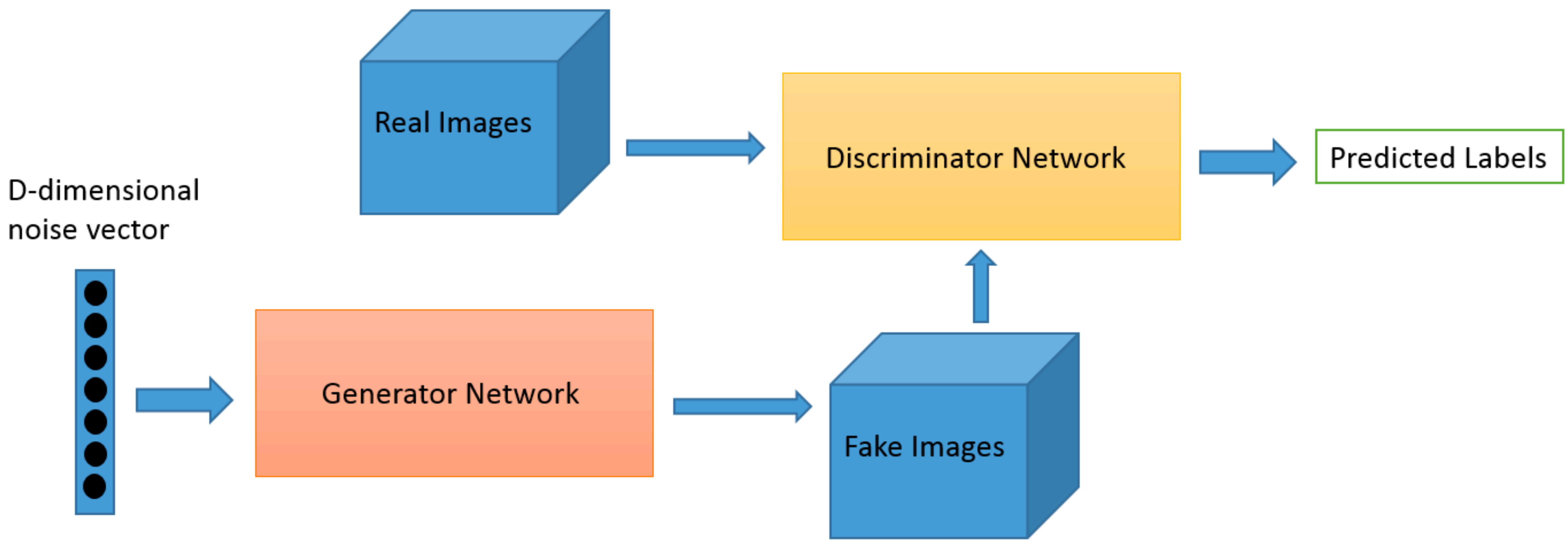


Generative models with neural networks

Random noice → **Artificial Neural network** →



Generative Adversarial Networks



Superresolution

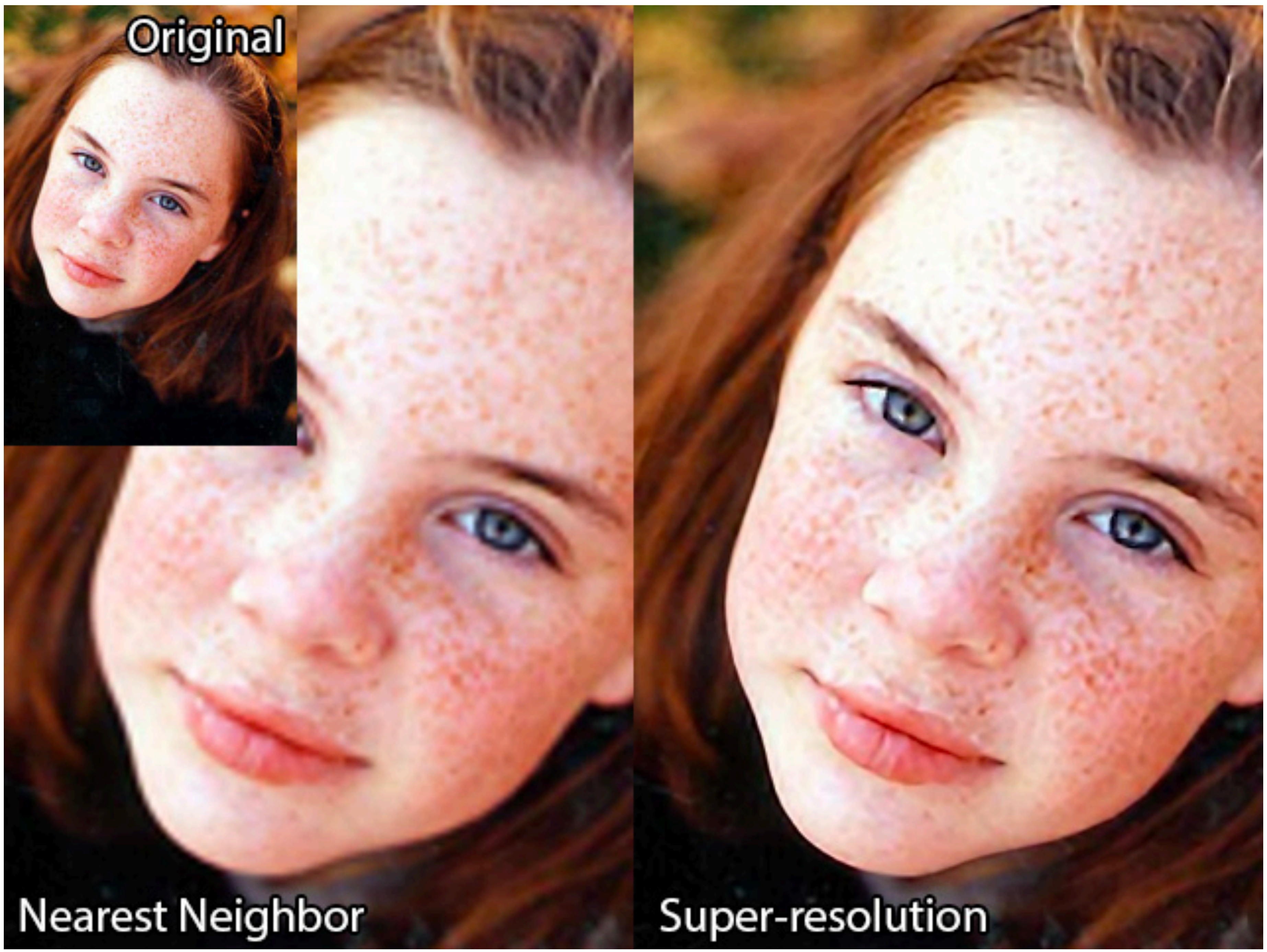


Image synthesis

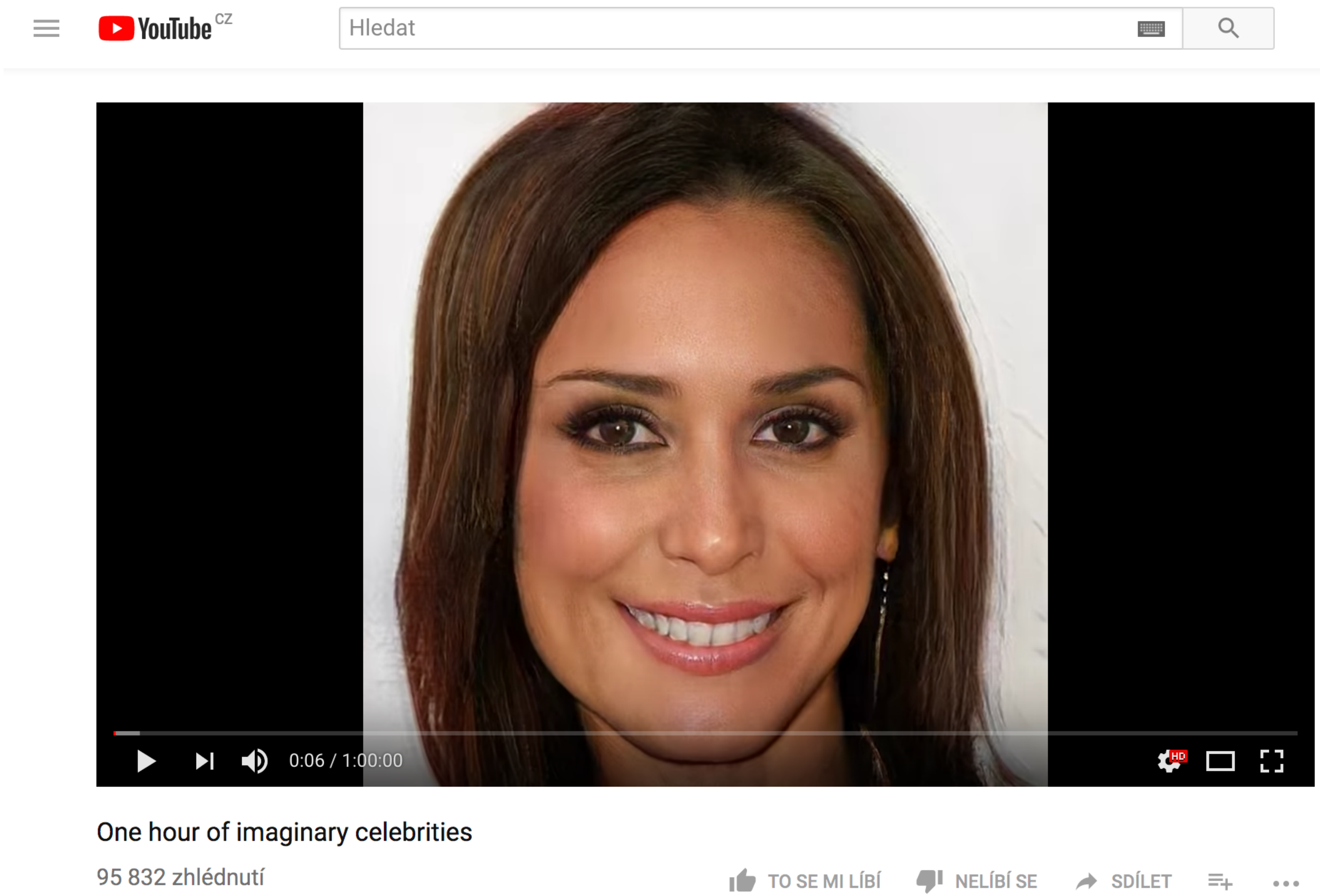
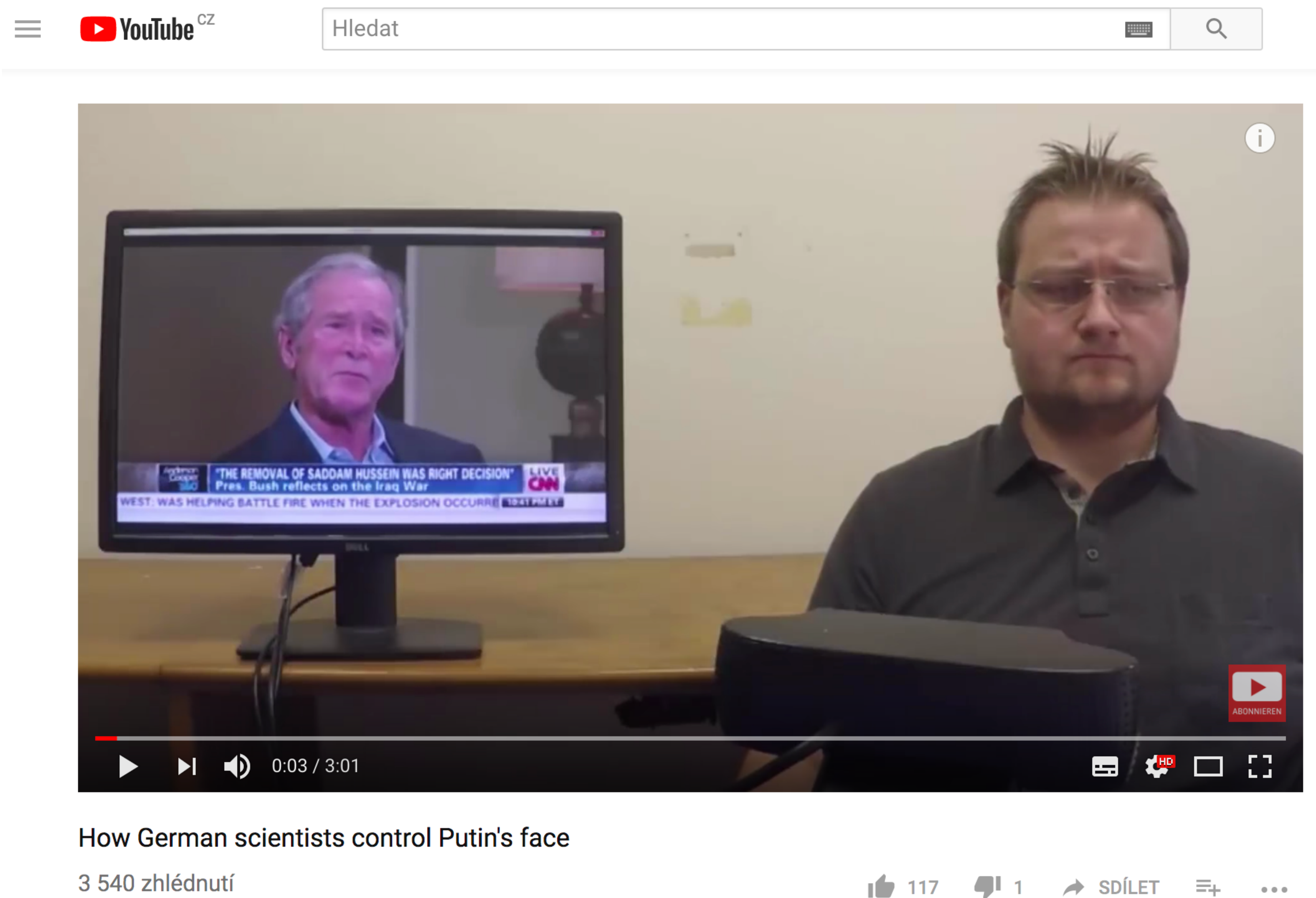


