Workshop AI: CNN, NLP, and deployment

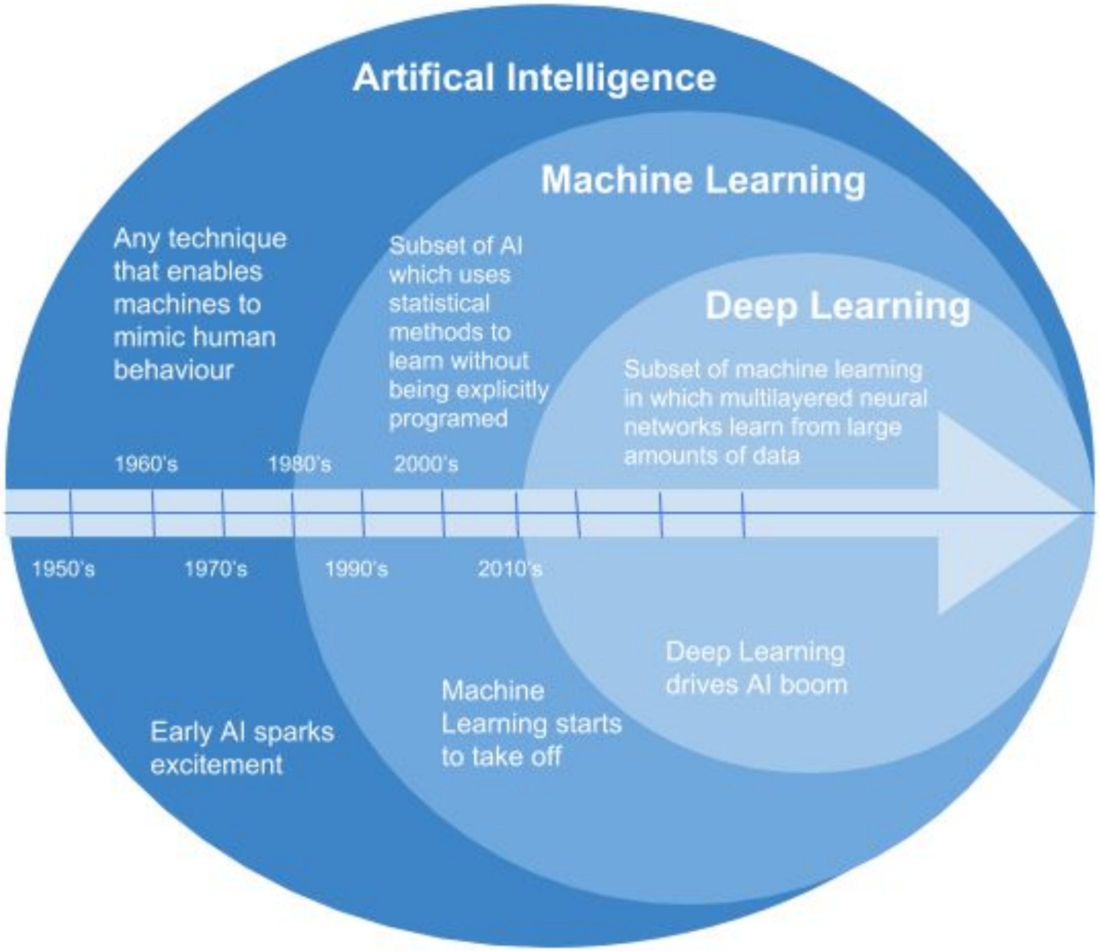
Nele Custers
nele.custers@pxl.be

Teaching:

- Programming in Python, Java, and Neural Networks.

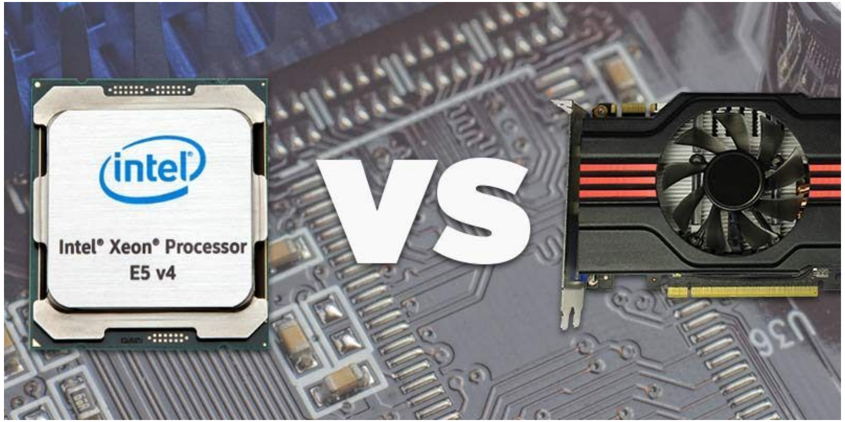
Educational Background:

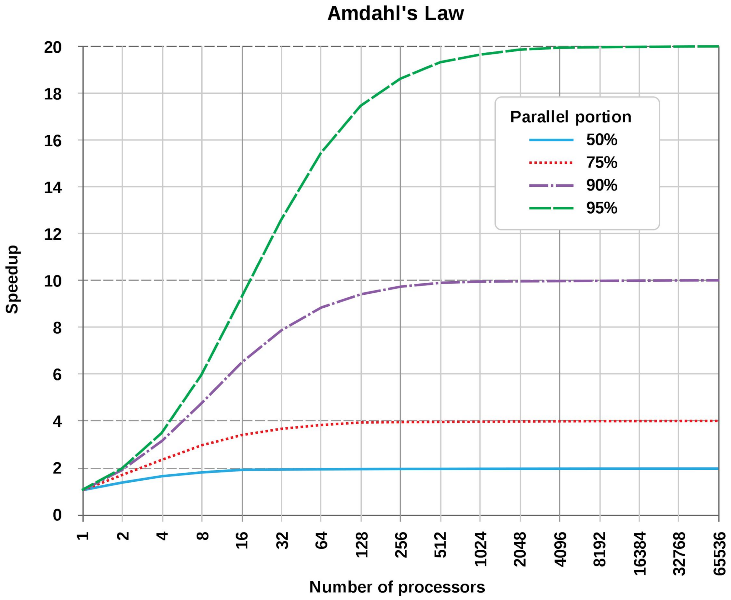
- Degree in Informatics
- Worked as a Java Consultant


Artifical Intelligence
Machine Learning
Deep Learning
Any technique that enables machines to mimic human behaviour
Subset of AI which uses statistical methods to learn without being explicitly programed
Subset of machine learning in which multilayered neural networks learn from large amounts of data
1960's
1980's
2000's
1950's
1970's
1990's
2010's
Early AI sparks excitement
Machine Learning starts to take off
Deep Learning drives AI boom

