Neural network - easy coding in Keras

From beginner to master in less then three hours

Startit Center Belgrade,

16.11.2019,

16:00-19:00

Milan M. Čugurović

- Entusiast interested in DSC, especially ML
- Master Mathematician and PhD Candidate in Informatics at Faculty of Mathematics, UB
- TA @ MATF, BG
- Member of ML Group at Faculty of Mathematics, UB
- Member of CS Department at Faculty of Mathematics, UB

Faculty of Mathematics, UB

- matf.bg.ac.rs/eng
- more than 6000 graduate mathematicians
- 400 doctors of mathematics and computer sciences
- 700 Magisters of Science
- many specialists etc.

ML@MATF

- machinelearning.math.rs
- Based at Faculty of Mathematics
- Includes researchers, students and partitioners from various research institutions and leading IT companies
- Led by professor dr Mladen Nikolić
- Lectures/talks and practical sessions are held on every 2 weeks

Schedule:

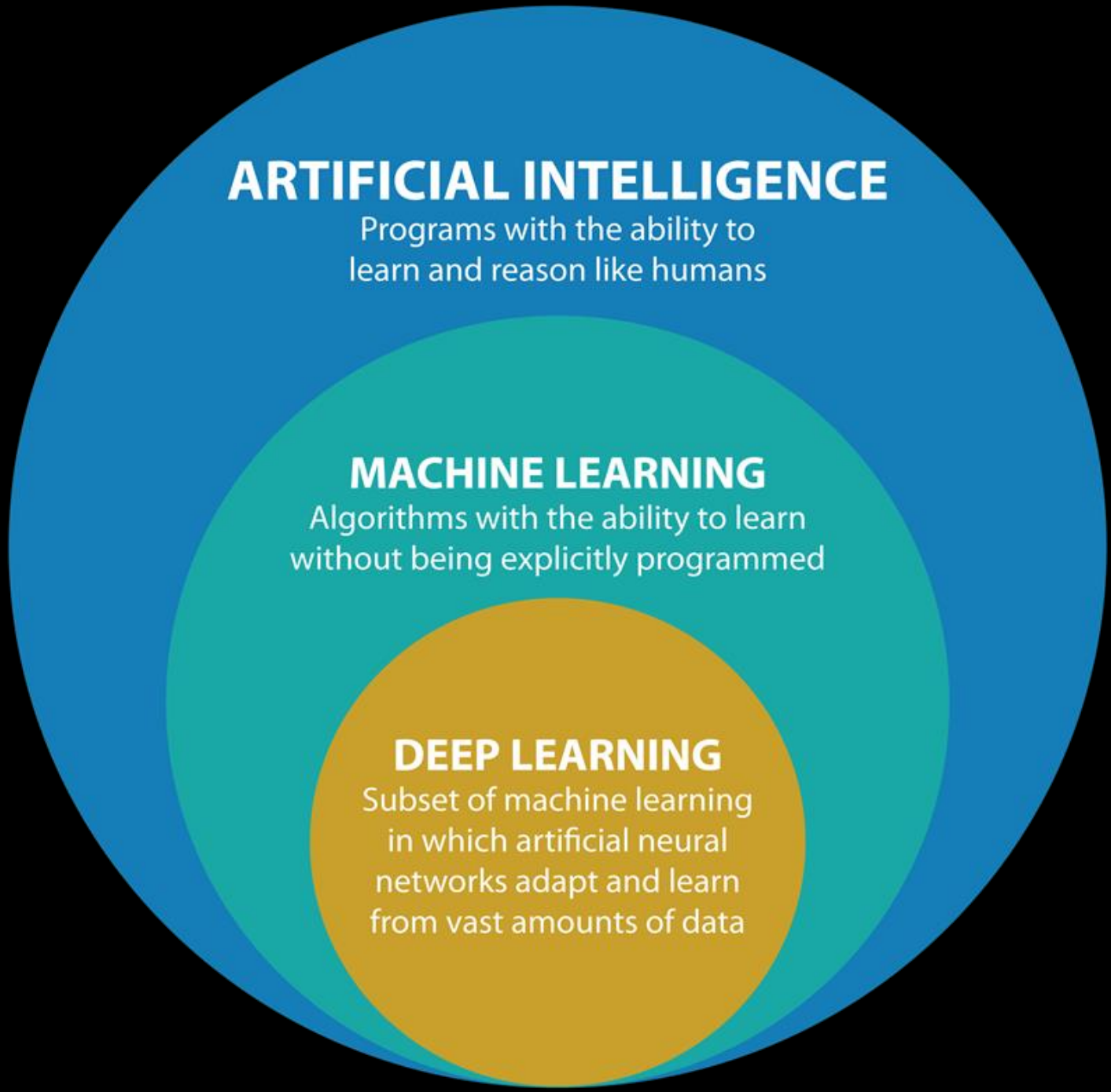
- Part I: ML basics / To fall in love with ML
- Part II: NN basics / To take part in any ML project/community
- Part III: Keras library / Learn basics of Keras

Lets meet these three goals and I will be pleased...