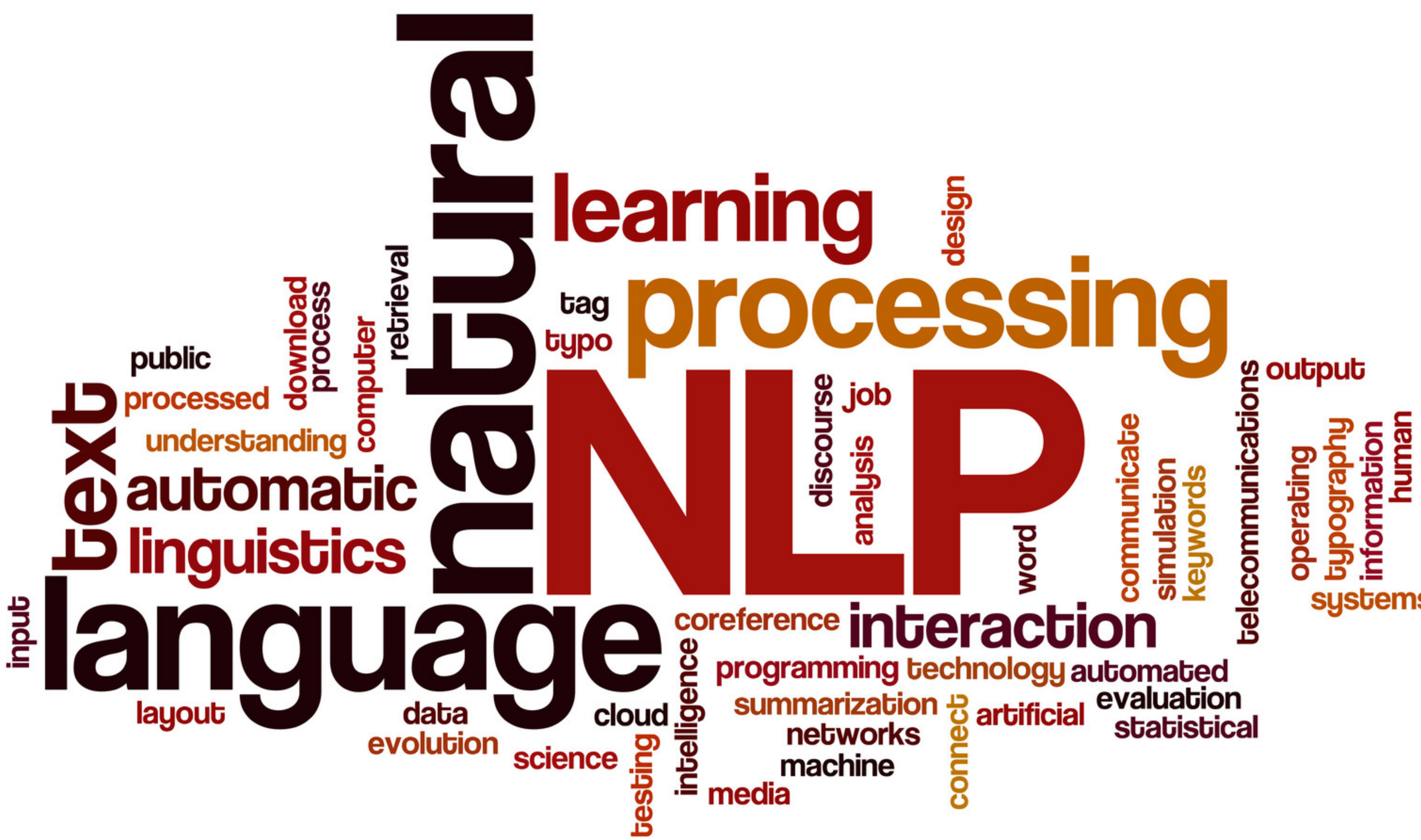
Image manipulation



Generative Adversarial Networks

06_GANs.ipynb

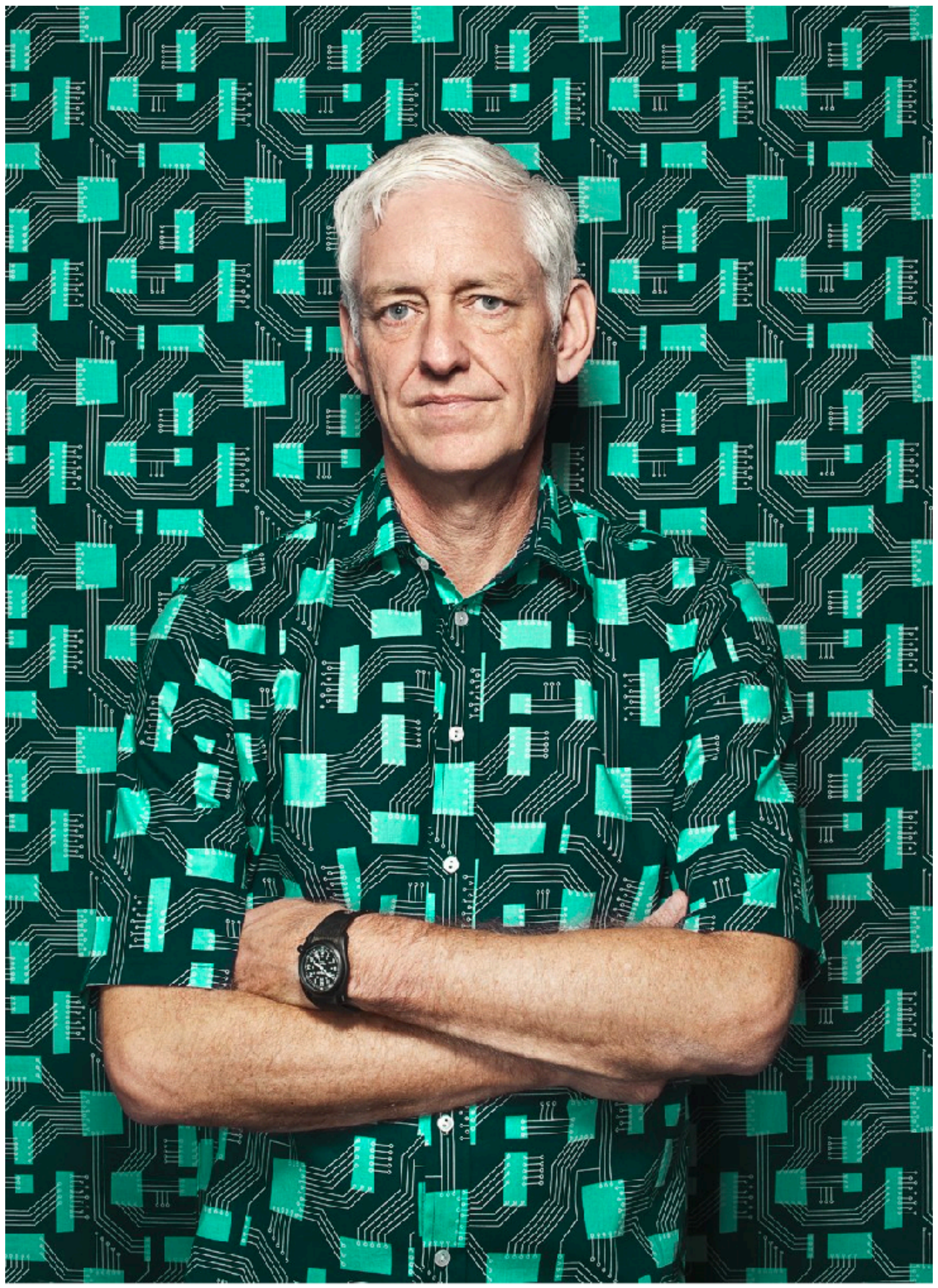
Natural Language Processing



Outline

- Introduction to natural language processing
- Computational linguistics
- Text document vectorization
- Practical document classification task
- Language modeling
- Practical tasks on language modeling
- Word embeddings
- Text generating
- Practical tasks on language modeling

Norvig vs. Chomsky



source: https://www.commarts.com

source: https://citaty.net

Text corpus

Sketch Engine | natural language | British National Corpus

Home
Search
Word list
Word sketch
Thesaurus
Sketch diff
Trends
Corpus info
My jobs
User guide

Save
Make subcorpus
View options
- KWIC
- Sentence

Sort
- Left
- Right
- Node
- References
- Shuffle

Sample

Query natural, language 255 (2.27 per million)

Page 1 of 13 Go Next | Last

J2K	nature of deixis (see Chapter 2 below) in	natural languages	, for sentences like (II) are true or false only
J2K	of the simple but immensely important fact that	natural languages	are primarily designed, so to speak, for use in
J2K	. The many facets of deixis are so pervasive in	natural languages	, and so deeply grammaticalized, that it is hard
J2K	the utterance, within the utterance itself.	Natural language	utterances are thus" anchored" directly to
J2K	semantics deals with certain	natural language	expressions. Suppose we identify the semantic
J2K	or self-referring expressions in	natural languages	, as in (12) and, arguably, in (13) (see Chapter 5
J2K	, is perhaps a philosophical red-herring.	Natural languages	, after all, just do have indexicals, and it is
J2K	. Semantics is then not concerned directly with	natural language	at all, but only with the abstract entities
J2K	to leave us with no term for all those aspects of	natural language	significance that are not in any way amenable to
J2K	of the deictic expressions that occur in	natural languages	, and we should now turn to consider linguistic
J2K	in familiar languages. Deictic systems in	natural languages	are not arbitrarily organized around the
J2K	. But this has the consequence, as we noted, that	natural languages	will only have a syntax and a pragmatics, and no
J2K	more or less directly on fragments of	natural language	(as initiated by Montague, 1974) would make
J2K	The semanticist who takes the other tack, that	natural language	senses are protean, sloppy and variable, is
J2K	offers a way out, for it allows one to claim that	natural language	expressions do tend to have simple, stable and
J2K	radical differences between logic and	natural language	seem to fade away. We shall explore this below
J2K	on what can be a possible lexical item in	natural languages	. Finally, the principles that generate
J0V	is meant any single document, or any stretch of	natural language	regarded as a self-contained unit for
J53	recognition and those that can understand	natural languages	, such as English, are known by the collective
HRK	through a dialogue, which approaches a	natural language	dialogue, or via a menu. In figure 6.2, the users

Page 1 of 13 Go Next | Last

Token & tokenization

This is a non-trivial English sentence: Ludolph's number is approx. 3.14.