Base conditions for AI

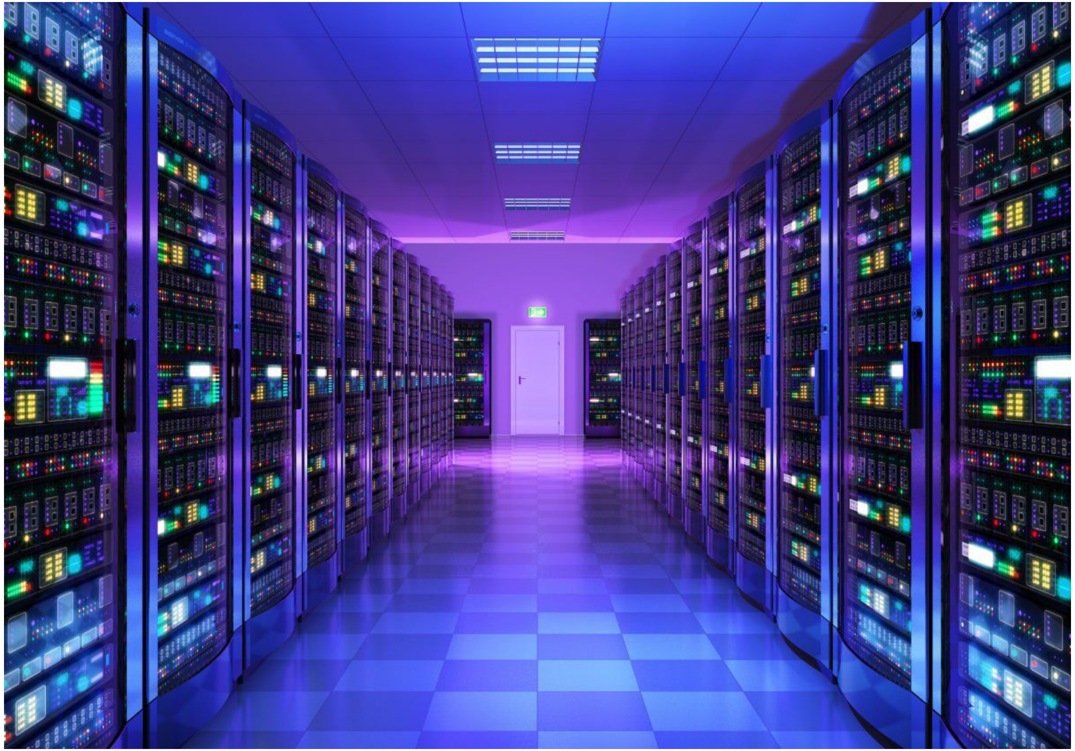
GPU





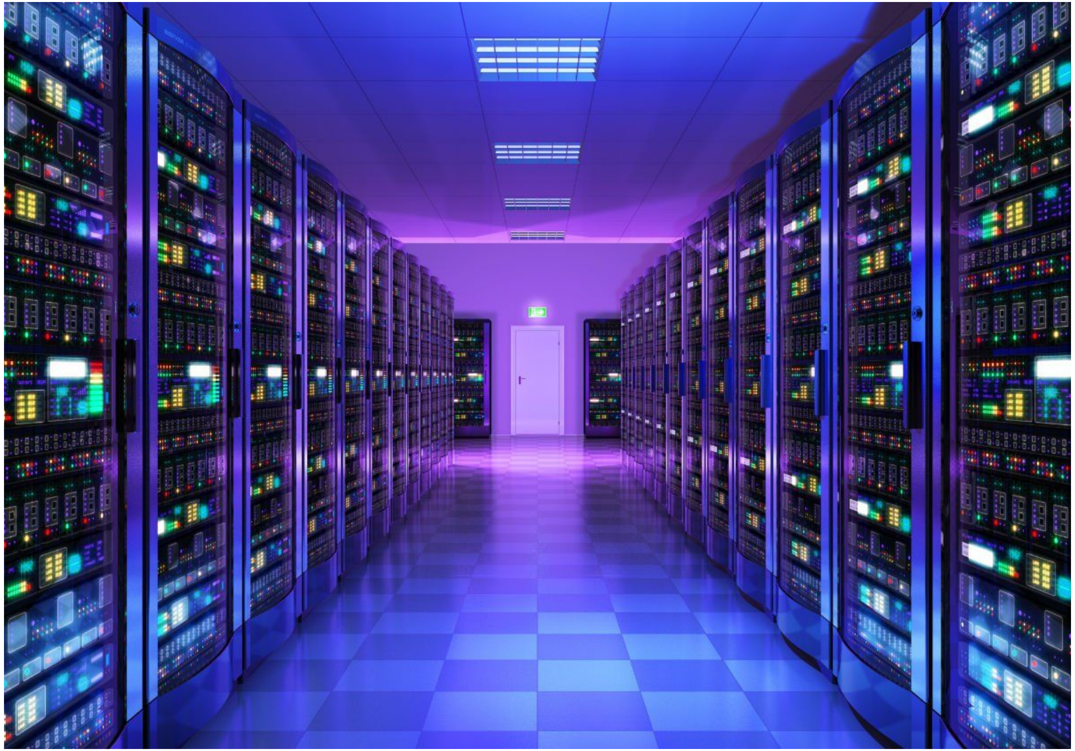
Base conditions for AI

Storage
Big Data

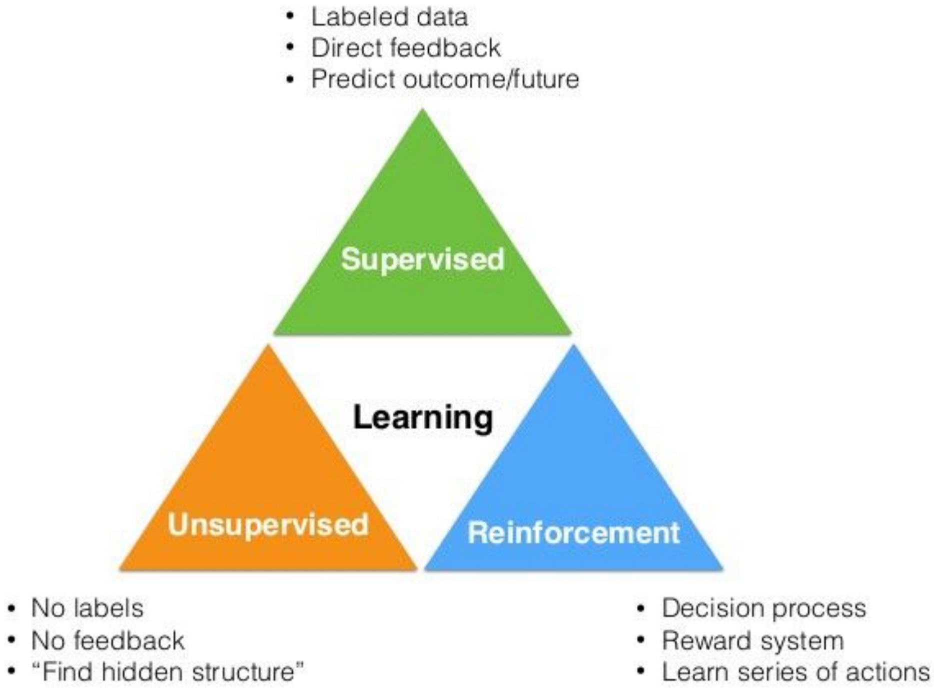


Base conditions for AI

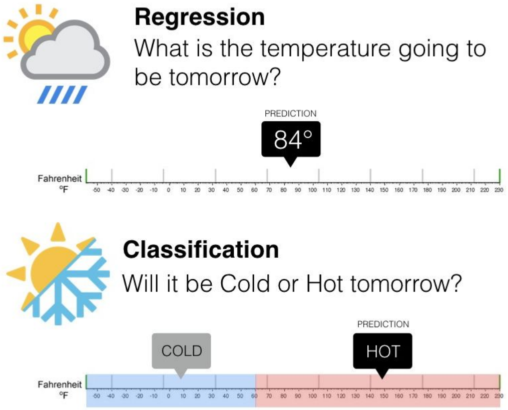
Base algorithms in the cloud

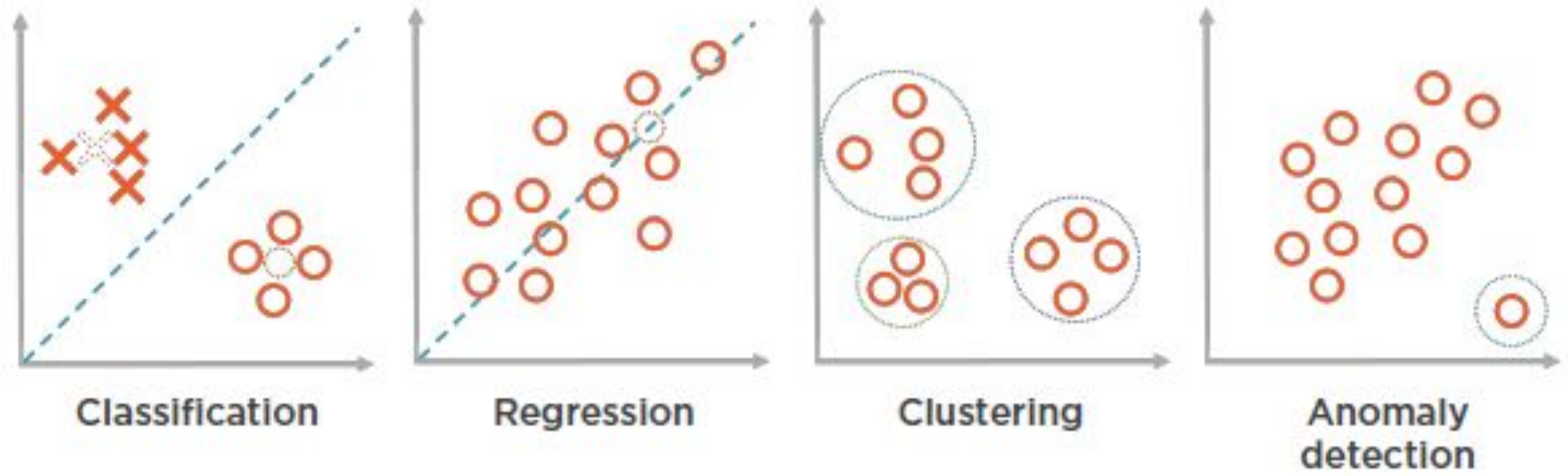


Machine learning algorithms

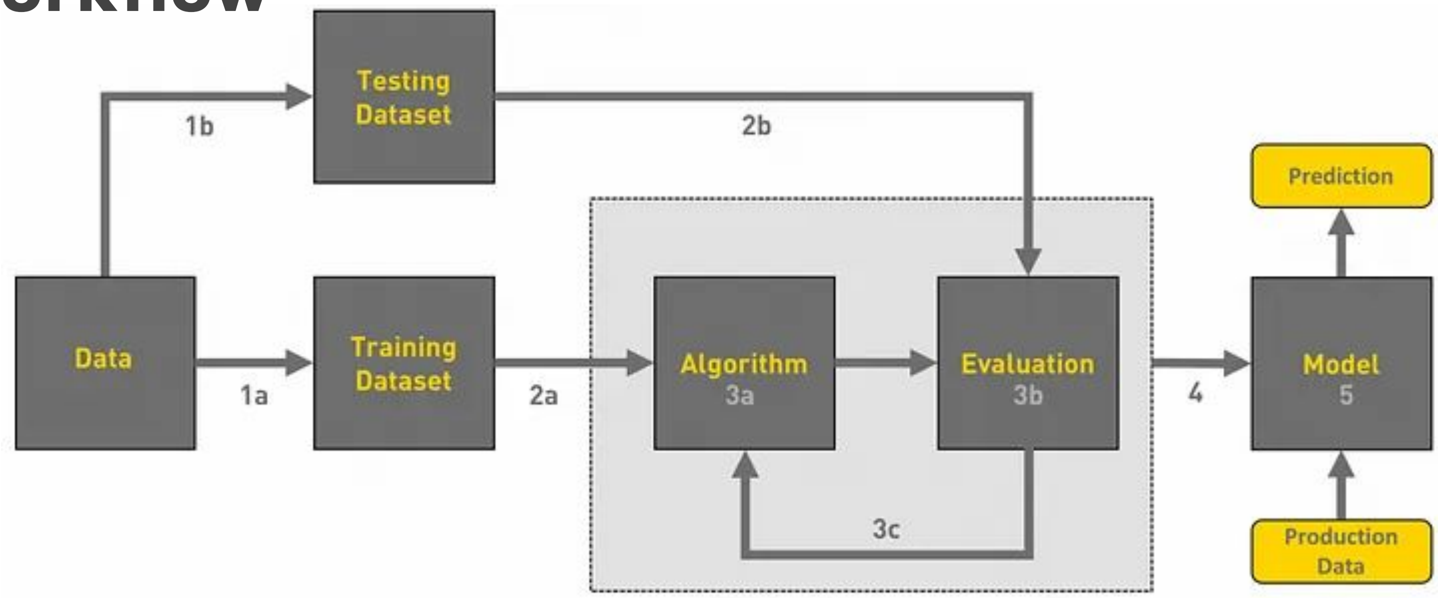


ML problems


Regression
What is the temperature going to be tomorrow?
PREDICTION
84°
Fahrenheit °F
Classification
Will it be Cold or Hot tomorrow?
PREDICTION
COLD
HOT
Fahrenheit °F


Classification
Regression
Clustering
Anomaly detection

ML workflow



Testing Dataset
1b
2b
Prediction
Data
1a
Training Dataset
2a
Algorithm 3a
Evaluation 3b
4
Model 5
3c
Production Data

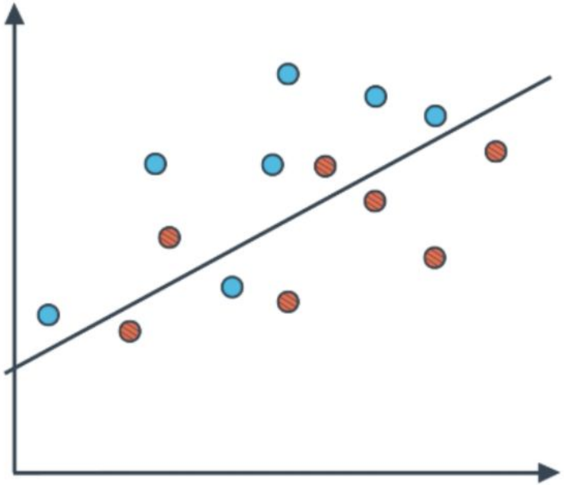
Training data

Year	Breakthroughs in AI	Datasets (First Available)	Algorithms (First Proposed)
1994	Human-level spontaneous speech recognition	Spoken Wall Street Journal articles and other texts (1991)	Hidden Markov Model (1984)
1997	IBM Deep Blue defeated Garry Kasparov	700,000 Grandmaster chess games, aka "The Extended Book" (1991)	Negascout planning algorithm (1983)
2005	Google's Arabic- and Chinese-to-English translation	1.8 trillion tokens from Google Web and News pages (collected in 2005)	Statistical machine translation algorithm (1988)