ML basics

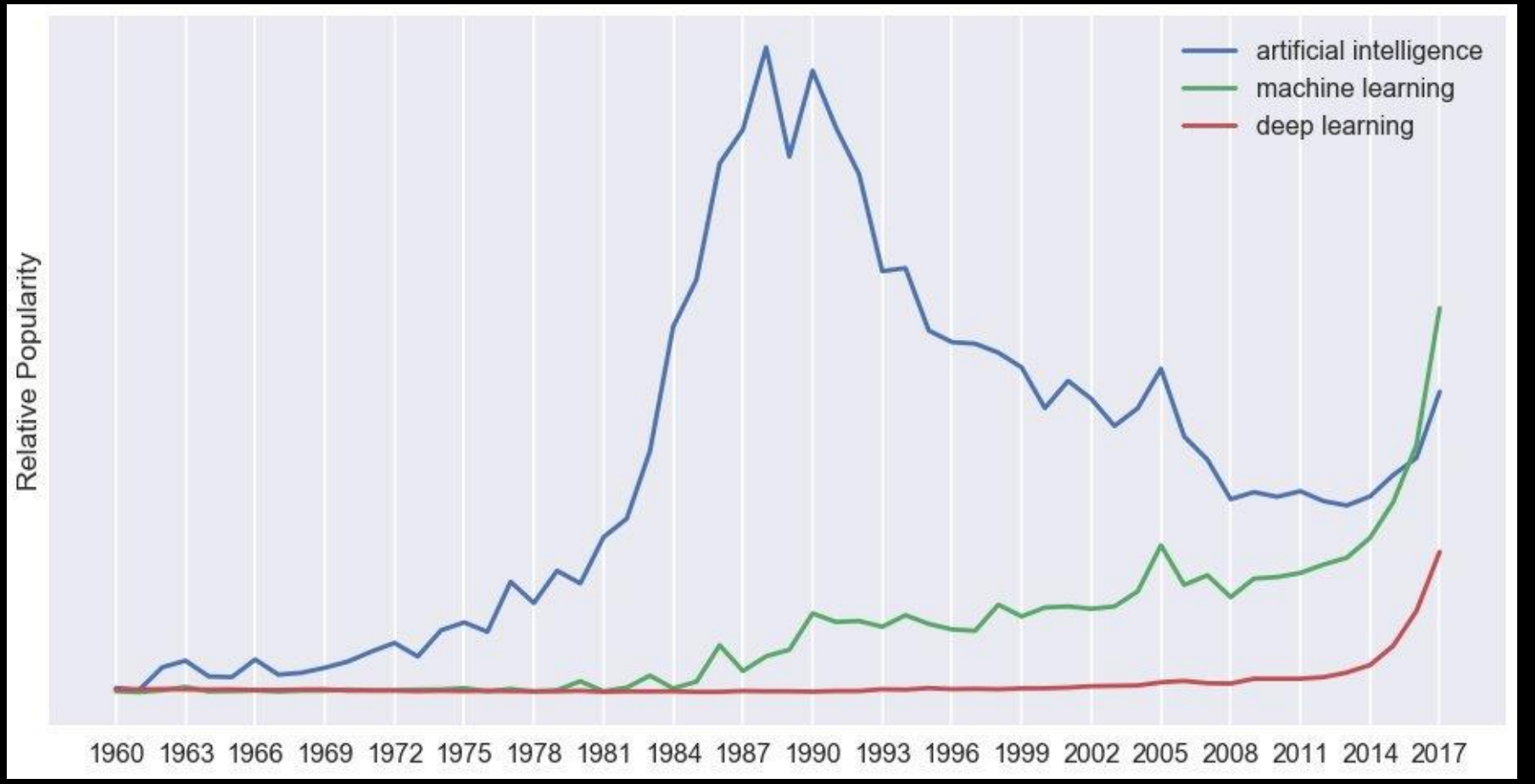
ML basics

- Hottest CS topic currently
- DL(NN) ⊂ ML ⊂ AI
- Very active research field today
- Has accomplished amazing results
- Built on multiple mathematical disciplines



A little bit of Rita is all I need…*

- Google trends
- 



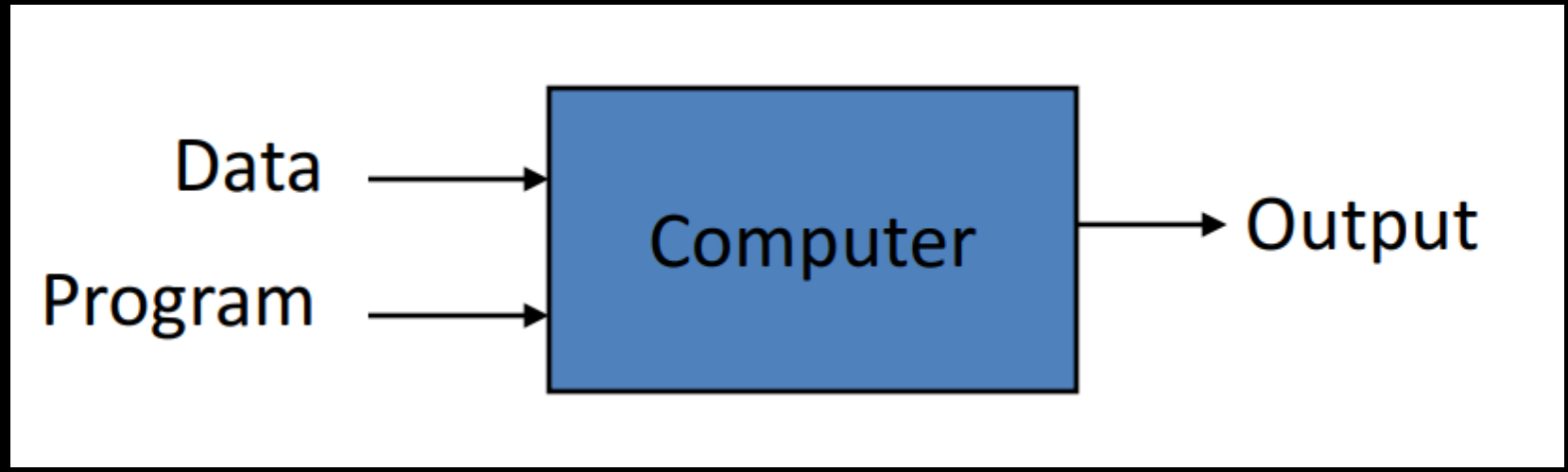
Definition of the term “learning”