Python library: http://www.nltk.org/

Stemming & lemmatization

Original	Stemming	Lemmatization
compensation	compens	compensation
compensations	compens	compensation
mouse	mous	mouse
mice	mice	mouse

Stemming & lemmatization

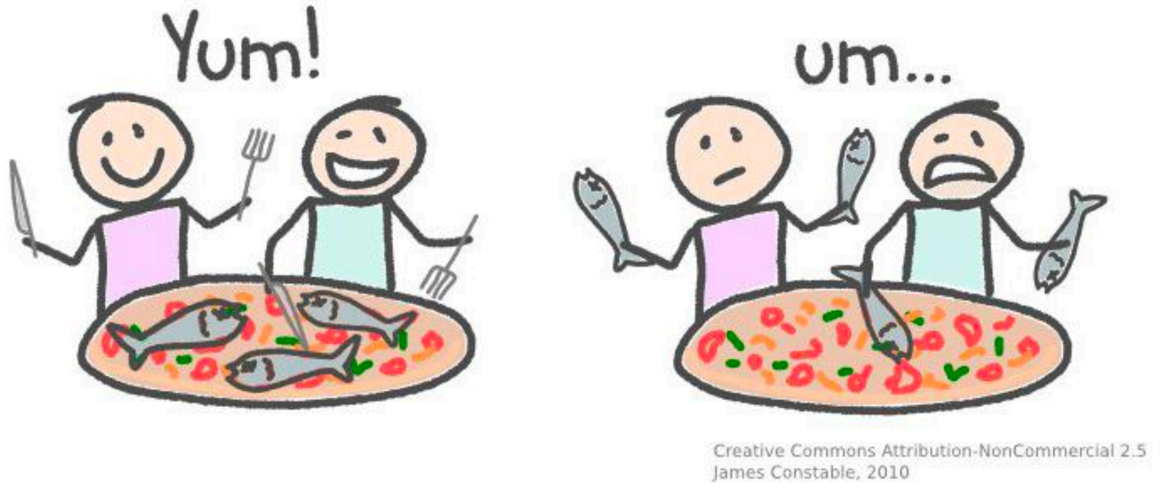
English:

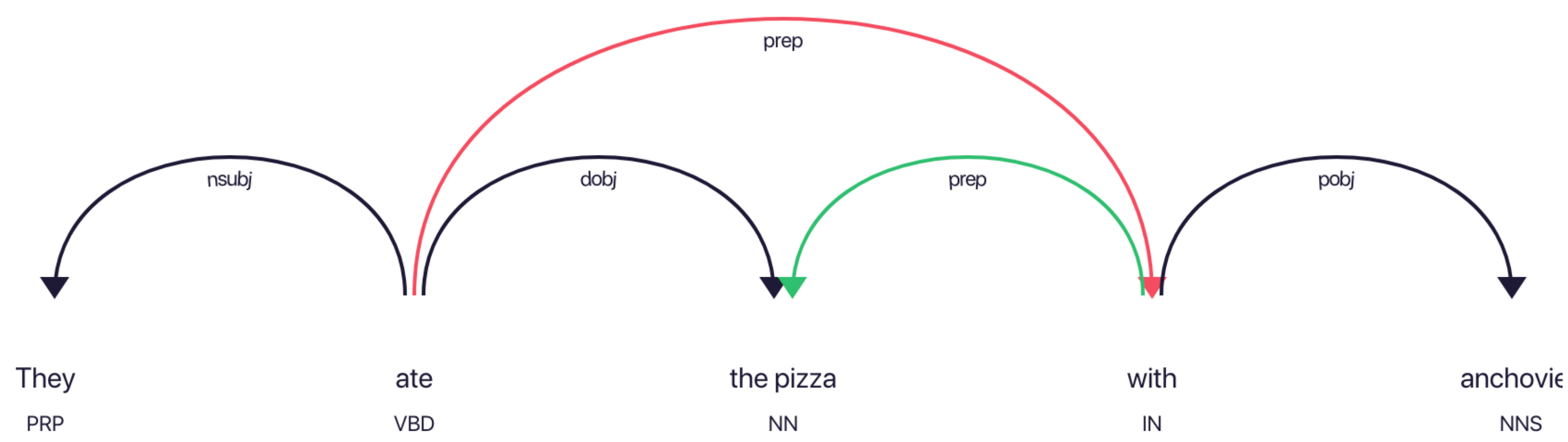
https://tartarus.org/martin/PorterStemmer/

http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/

Parsing

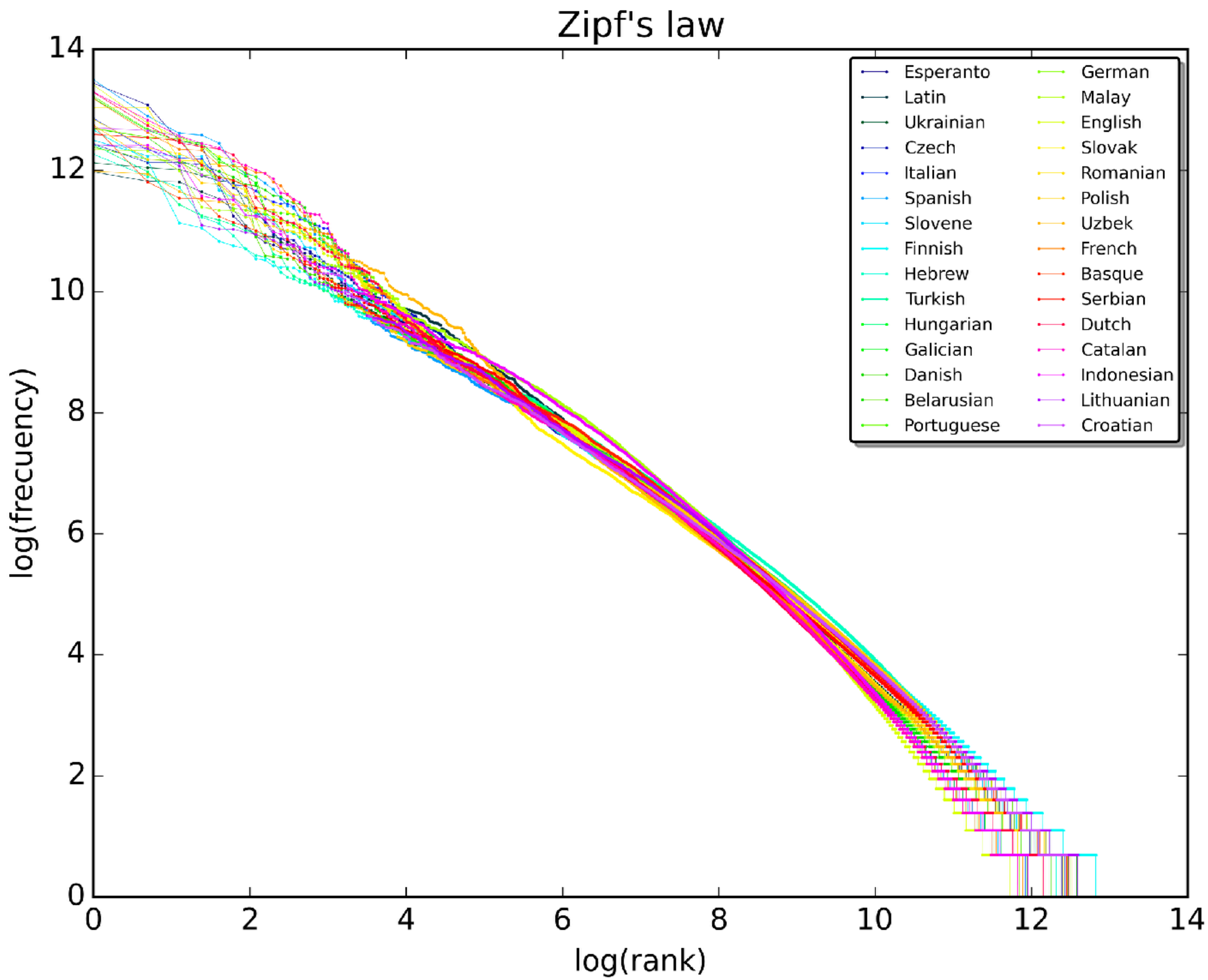
```
They ate the pizza with anchovies
```





source: https://explosion.ai/blog/syntaxnet-in-context

Zipf’s law & long tail



Publicly available corpora

British National Corpus: http://www.natcorp.ox.ac.uk/

Common Crawl: http://commoncrawl.org/the-data/get-started/

Wikipedia: https://dumps.wikimedia.org/

Feature extraction for NLP

1. *the man walked the dog*
2. *the man took the dog to the park*
3. *the dog went to the park*

[dog, man, park, the, to, took, walked, went]

1. [1, 1, 0, 1, 0, 0, 1, 0]
2. [1, 1, 1, 1, 1, 1, 0, 0]
3. [1, 0, 1, 1, 1, 0, 0, 1]

$$\mathrm{idf}(t, D) = \log \frac{N}{|\{d \in D : t \in d\}|}$$

1. [1, 1, 0, 2, 0, 0, 1, 0]
2. [1, 1, 1, 3, 1, 1, 0, 0]
3. [1, 0, 1, 2, 1, 0, 0, 1]

1. [0, 0.18, 0, 0, 0, 0, 0.48, 0]
2. [0, 0.18, 0.18, 0, 0.18, 0.48, 0, 0]
3. [0, 0, 0.18, 0, 0.18, 0, 0, 0.48]

NLP Introduction task

07-text-classification-introduction.ipynb

Language models

- spell checking
- speech recognition
- machine translation
- ...

n-gram models — example

P(<s>, machine, learning, college, </s>) =

P(machine|<s>)P(learning | machine)P(college | learning).P(<s/>|college)

P(learning | machine) = count(machine, learning)/count(machine)