2011	IBM Watson became the world Jeopardy! champion	8.6 million documents from Wikipedia, Wiktionary, Wikiquote, and Project Gutenberg (updated in 2010)	Mixture-of-Experts algorithm (1991)
2014	Google's GoogLeNet object classification at near-human performance	ImageNet corpus of 1.5 million labeled images and 1,000 object categories (2010)	Convolution neural network algorithm (1989)
2015	Google's Deepmind achieved human parity in playing 29 Atari games by learning general control from video	Arcade Learning Environment dataset of over 50 Atari games (2013)	Q-learning algorithm (1992)
Average No. of Years to Breakthrough:		**3 years**	**18 years**

Wissner-Gross, Alexander (2016). Datasets Over Algorithms.

"Goodness of a model" - Confusion Matrix

Confusion Matrix Summary

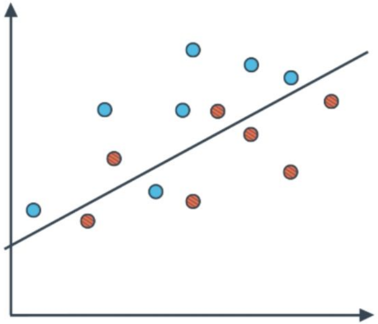


	Guessed Positive	Guessed Negative
Positive	True Positives	False Negatives
Negative	False Positives	True Negatives

Blue Points — labelled positive; Red Points — labelled negative

“Goodness of a model” - Confusion Matrix

Confusion Matrix Summary



	Guessed Positive	Guessed Negative
Positive	True Positives	False Negatives
Negative	False Positives	True Negatives

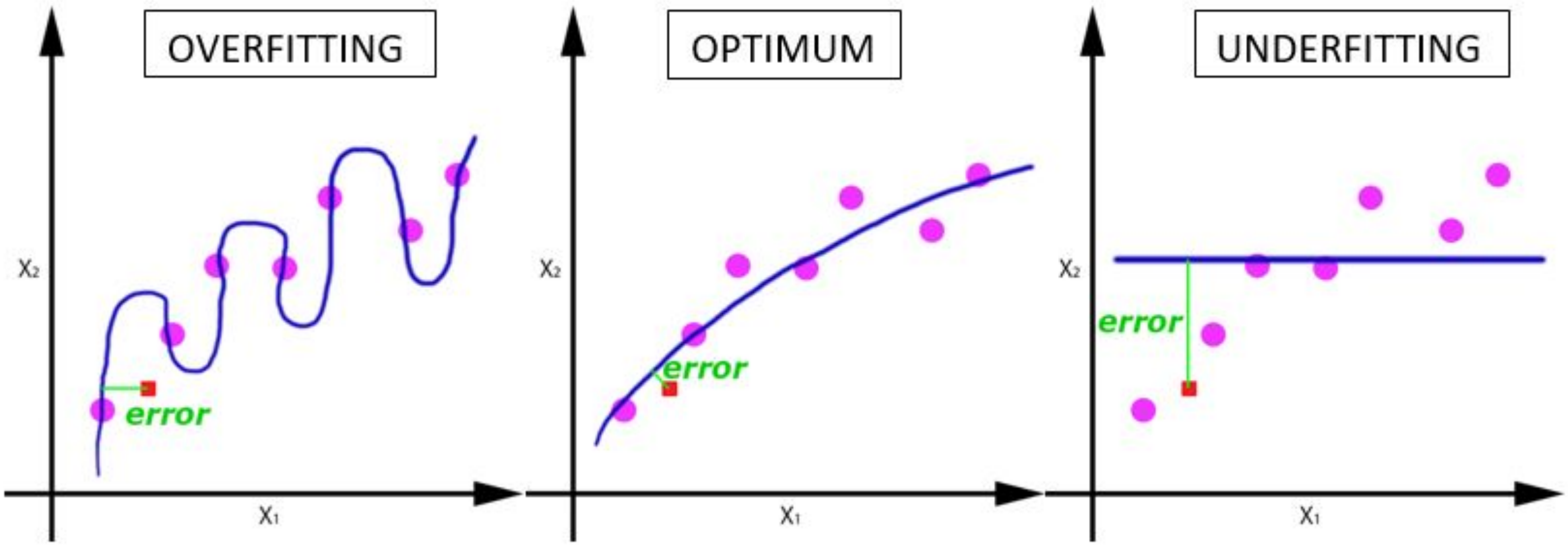
Blue Points — labelled positive; Red Points — labelled negative