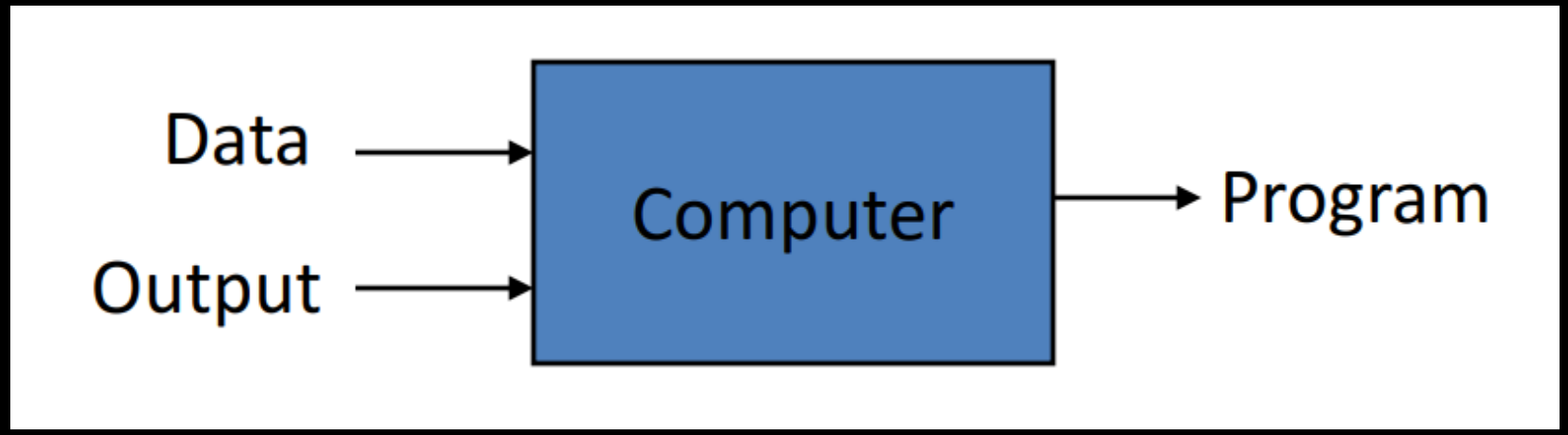
- How would you define the term “learning” ?

Definition of the term “learning”

- Arthur Samuel (1959):
  - “Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed”
- Herbert Simon (1970):
  - “Learning is any process by which a system improves performance from experience.”
- Tom M. Mitchell (1998):
  - Machine Learning is the study of algorithms that:
    - improve their performance P
    - at some task T
    - with experience E
  - A well-defined learning task is given by <P, T, E>

ML vs Traditional Programming





Focus on me,
Focus on me…*

- Google: processes 24 peta bytes of data per day.
- Twitter: 400 million tweets per day.
- Facebook: 10 million photos uploaded every hour.
- YouTube: 1 hour of video uploaded every second.
- Astronomy: Satellite data is in hundreds of PB.
- “By 2020 the digital universe will reach 44 zettabytes…” (44 trillion gigabytes)
  - The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things, April 2014

Some ML results

- 1992 - TD-Gammon, computer program developed by Gerald Tesauro able to play backgammon
- 2011 - IBM's Watson wins in quiz Jeopardy!
- 2012 - Google X creates system able to recognize cats on video recordings
- 2015 - Classification error for images reduced to 3.6% (5-10% is the error made by humans)
- 2016 - Google creates AlphaGo, agent able to play Go who beats the world champion 4:1
- 2017 - AlphaGo plays against its 2016 version and wins 100/100 games
- October 2017 - AlphaGo Zero learnt to play the game of Go simply by playing games against itself, starting from scratch (40 days)
- 2018 - Humans are miserable compared to ML (any area)

But why is it so successful and popular today?