Language model smoothing

- Laplace smoothing (plus one)

$$P(w_i|w_{i-1}) = \frac{count(w_{i-1}, w_i) + 1}{count(w_{i-1}) + V}$$

- interpolation
- Good-Turing
- Witten-Bell
- …

Perplexity

$$PP(W) = P(w_1, w_2, \ldots, w_N)^{-\frac{1}{N}}$$

$$= \sqrt[N]{\frac{1}{P(w_1, w_2, \ldots, w_N)}}$$

$$= \sqrt[N]{\prod_{i=1}^{N} \frac{1}{P(w_i | w_1, \ldots, w_{i-1})}}$$

$$= 2^{-\frac{1}{N} \sum_{i=1}^{N} log_2 P(w_i | w_1, \ldots, w_{i-1})}$$

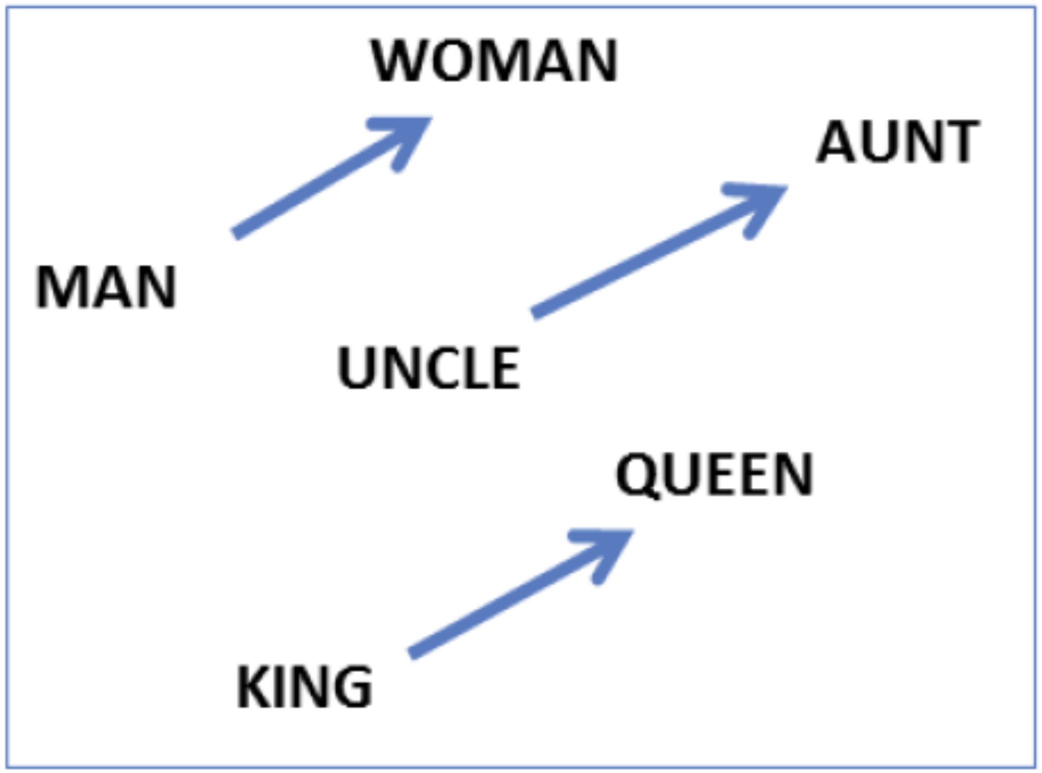
Language detection using language models

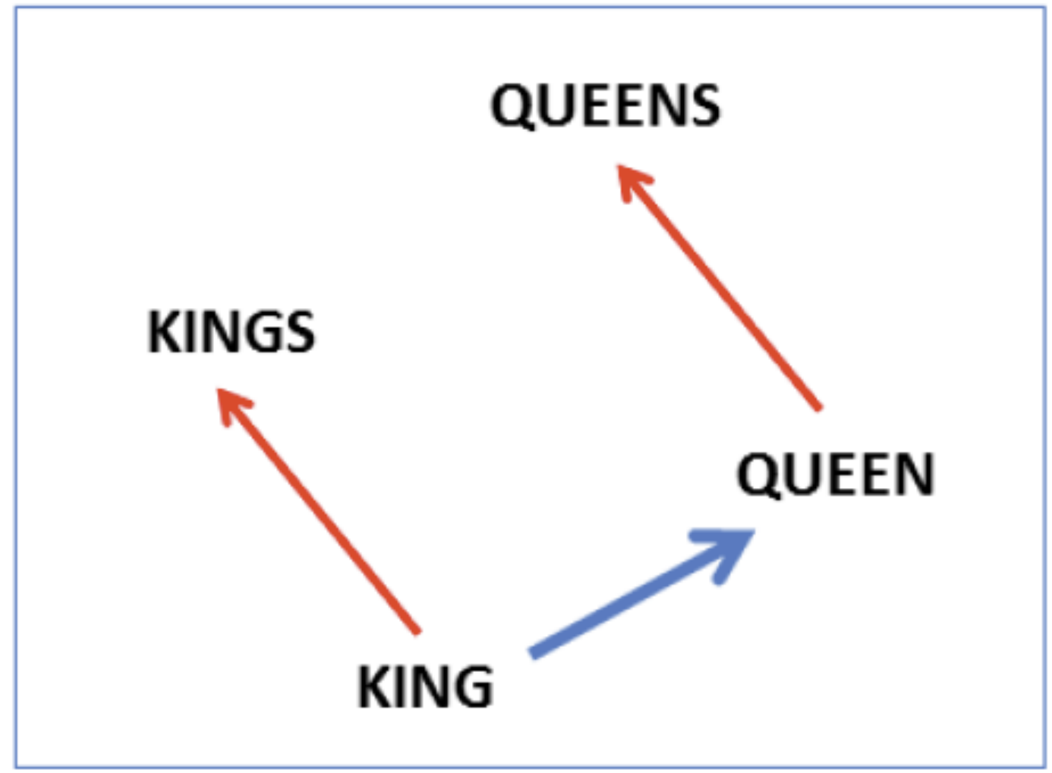
08-Language-detection-assignment.ipynb

Travel agency review classification

09-Review-classification-assignment.ipynb

word2vec

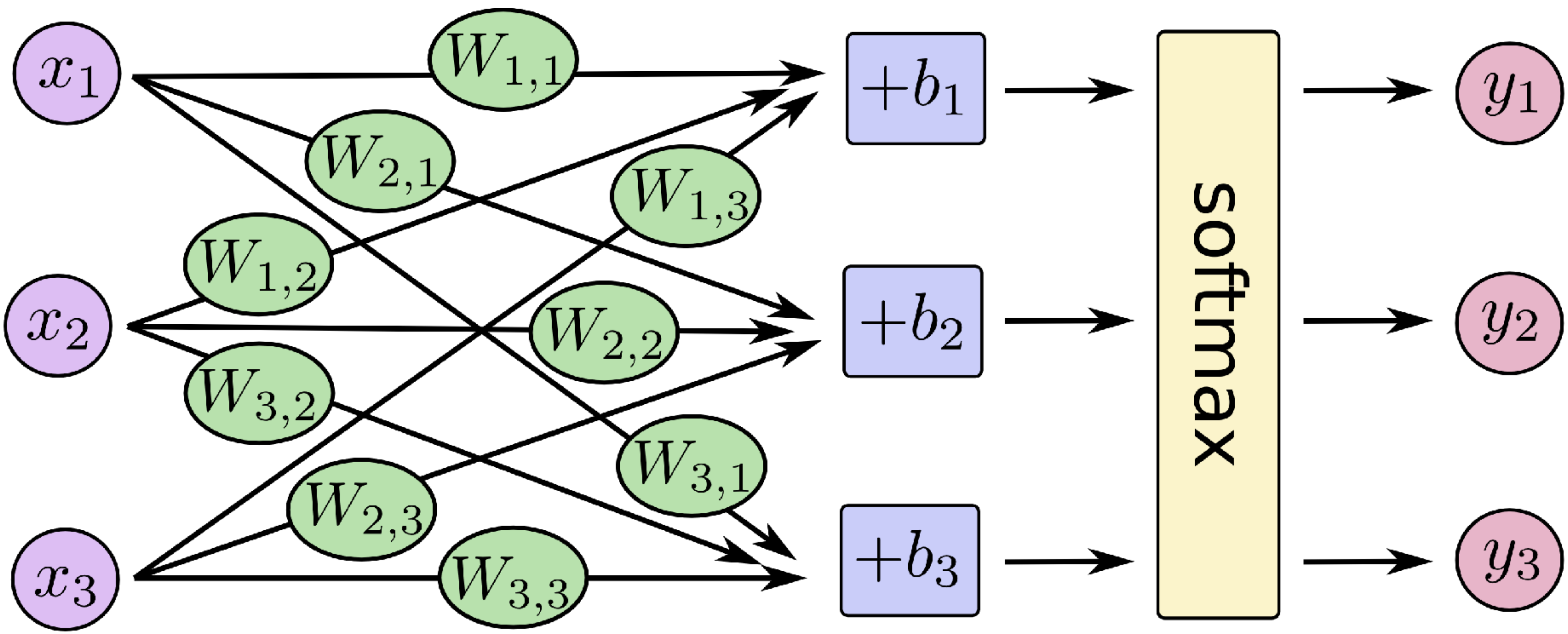




king is to **kings** as **queen** to ***?***.

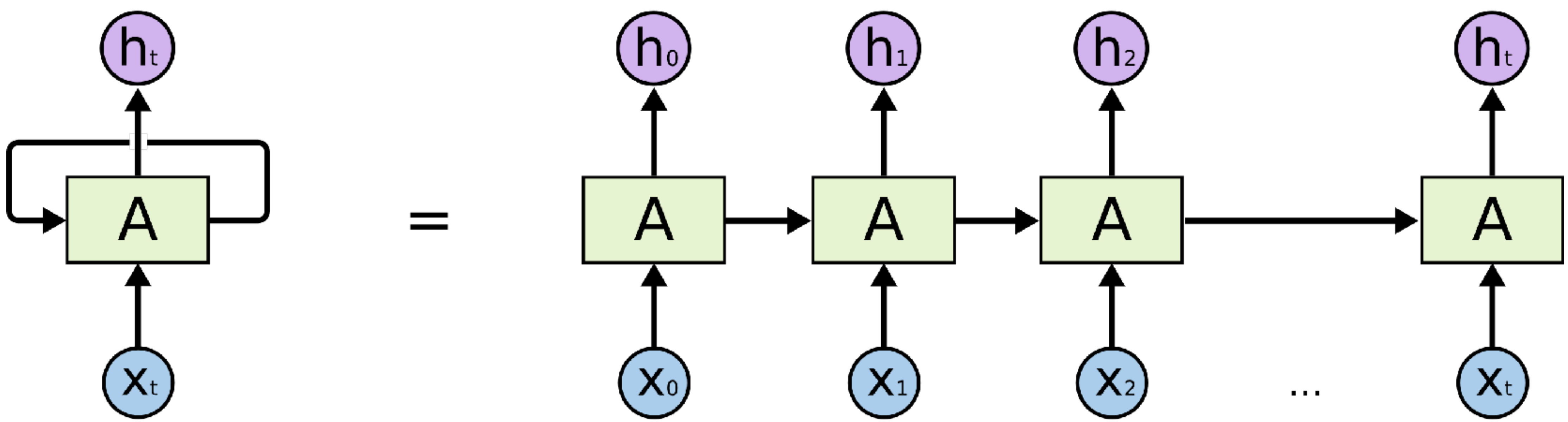
v(**kings**) - v(**king**) = v(**queens**) - v (**queen**)