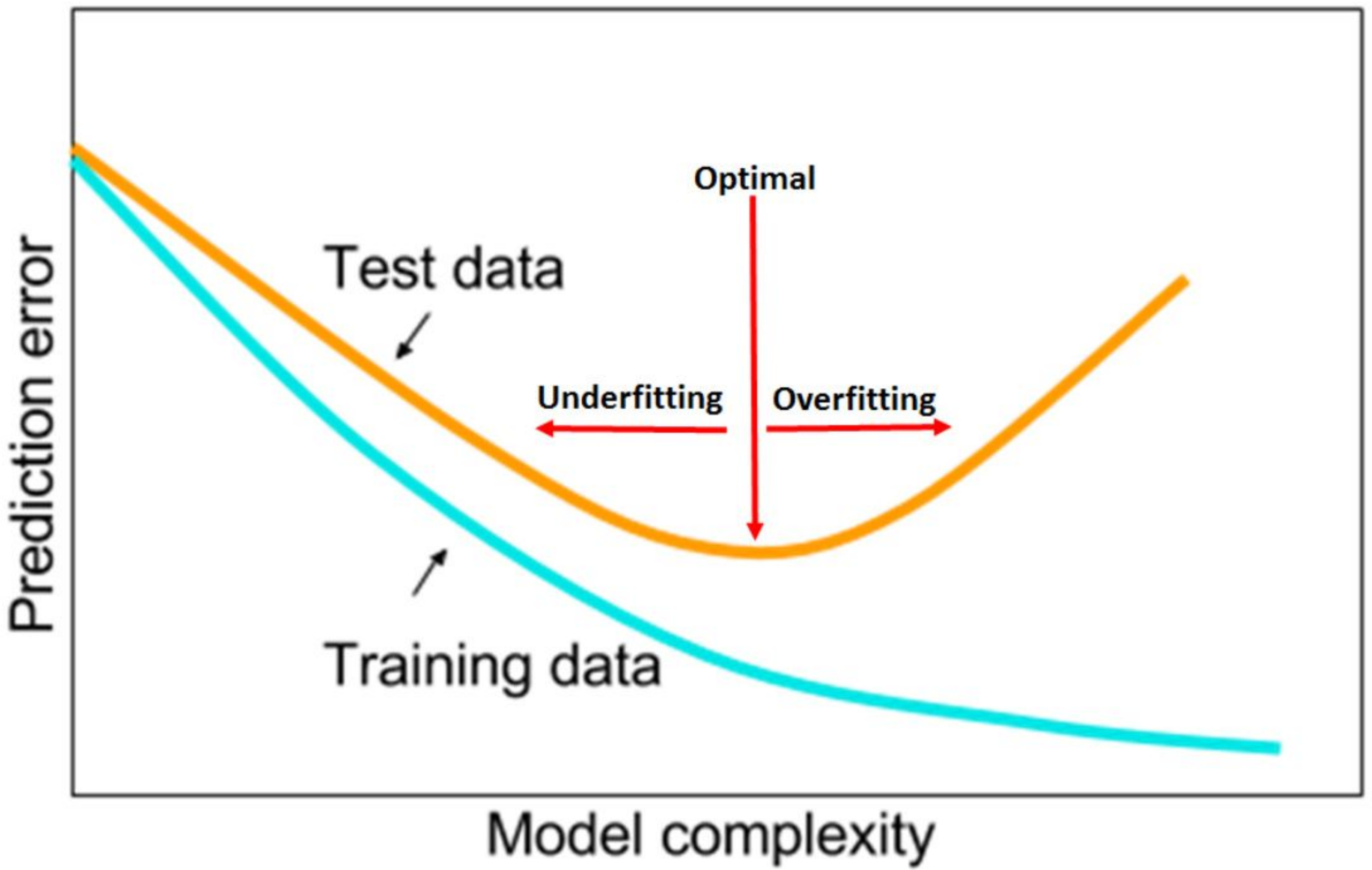
$$accuracy = (True\ Positives + True\ Negatives) \,/\, Total\ Points$$

$$precision = True\ Positives \,/\, (True\ Positives + False\ Positives)$$

$$recall = True\ Positives \,/\, (True\ Positives + False\ Negatives)$$

Bias and variance




Optimal
Test data
Underfitting
Overfitting
Training data
Prediction error
Model complexity

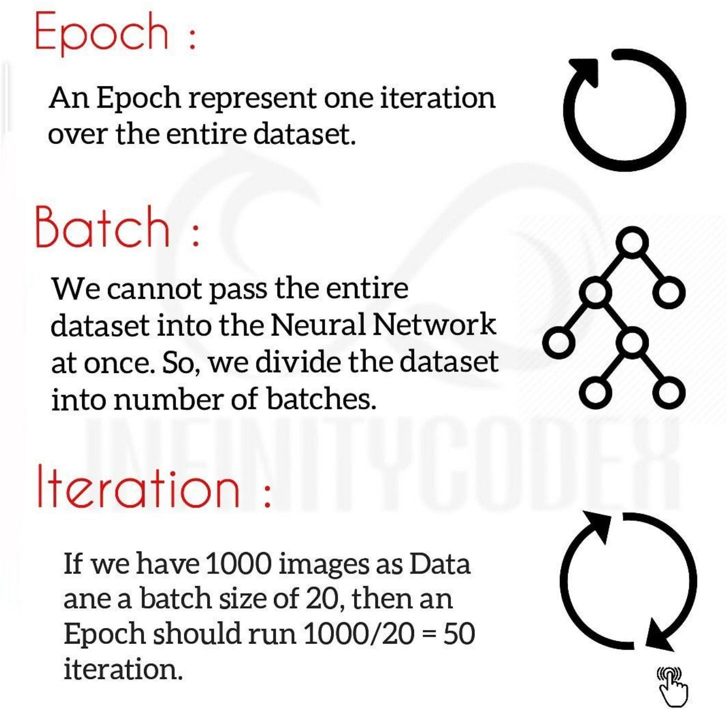
Bias-Variance Dilemma and No. of Features

high bias	high variance
pays little attention to data	pays too much attention to data (does not generalize well)
oversimplified	overfits
high error on training set (low r^2, large SSE)	much higher error on test set than on training set

Bias all around...

Coded bias (trailer): https://www.youtube.com/watch?v=jZI55PsfZJQ

Terminology



The typically mini-batch sizes are 64, 128, 256 or 512.
And, in the end, make sure the minibatch fits in the CPU/GPU.

Datasets

https://www.kaggle.com/datasets/grassknoted/asl-alphabet

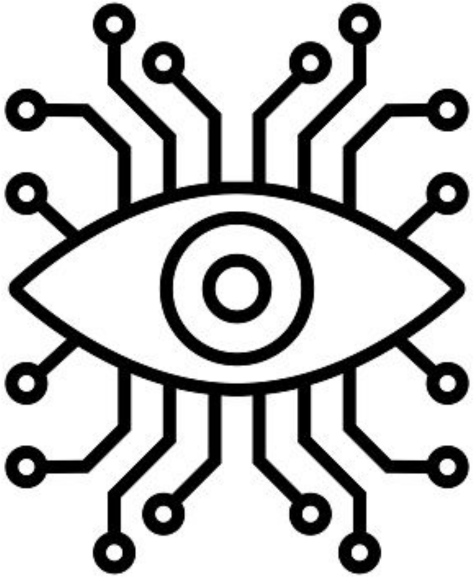
https://www.kaggle.com/competitions/asl-signs

http://facundoq.github.io/guides/sign_language_datasets/slr

https://www.kaggle.com/datasets/datamunge/sign-language-mnist

https://www.youtube.com/watch?v=a99p + https://github.com/kinivi/hand-gesture-recognition-mediapipe

Computer vision and NLP



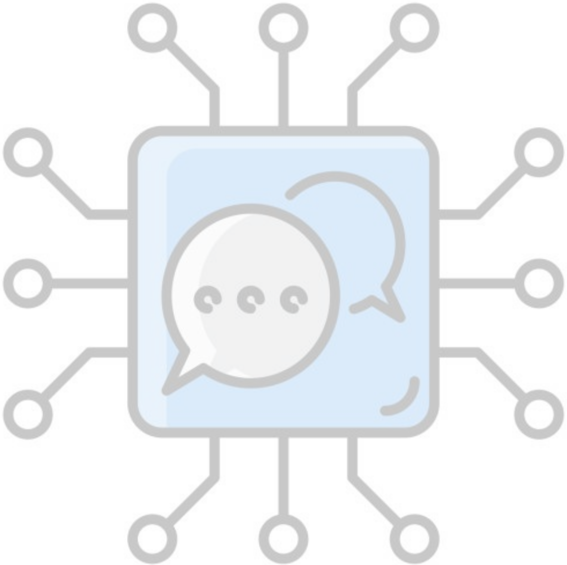
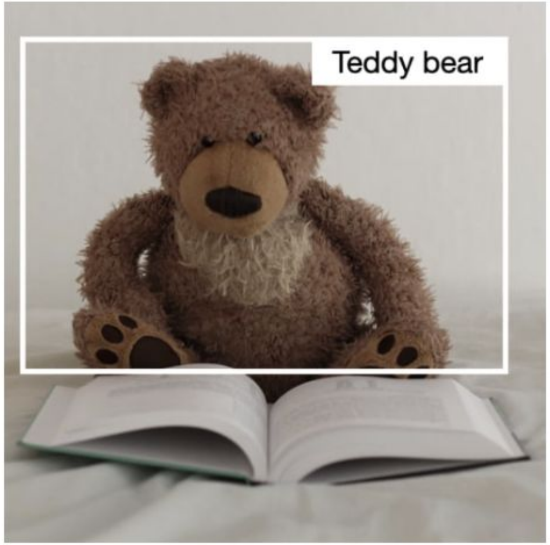
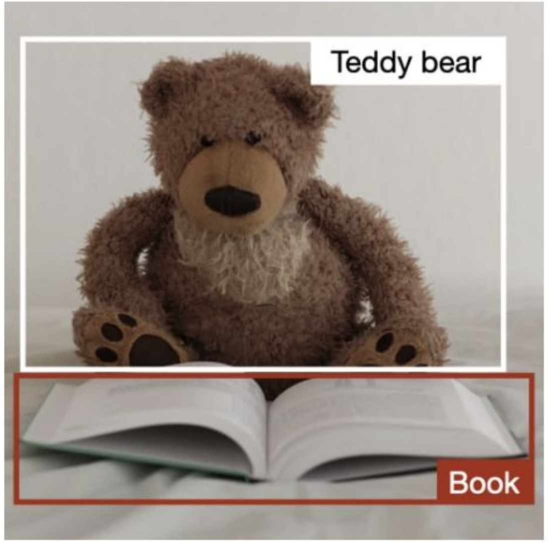
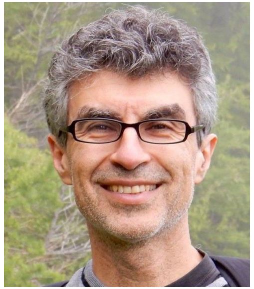