- There is serious amount of mathematics behind [Murphy, 2012, Bishop, 2006, Hastie et al., 2001, Shalev-Shwartz and Ben-David, 2014, Vapnik, 1995]

$$P(\sup_{f\in\mathcal{F}}(R(f)-E(f))>\varepsilon)\leq$$

$$2P(\sup_{f\in\mathcal{F}}(E'(f)-E(f))>\varepsilon/2)=$$

$$2P\left(\sup_{l\in\mathcal{L}_{z_1,\ldots,z_N,z'_1,\ldots,z'_N}}\left(\frac{1}{N}\sum_{i=N+1}^{2N}l_i-\frac{1}{N}\sum_{i=1}^{N}l_i\right)>\varepsilon/2\right)\leq$$

$$2\sum_{l\in\mathcal{L}_{z_1,\ldots,z_N,z'_1,\ldots,z'_N}}P\left(\frac{1}{N}\sum_{i=N+1}^{2N}l_i-\frac{1}{N}\sum_{i=1}^{N}l_i>\varepsilon/2\right)=$$

$$2\sum_{l\in\mathcal{L}_{z_1,\ldots,z_N,z'_1,\ldots,z'_N}}P\left(\frac{1}{N}\sum_{i=N+1}^{2N}l_i-R(f)+R(f)-\frac{1}{N}\sum_{i=1}^{N}l_i>\varepsilon/2\right)\leq$$

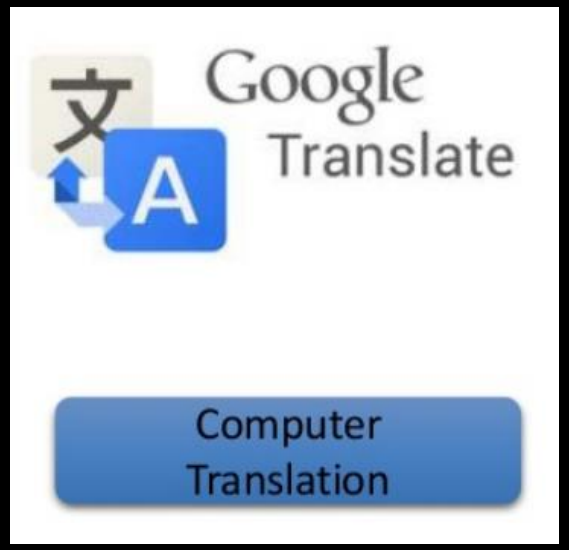
$$2\sum_{l\in\mathcal{L}_{z_1,\ldots,z_N,z'_1,\ldots,z'_N}}\left(P\left(\frac{1}{N}\sum_{i=N+1}^{2N}l_i-R(f)>\varepsilon/2\right)+P\left(R(f)-\frac{1}{N}\sum_{i=1}^{N}l_i>\varepsilon/2\right)\right)\leq$$

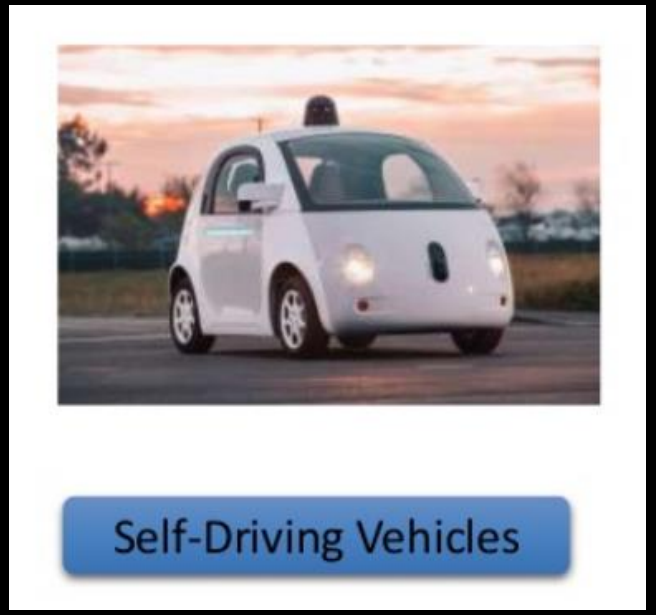
$$2\sum_{l\in\mathcal{L}_{z_1,\ldots,z_N,z'_1,\ldots,z'_N}}(\exp(-N\varepsilon^2/2)+\exp(-N\varepsilon^2/2))=$$

$$4\sum_{l\in\mathcal{L}_{z_1,\ldots,z_N,z'_1,\ldots,z'_N}}\exp(-N\varepsilon^2/2)=$$

Some ML Applications

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- Robotics
- Information extraction
- Social networks
- Debugging software
- [Your favorite area]