Feed-Forward Neural Network



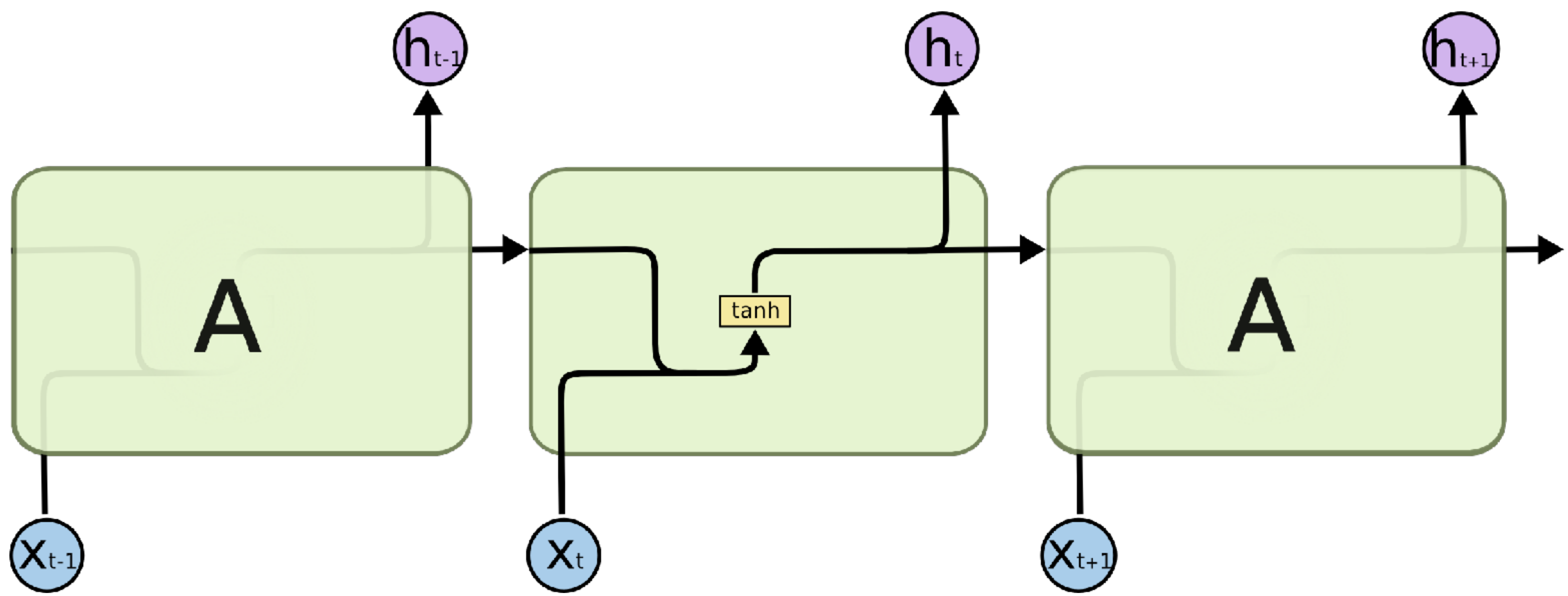
source: https://www.tensorflow.org

Recurrent Neural networks 1/2



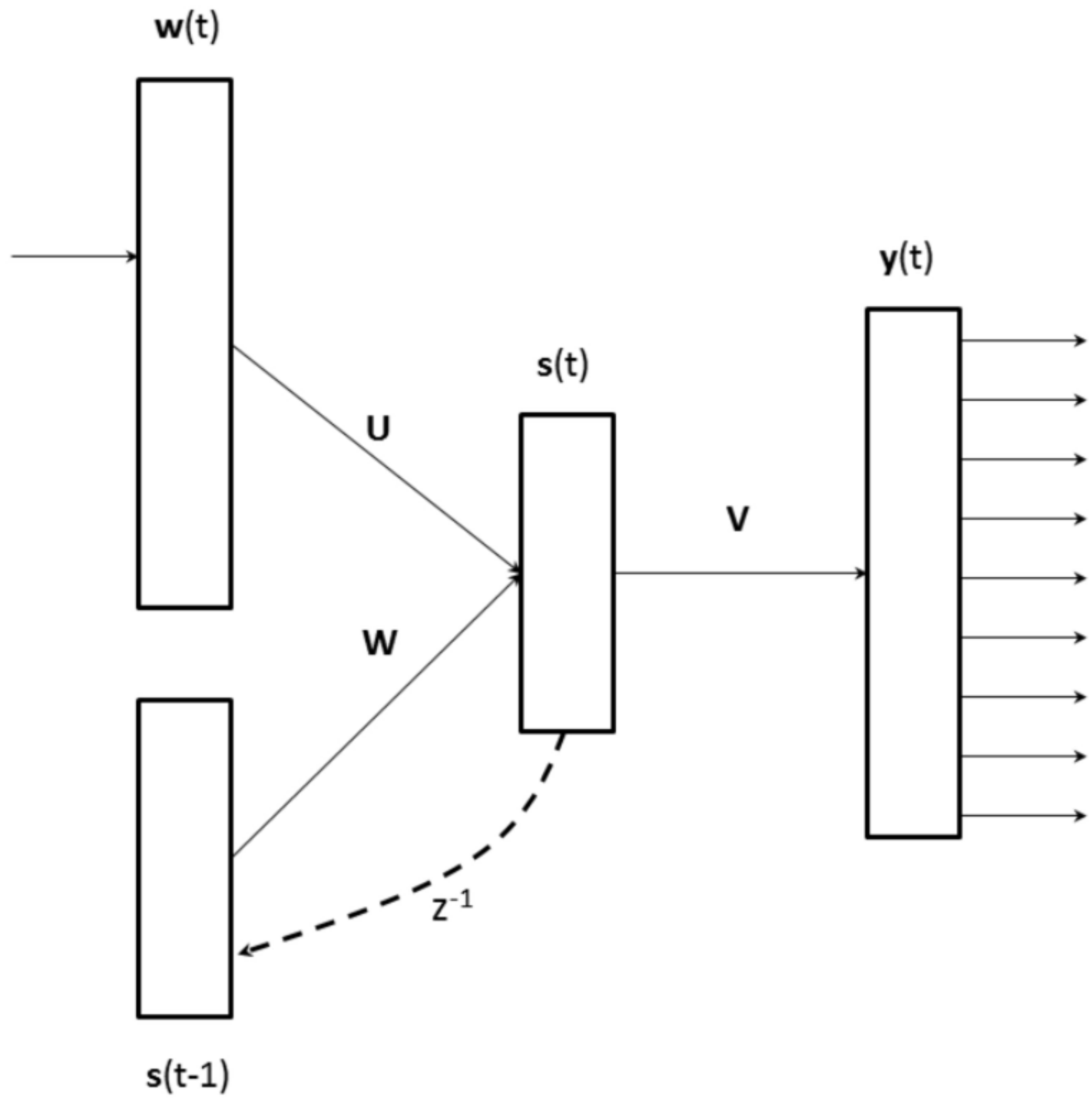
source: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Recurrent Neural Networks 2/2



https://www.tensorflow.org

Recurrent Neural Network Language Modeling Toolkit



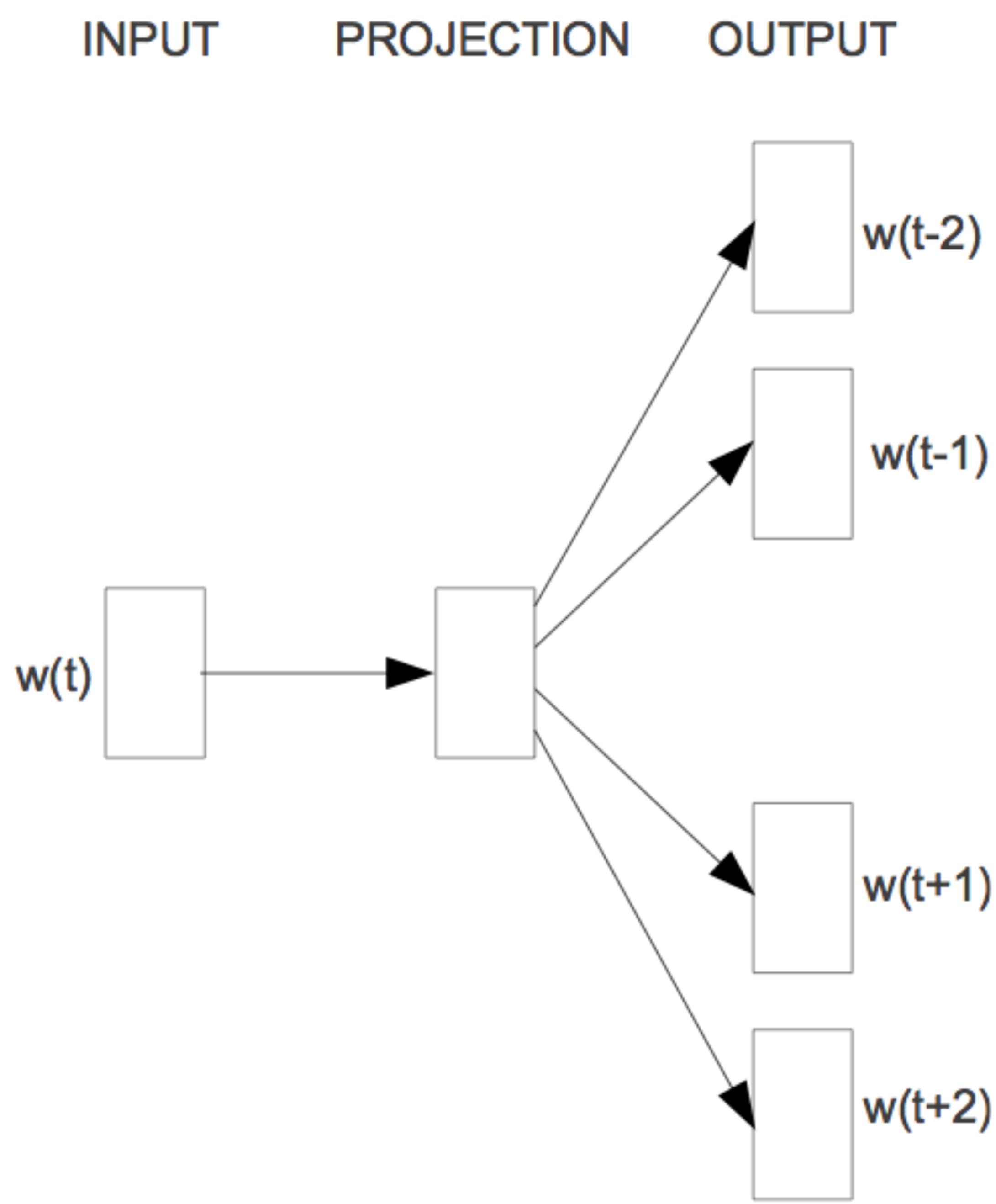
$$\mathbf{s}(t) = f\left(\mathbf{U}\mathbf{w}(t) + \mathbf{W}\mathbf{s}(t-1)\right)$$

$$\mathbf{y}(t) = g\left(\mathbf{V}\mathbf{s}(t)\right),$$

$$f(z) = \frac{1}{1+e^{-z}}, \quad g(z_m) = \frac{e^{z_m}}{\sum_k e^{z_k}}.$$

source: http://www.fit.vutbr.cz/~imikolov/rnnlm/

The skip-gram model



Experiments with word2vec

10-Word2vec-in-gensim.ipynb

11-Review-classification-w2v-assignment.ipynb

Language models for text generating

LISTOPAD

usínám, pláču, umírám, přemýšlím
co cítíš ty?
cítím tvou slabost
a whisky

NOVEMBER

I am falling asleep, crying, dying, thinking
what do you feel?
I feel your weakness
and whisky

SPRAVEDLNOST

na tvou dekadentní duši
ráno i v poledne
bůh má připravenou kuši

JUSTICE

for your decadent soul
in the morning, in the evening
the god has prepared a crossbow