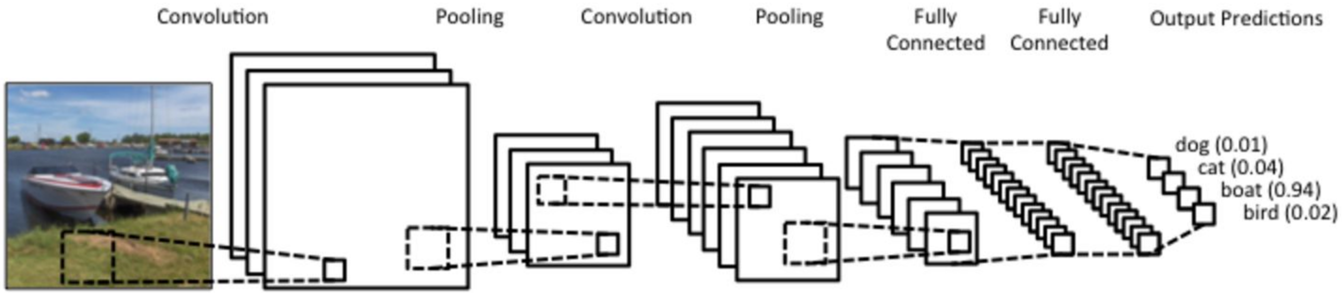
Image classification	Classification w. localization	Detection
		
• Classifies a picture • Predicts probability of object	• Detects an object in a picture • Predicts probability of object and where it is located	• Detects up to several objects in a picture • Predicts probabilities of objects and where they are located
Traditional CNN	Simplified YOLO, R-CNN	YOLO, R-CNN

Convolutional Neural Network

Convolutional networks for images, speech, and time series

Yann LeCun & Yoshua Bengio (1995)

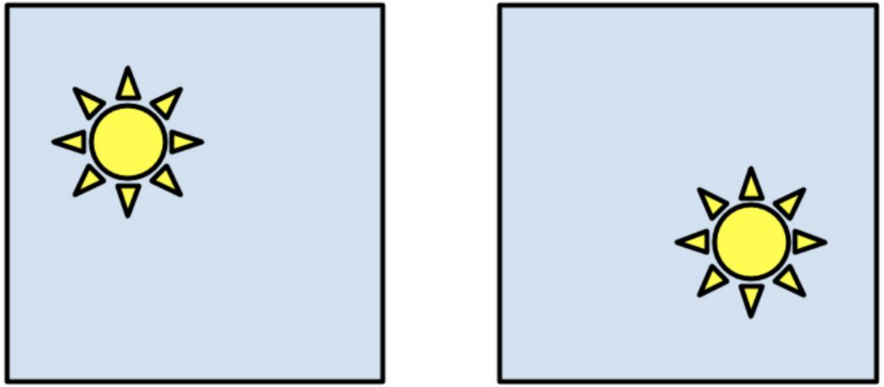




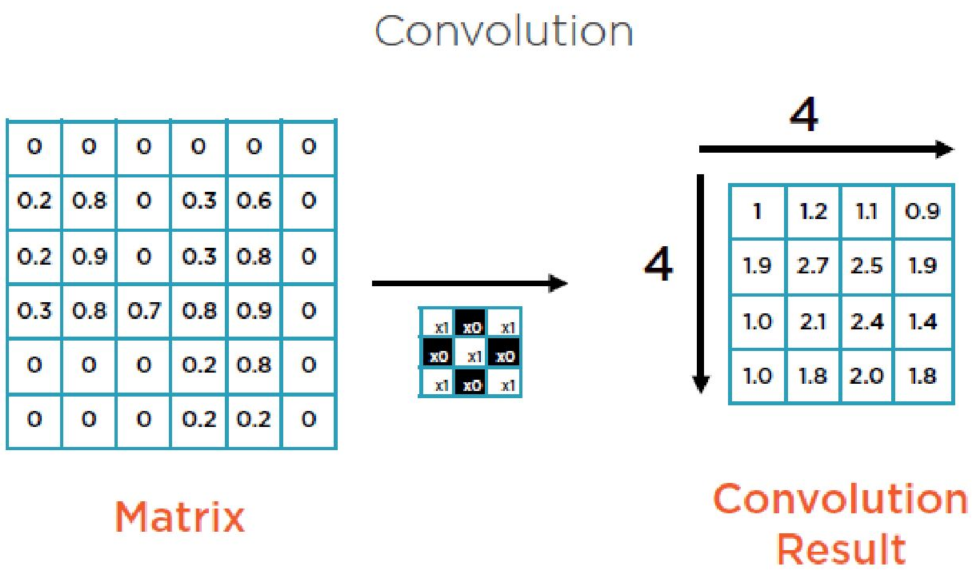
Motivation CNNs

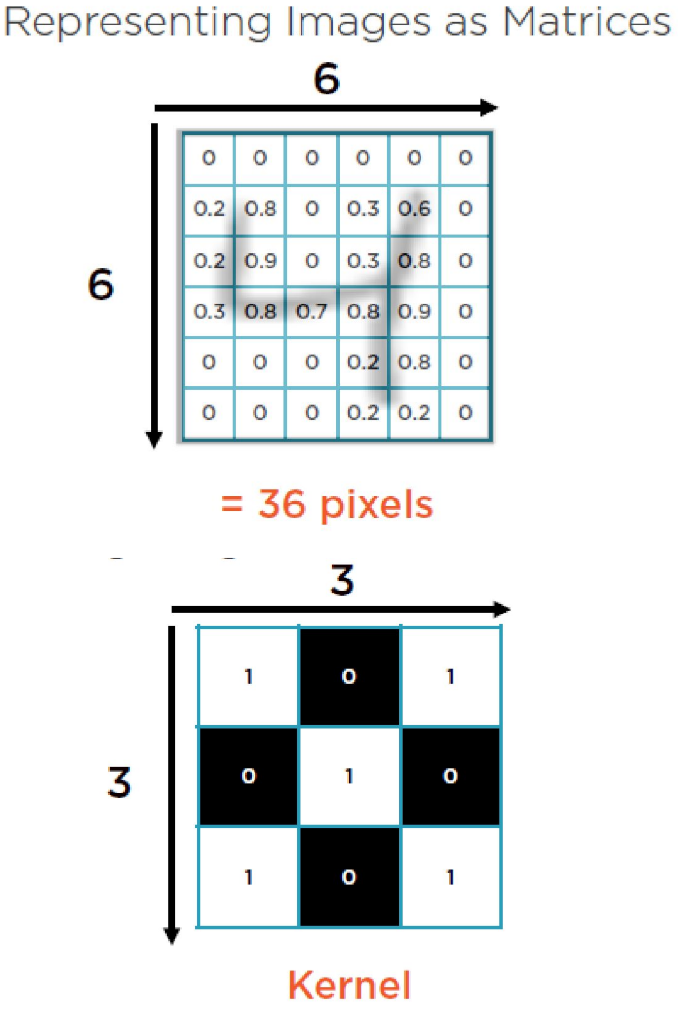
Reason 1: Images are big

Reason 2: Positions can change



What is a convolution?





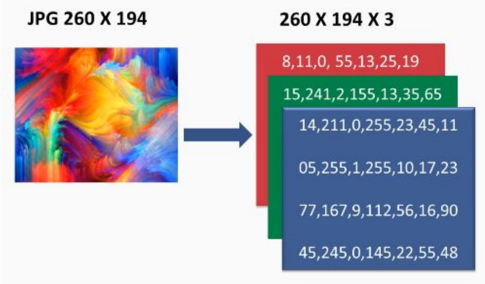
Kernels

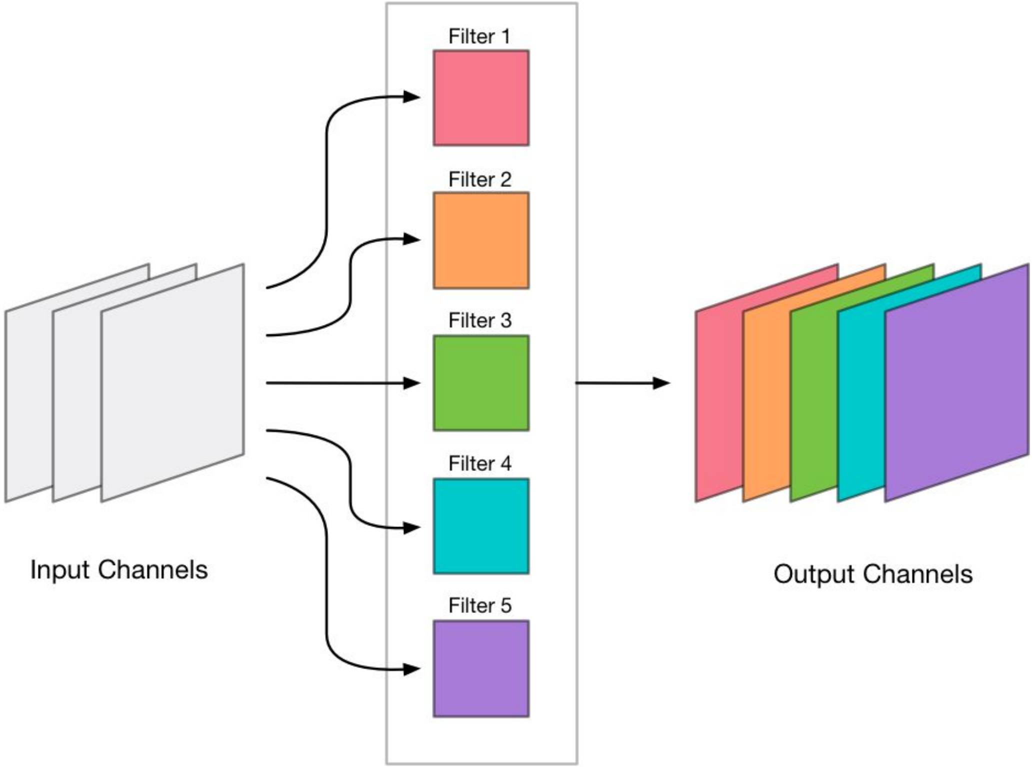
Try out at:
https://setosa.io/ev/image-kernels/