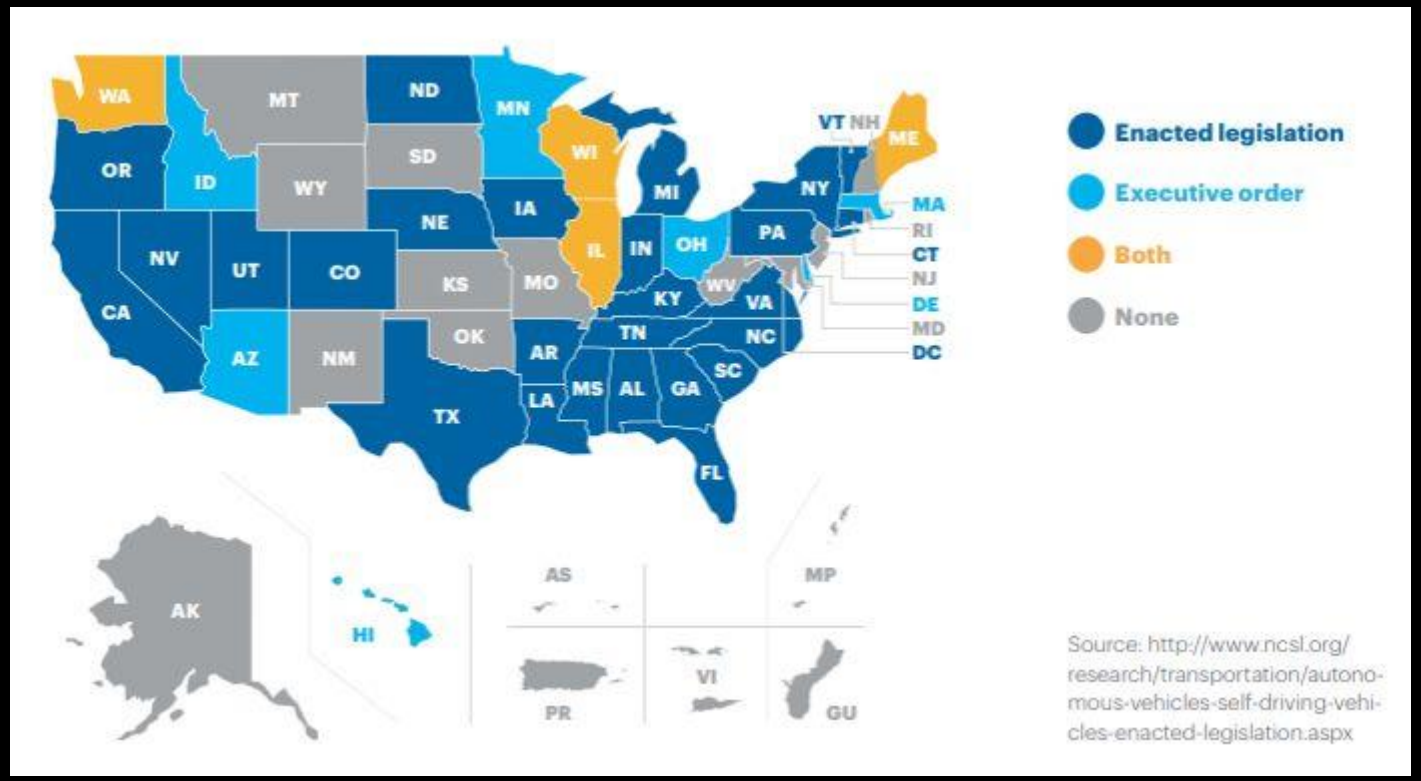




Example 1: Autonomous Cars
youtube.com/watch?v=NLzakmRgEAs (Gyratory traffic)

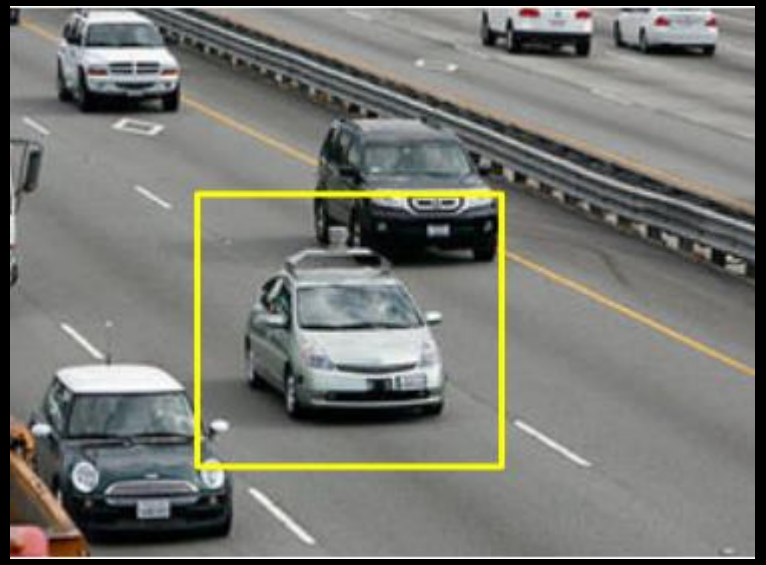
Example 1: Autonomous Cars

- June 2011: Nevada made it legal for autonomous cars to drive on roads
- 2013: Nevada, Florida, California, and Michigan have legalized autonomous cars
- Now:

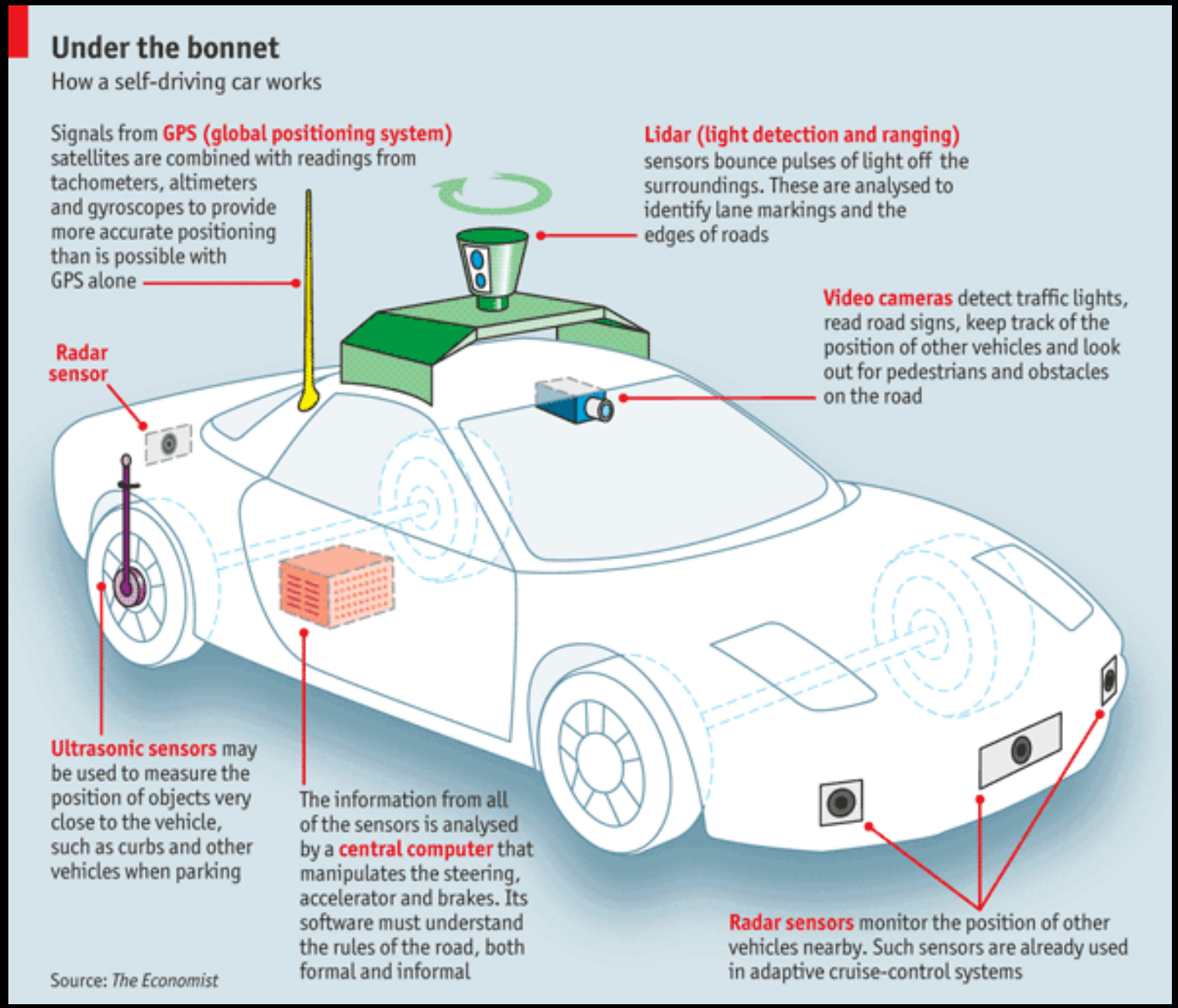


Example 1: Autonomous Cars

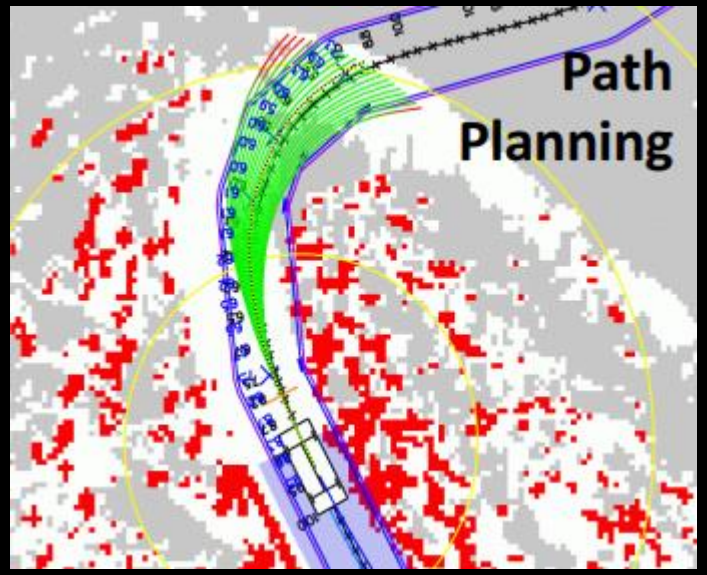
- Detecting it
- Penn's Autonomous Car ((Ben Franklin Racing Team)