Metaphores

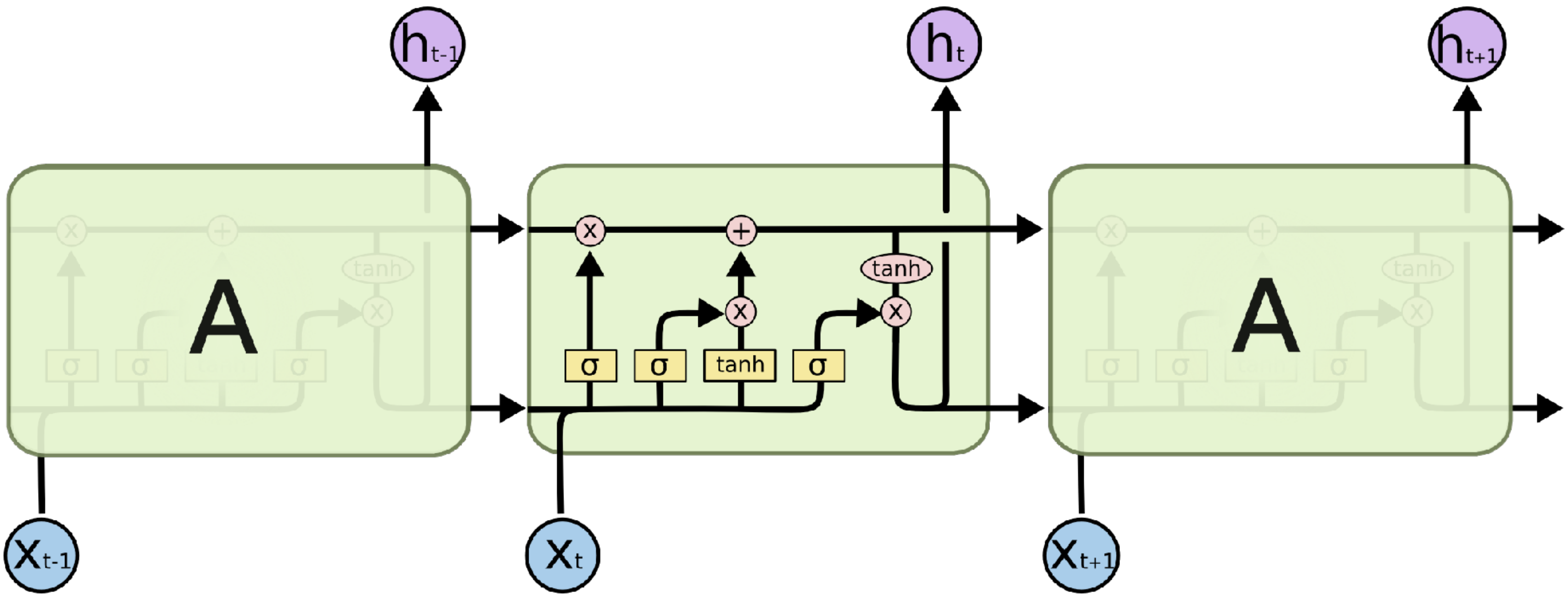
...tělo plné červánků...

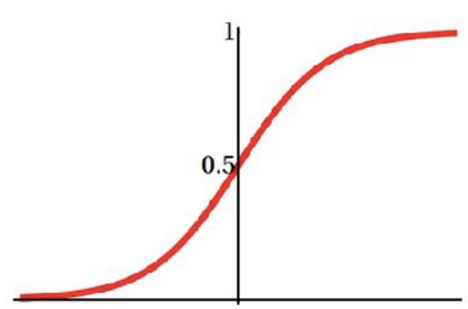
...body full of blush of dawn...

...tak vzácný jako listí...

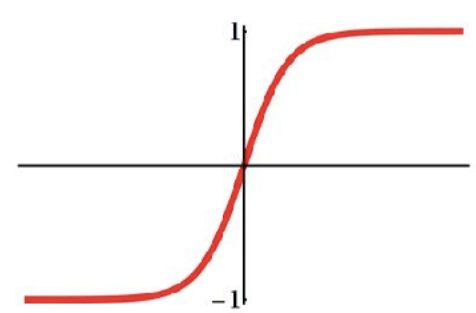
...as rare as leaves of trees...