Operation	Kernel ω	Image result g(x,y)
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	

https://en.wikipedia.org/wiki/Kernel_(image_processing)

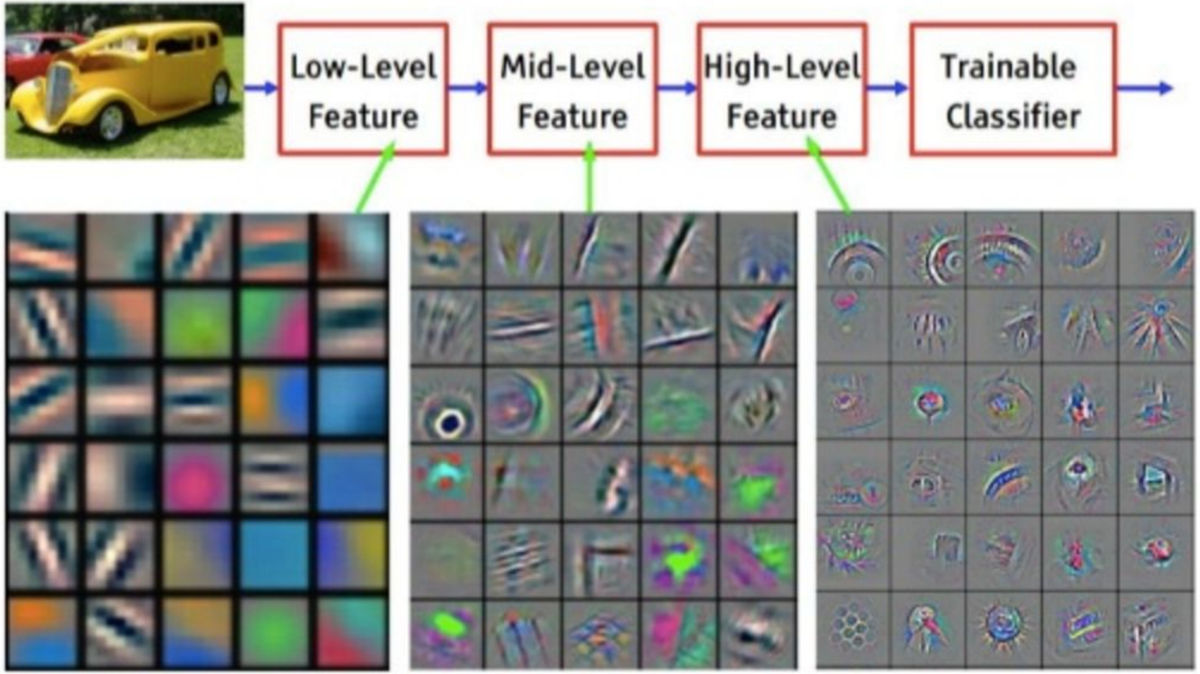
Filters



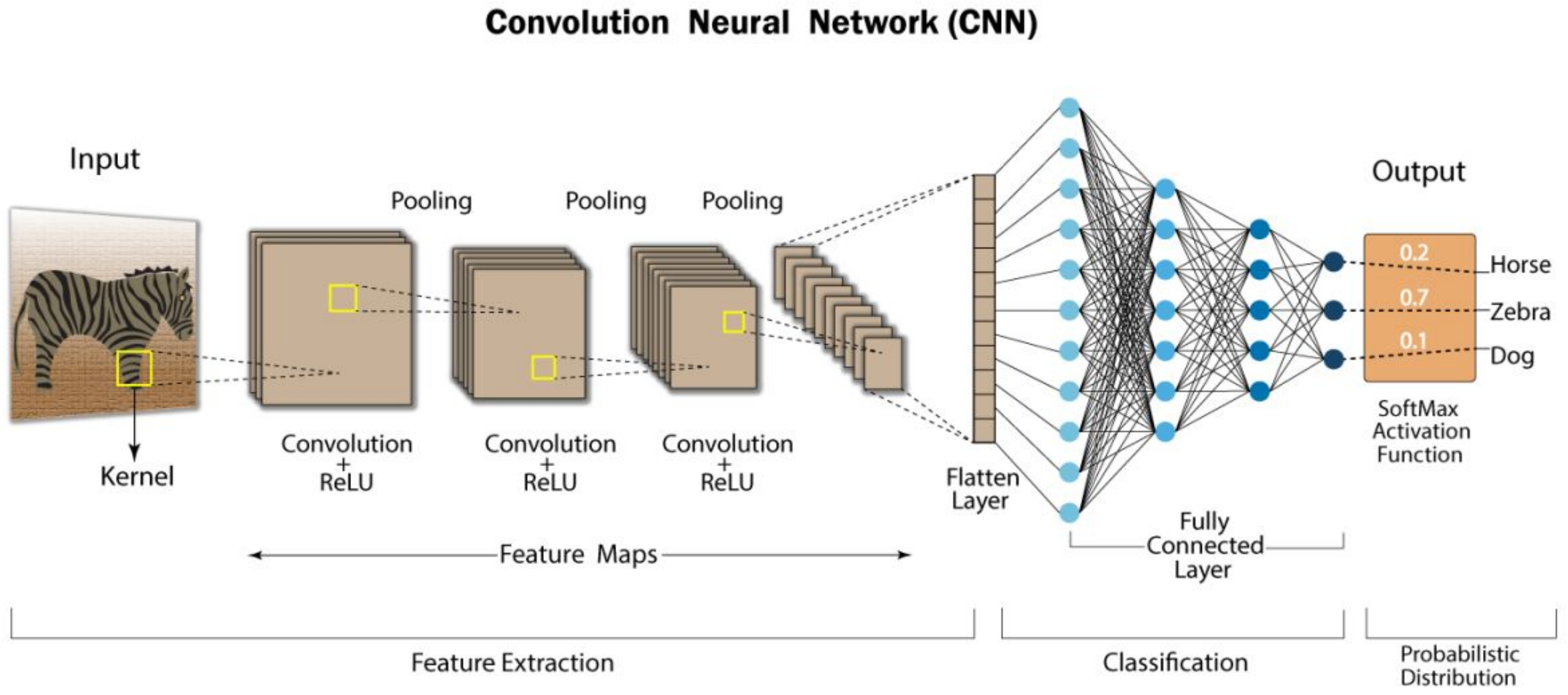


CNN doesn't only learn one filter, it learns multiple filters. In fact, it even learns multiple filters in each layer! Every filter learns a specific pattern, or feature.