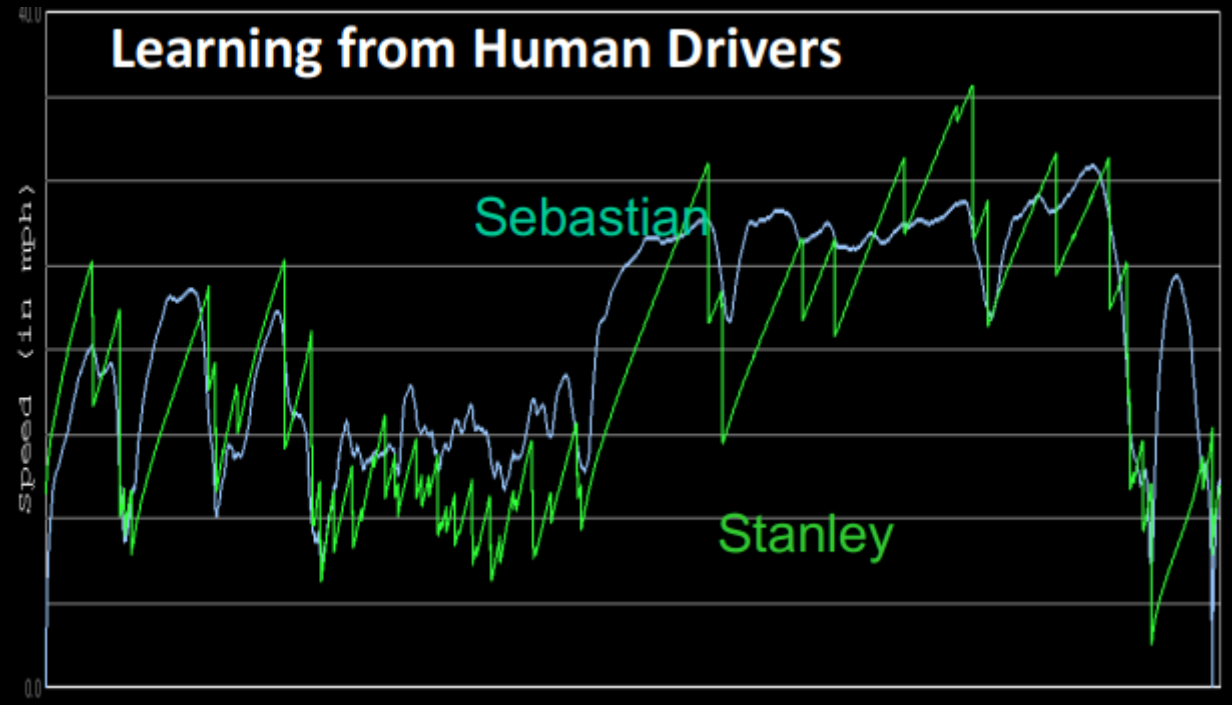


Example 1: Autonomous Cars




Laser Terrain Mapping


Path
Planning


Learning from Human Drivers
Sebastian
Stanley
Speeds (in mph)
40.0
0.0


Adaptive Vision

Example 1: Autonomous Cars

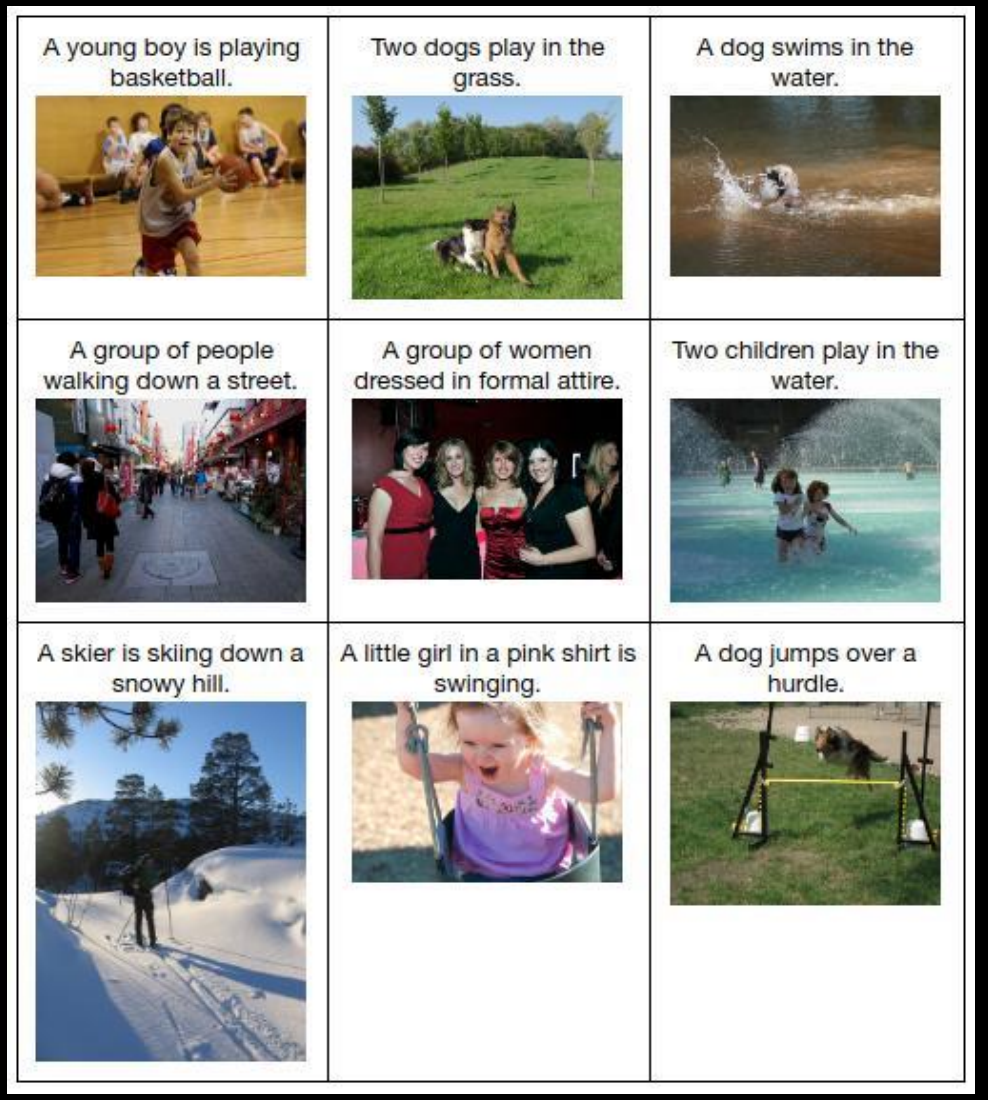
- Tesla: https://www.youtube.com/watch?v=tlThdr3O5Qo
- Wayve: https://www.youtube.com/watch?v=26Or4QbLbMM
- Audi A8 Parking: https://www.youtube.com/watch?v=Cwcoi-HlZ-M