Long Short-Term Memory





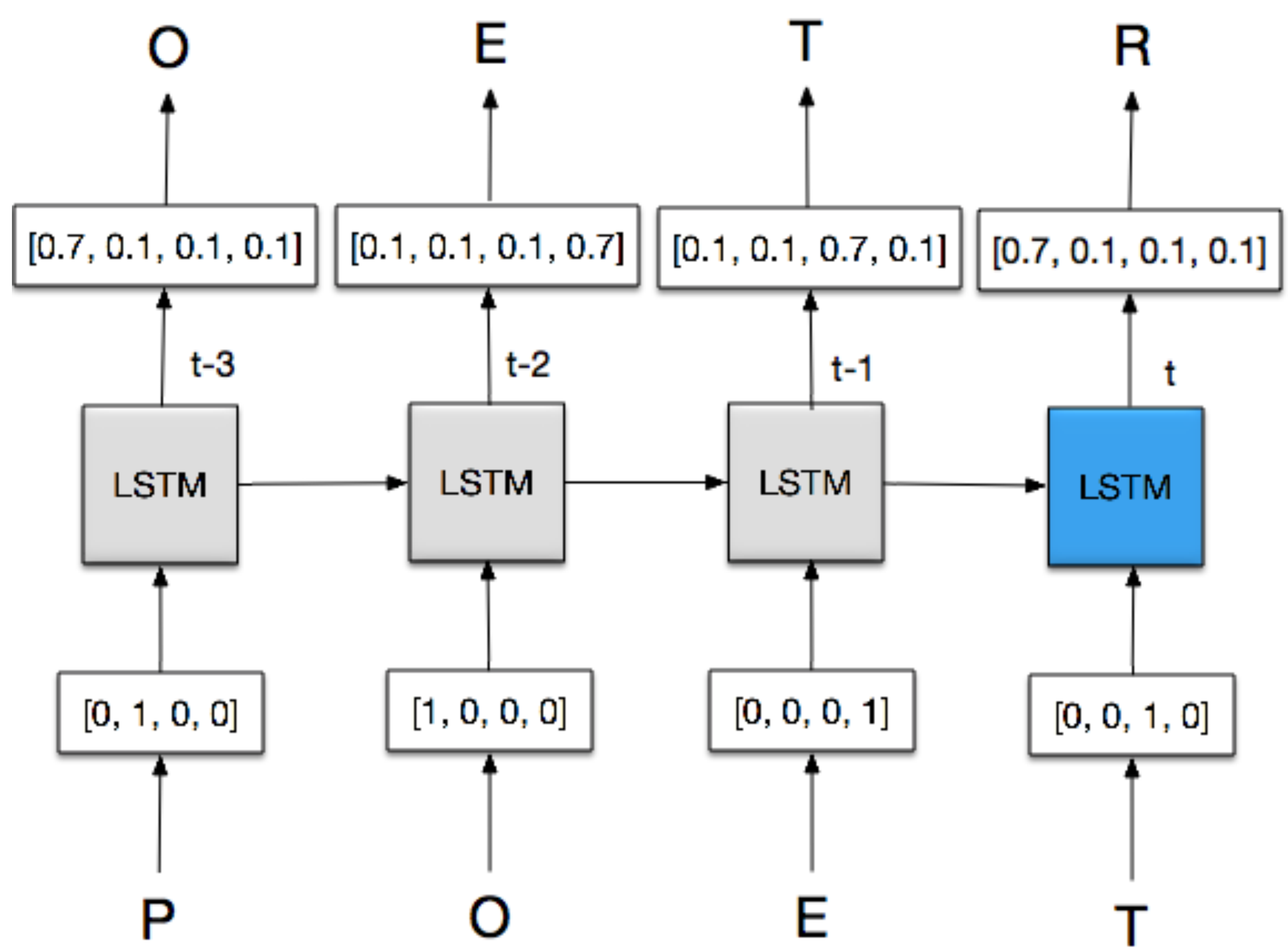
sigmoid

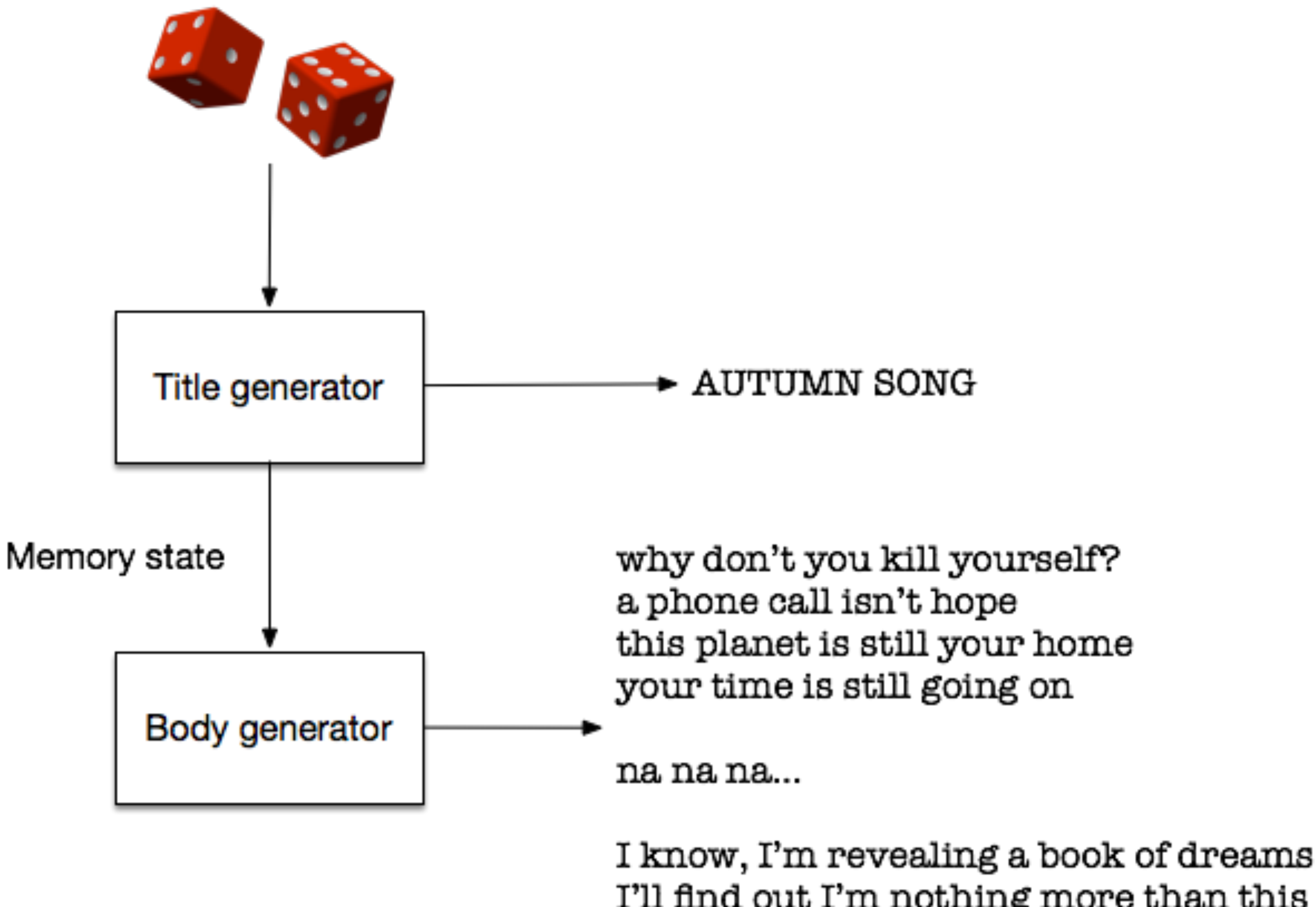


tanh

Source: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

LSTM language model

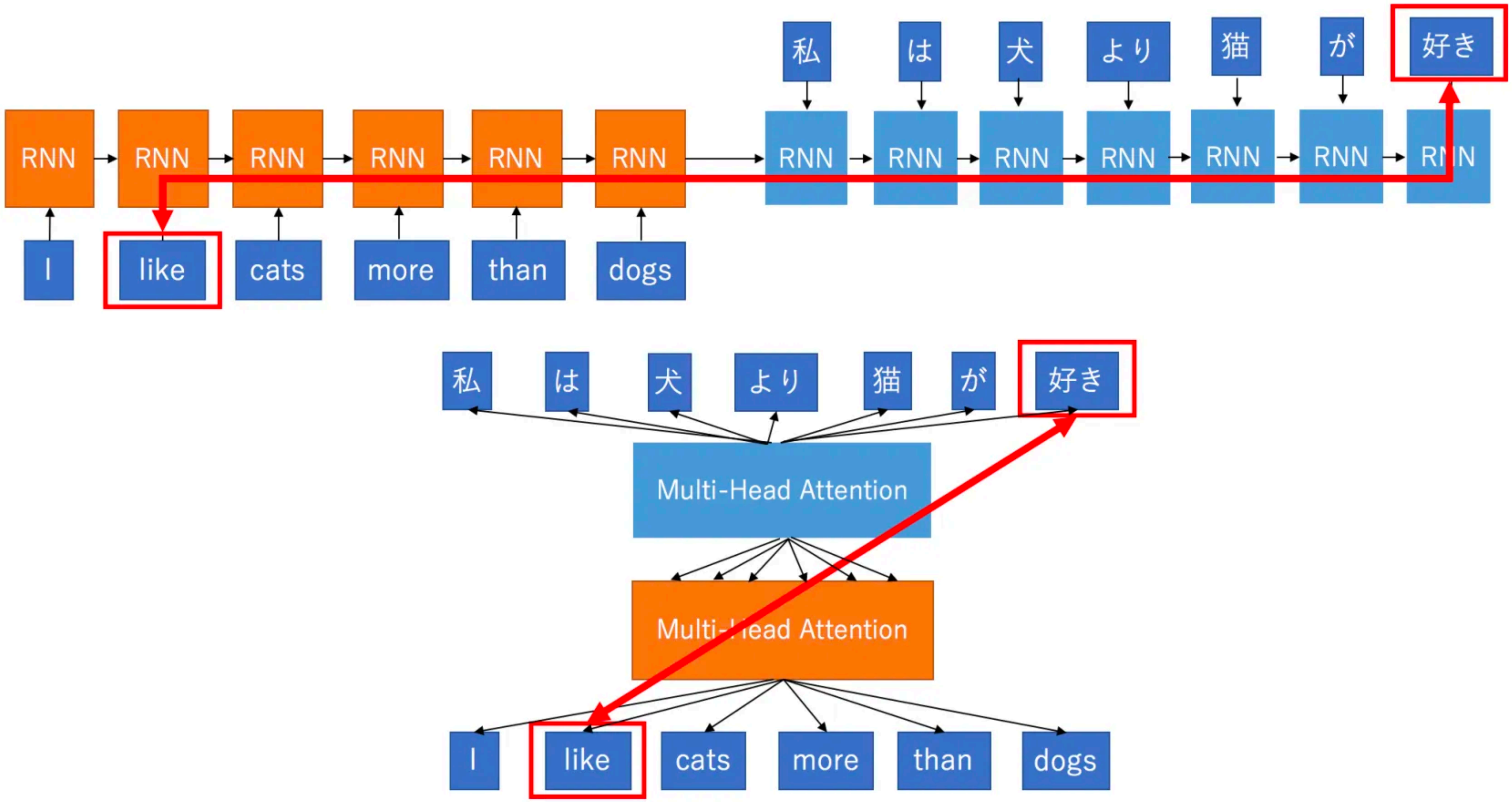



Title generator
AUTUMN SONG
Memory state
Body generator
why don't you kill yourself?
a phone call isn't hope
this planet is still your home
your time is still going on
na na na...
I know, I'm revealing a book of dreams
I'll find out I'm nothing more than this

LSTM review generator

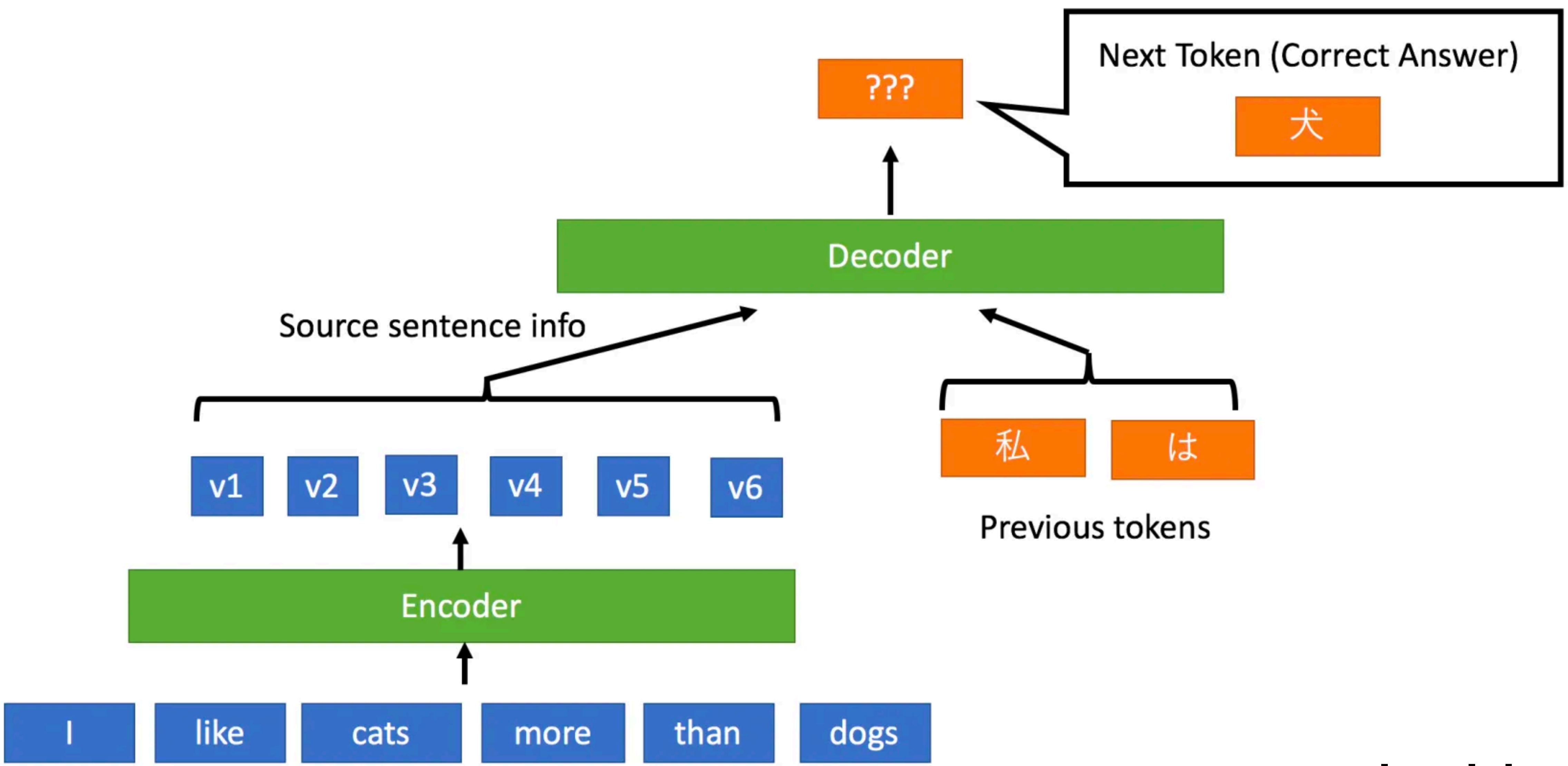
12-Review-generator.ipynb

Transformer



source: www.mlexplained.com

Translation with Transformers



GPT-2 Language model

Donald Trump told...

Demo: https://talktotransformer.com/

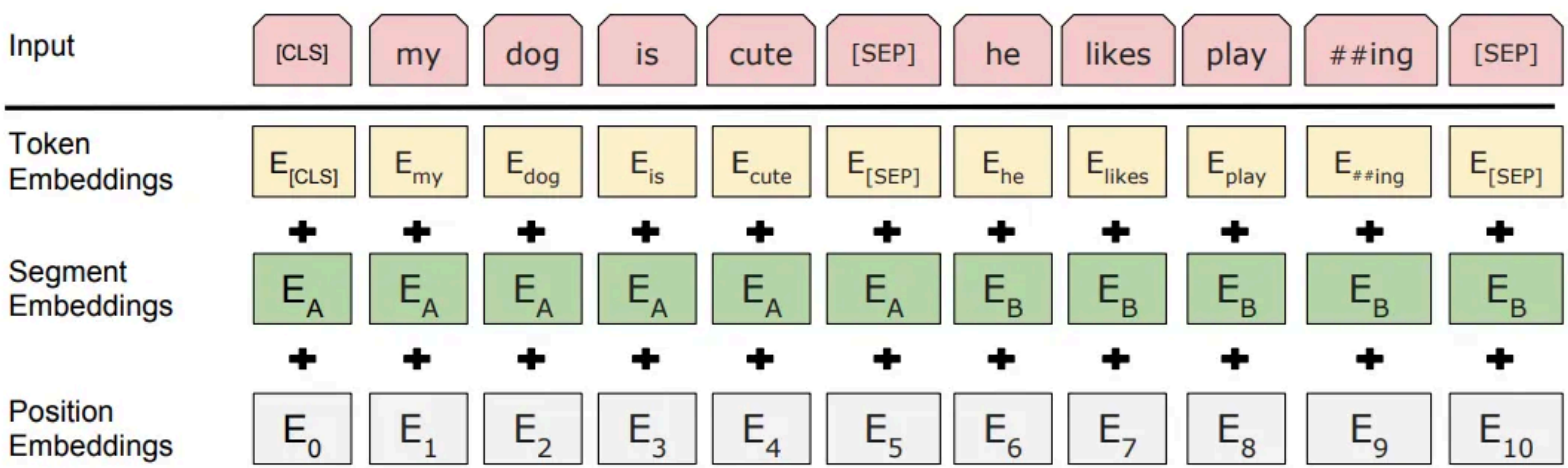
GPT-2 Language model

Donald Trump told the Times he is preparing a "major speech" on his economic plans, but did not provide details on what it will entail.