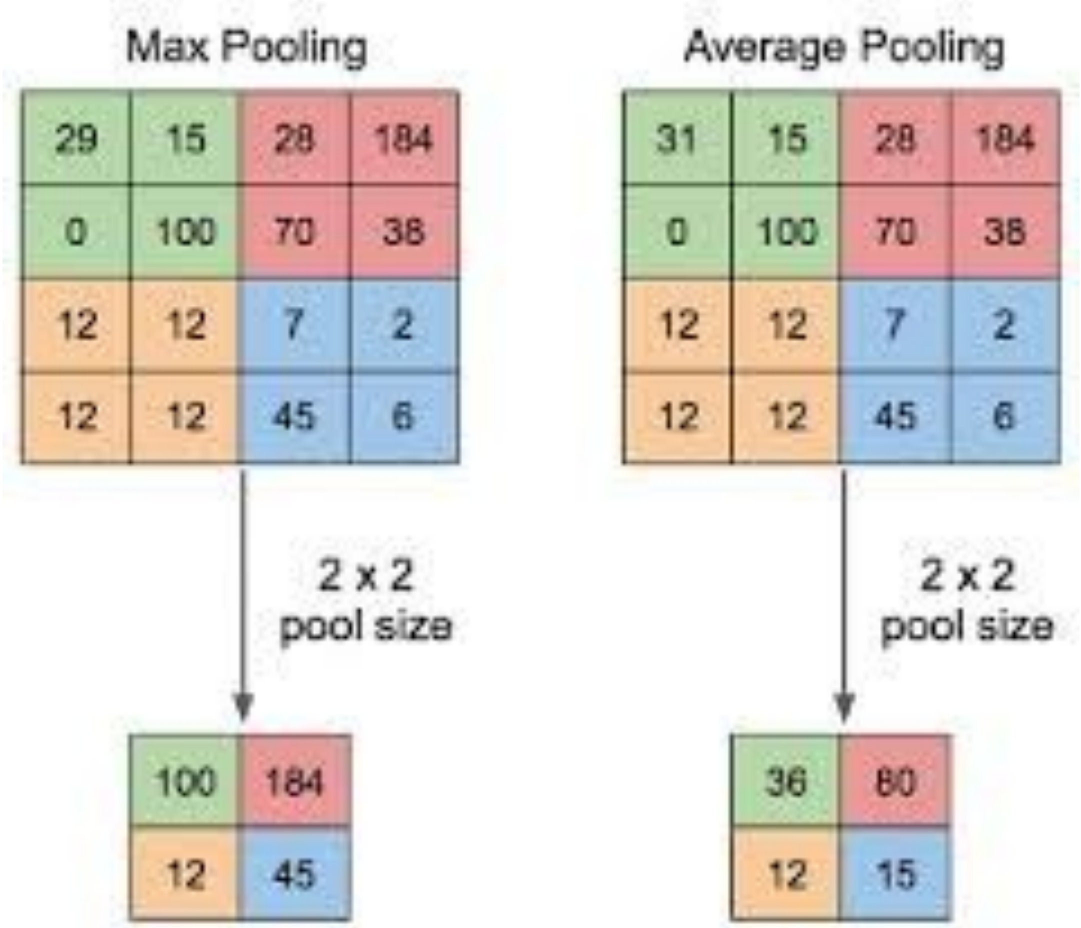
Convolutional Neural Network



Activation functions

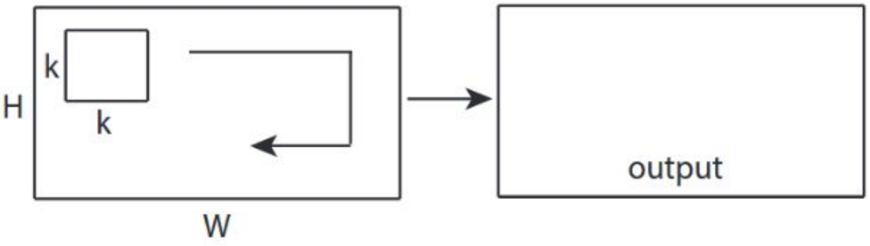


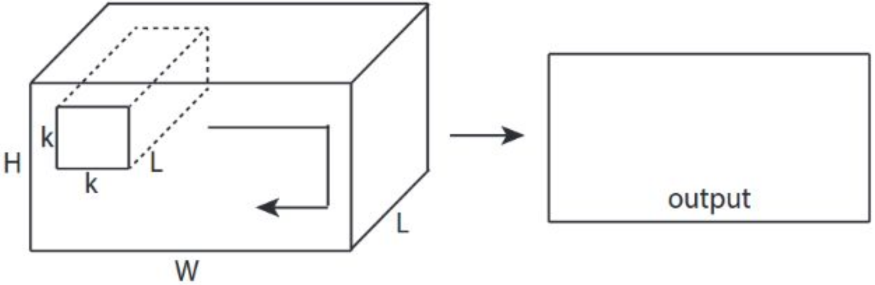
Pooling

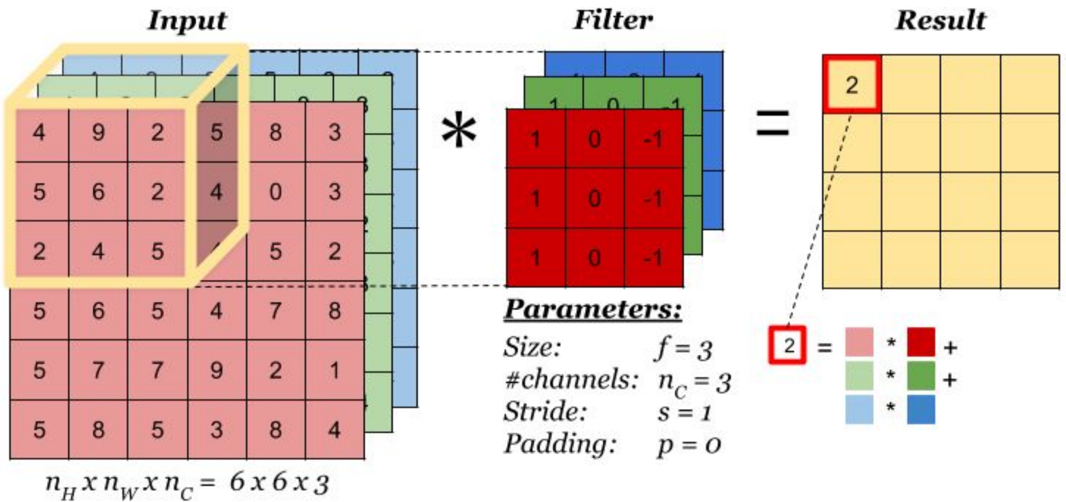


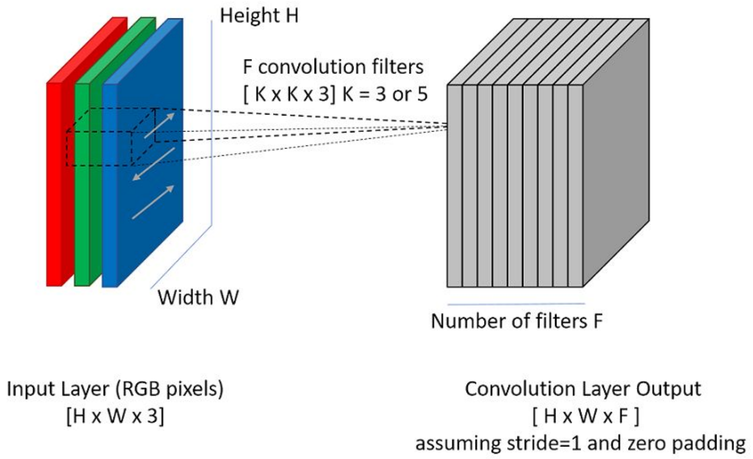
Max Pooling
29
15
28
184
0
100
70
38
12
12
7
2
12
12
45
6
2 x 2
pool size
100
184
12
45
Average Pooling
31
15
28
184
0
100
70
38
12
12
7
2
12
12
45
6
2 x 2
pool size
36
80
12
15

2D Convolution (images)

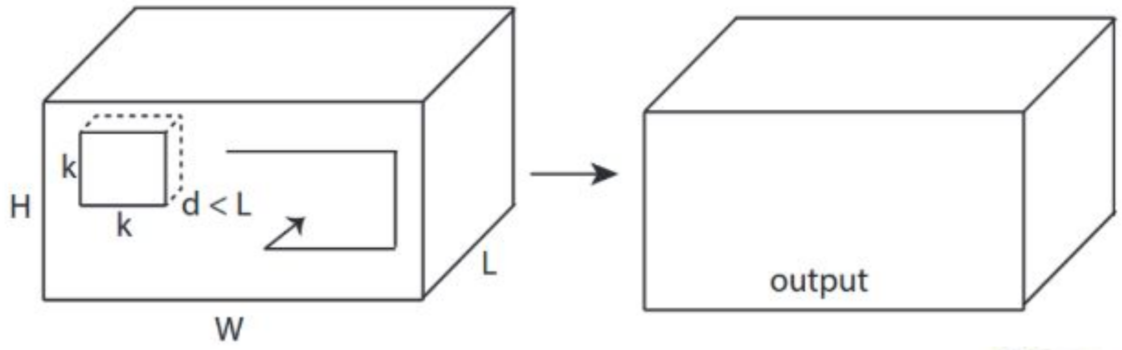


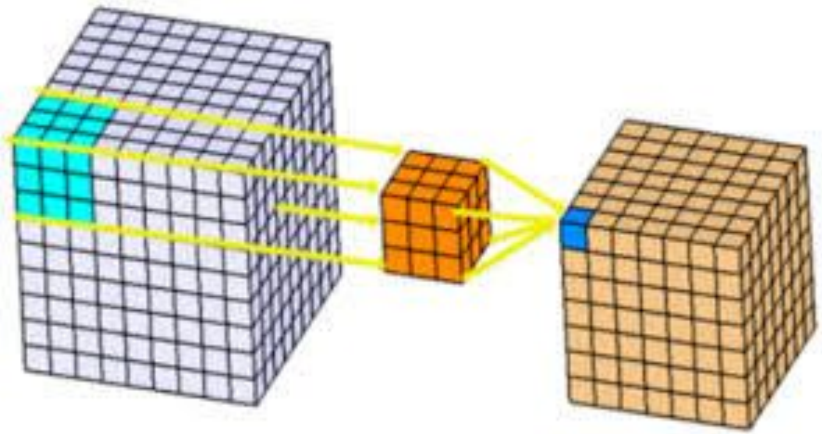




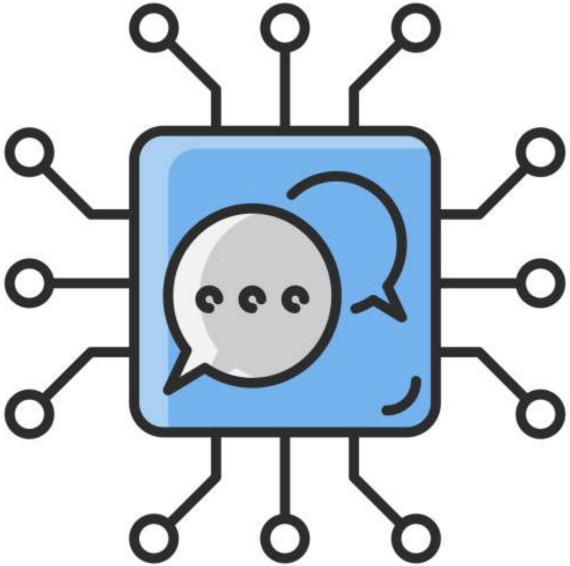


3D Convolutions (e.g. movies)





Computer vision and NLP



Language model

A language model in NLP is a model that computes probability of a sentence(sequence of words) or the probability of a next word in a sequence.