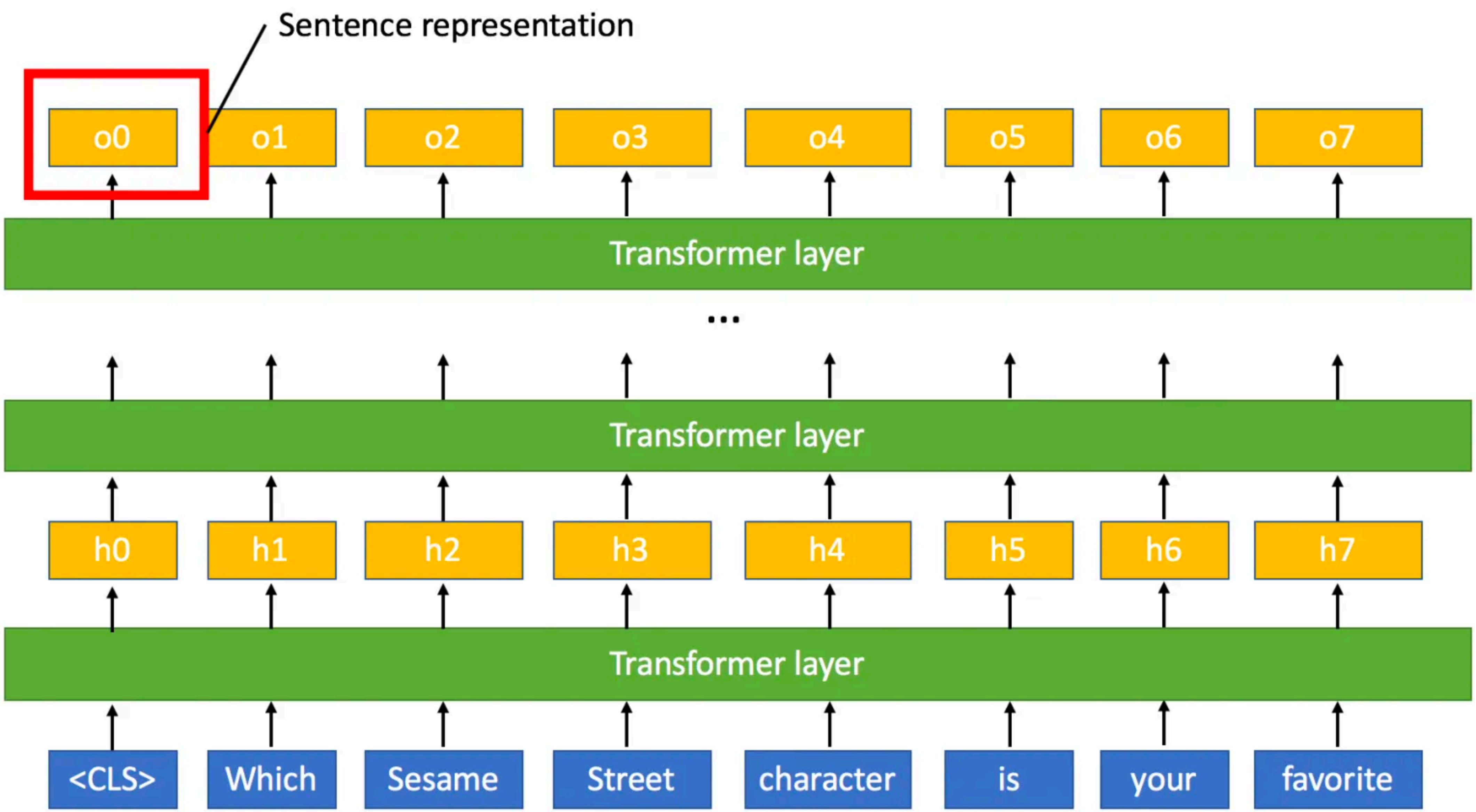
“I’m getting ready for the speech. And I will have a major speech on Tuesday.” Trump said during an interview in the White House residence.

Demo: https://talktotransformer.com/

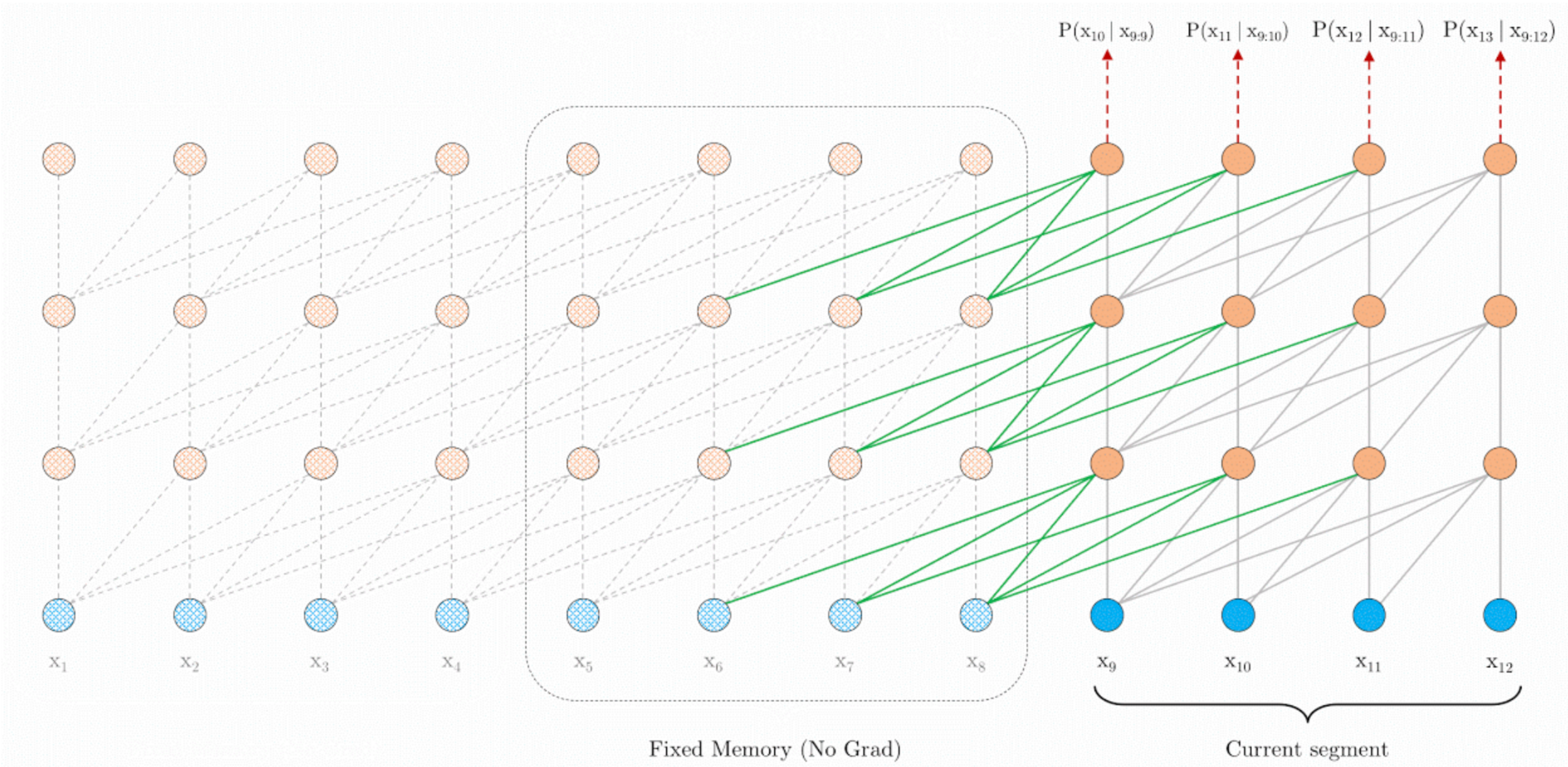
BERT
(input encoding)



BERT (classification)

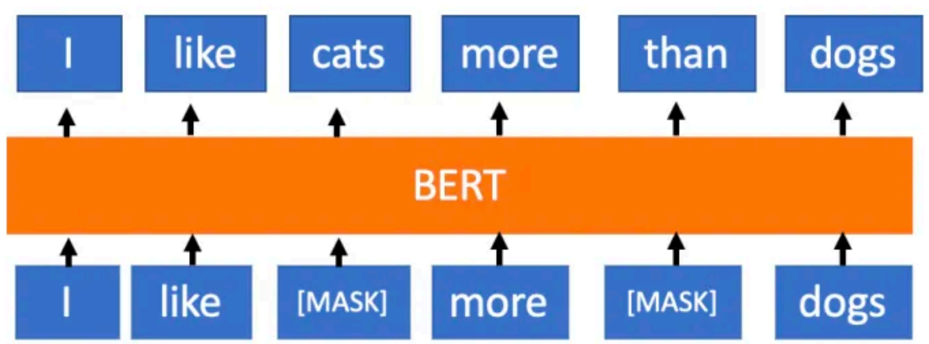


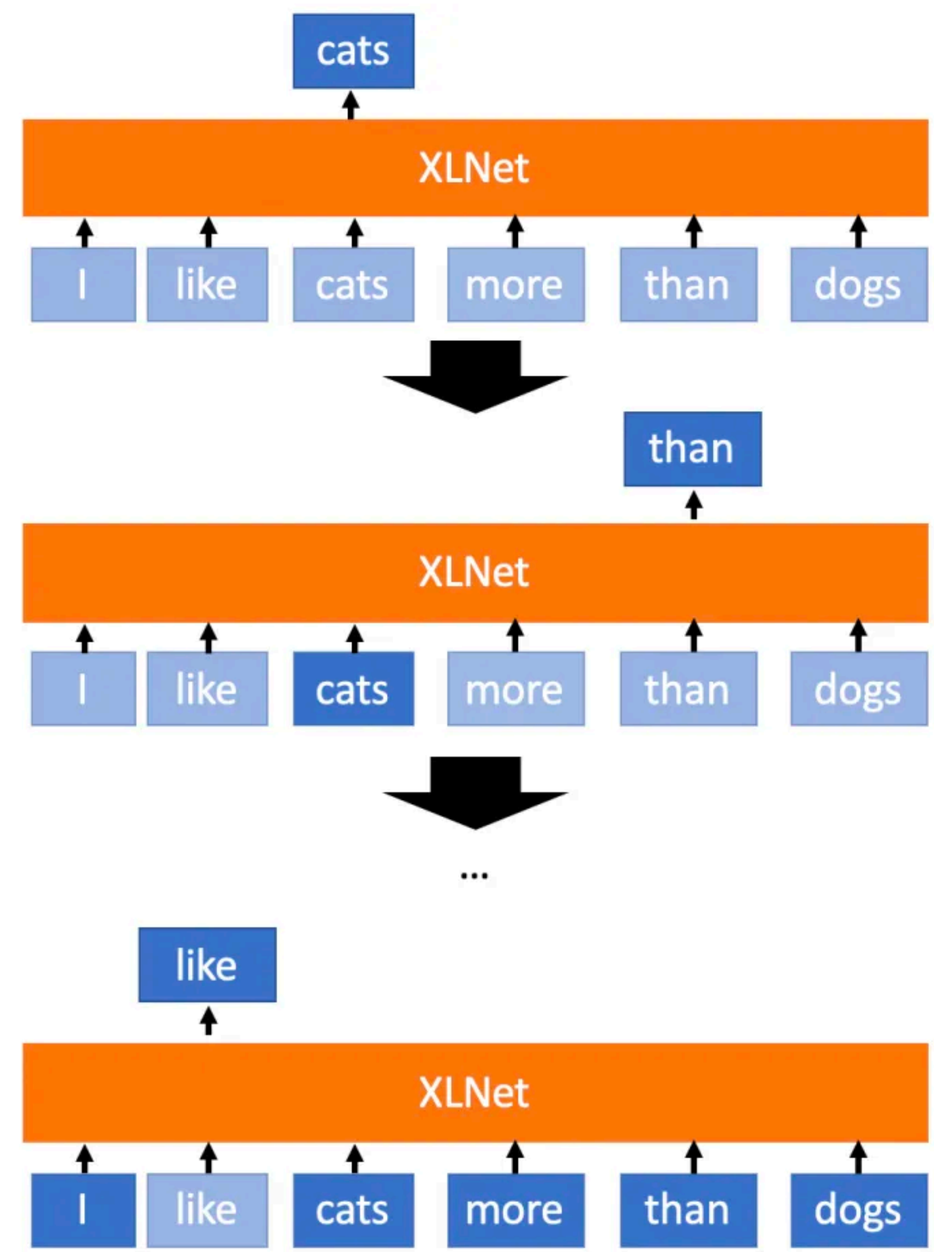
Transformer XL



source: ai.googleblog.com

XLNet
(permutation language model)





Thank you for your attention

e-mail: jiri@mlguru.com

Web: www.mlguru.com

Twitter: @JiriMaterna

Facebook: https://www.facebook.com/maternajiri

LinkedIn: https://www.linkedin.com/in/jirimaterna/