- Sentimental Analysis
- Question-Answering
- Summarization
- Machine Translation

$$P(w_5|w_1, w_2, w_3, w_4)$$

Word embeddings

Vocabulary:
Man, woman, boy, girl, prince, princess, queen, king, monarch

	1	2	3	4	5	6	7	8	9
man	1	0	0	0	0	0	0	0	0
woman	0	1	0	0	0	0	0	0	0
boy	0	0	1	0	0	0	0	0	0
girl	0	0	0	1	0	0	0	0	0
prince	0	0	0	0	1	0	0	0	0
princess	0	0	0	0	0	1	0	0	0
queen	0	0	0	0	0	0	1	0	0
king	0	0	0	0	0	0	0	1	0
monarch	0	0	0	0	0	0	0	0	1

Each word gets a 1x9 vector representation

Try to build a lower dimensional embedding

Vocabulary:
Man, woman, boy, girl, prince, princess, queen, king, monarch

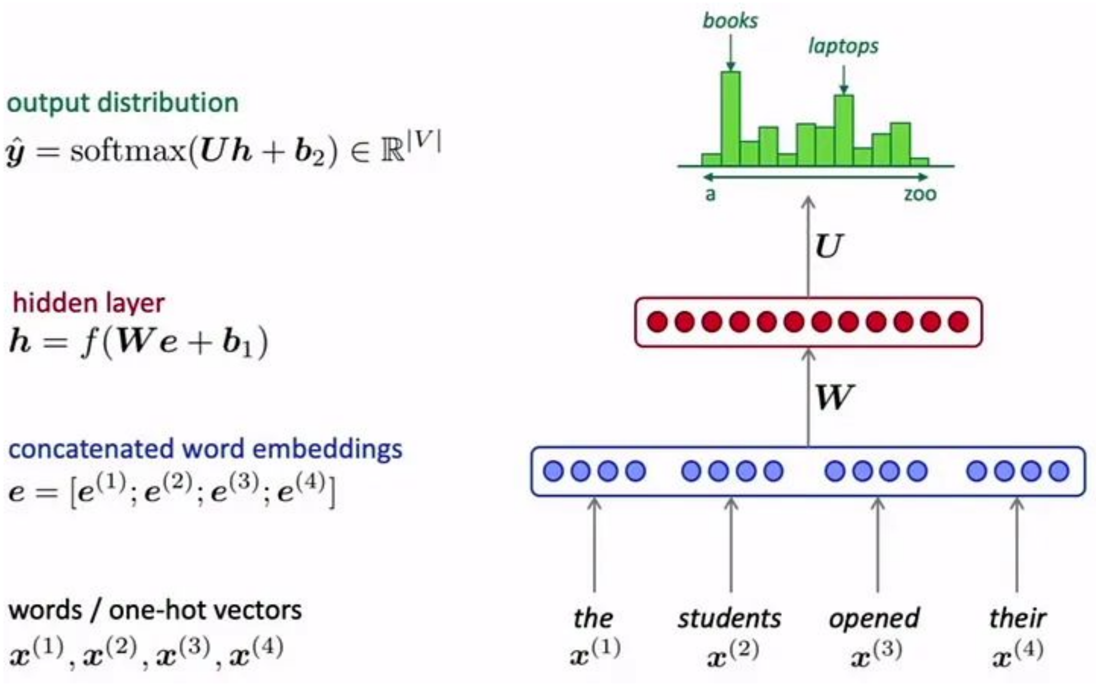
	Femininity	Youth	Royalty
Man	0	0	0
Woman	1	0	0
Boy	0	1	0
Girl	1	1	0
Prince	0	1	1
Princess	1	1	1
Queen	1	0	1
King	0	0	1
Monarch	0.5	0.5	1

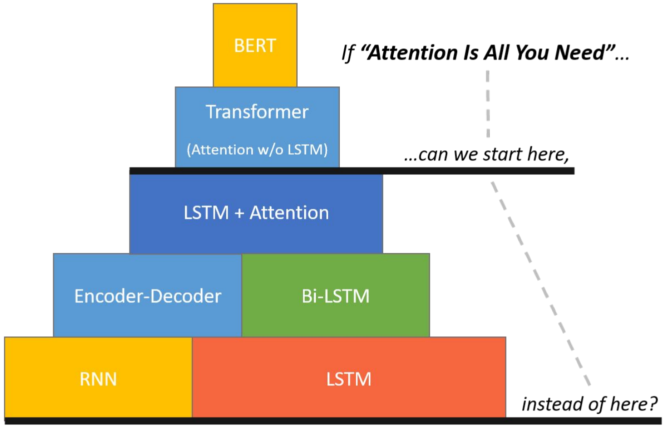
Each word gets a 1x3 vector

Similar words... similar vectors

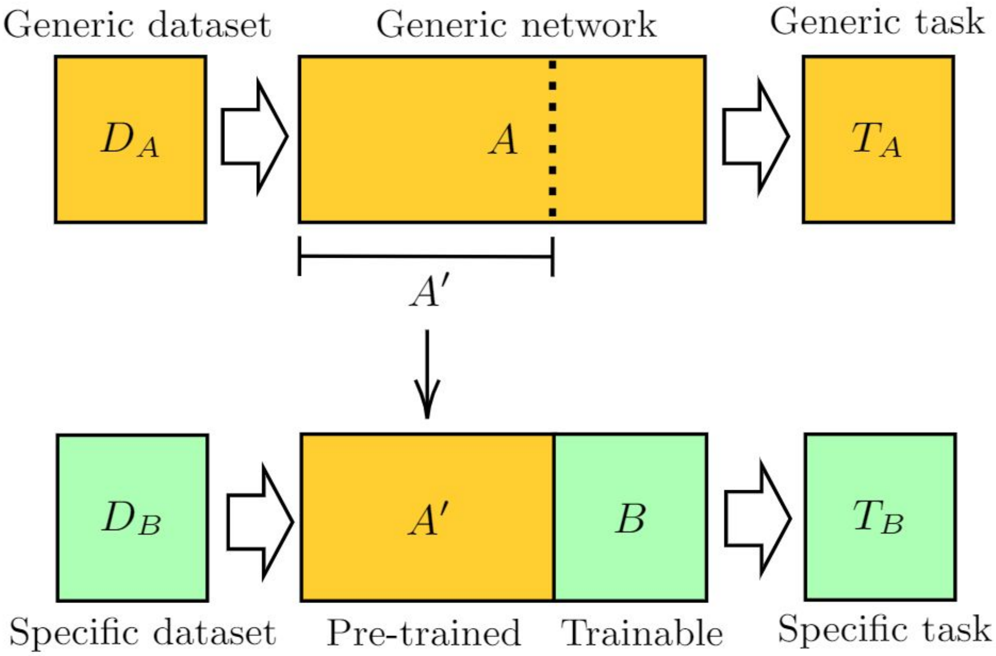
@shane_a_lynn | @TeamEdgeTier

Models for NLP

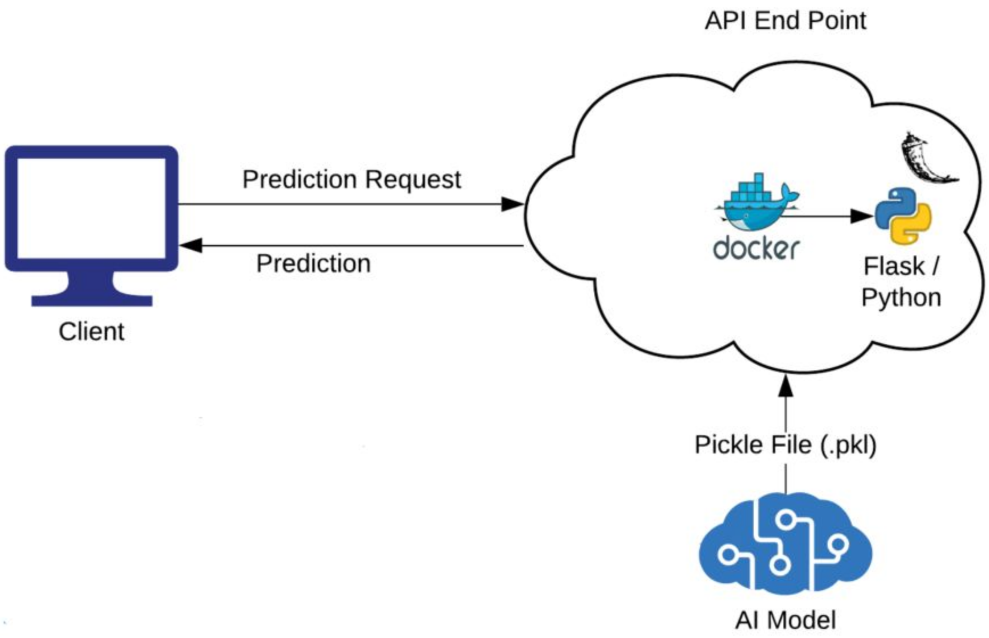




Transfer learning



Deploying


API End Point
Prediction Request
Prediction
Client
docker
Flask /
Python
Pickle File (.pkl)
AI Model


DEMO
Shift

Additional resources

https://wandb.ai/site/experiment-tracking

https://youtube.com/playlist?list=PLam9sigHPGwOBuH4_4fr-XvDbe5uneaf6

https://segment-anything.com/