Example 2: Scene Parsing/Labeling via DL

- Background/Motivation?
- Scene parsing: labeling each pixel in the image with category of the object to which it belongs
- Important step toward image understanding

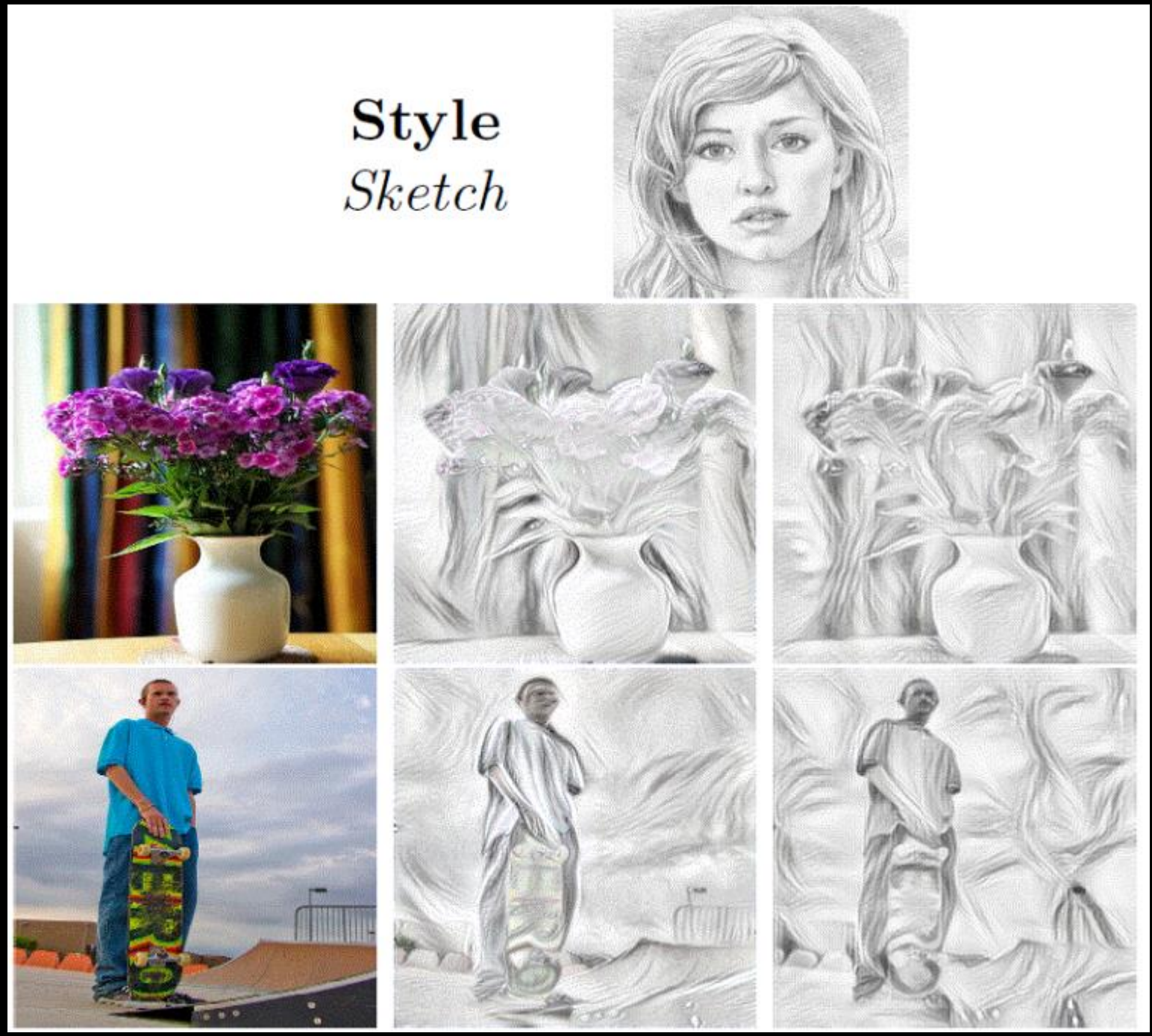
Example 2: Scene Parsing/Labeling via DL

- Image -> Video
  - https://www.youtube.com/watch?v=N1f7i-oItWU
- To Train DL Models on Satellite Imagery
- Image captioning



Example 3: Style Transfer

- My lecture at ML@MATF Seminary:
  - http://machinelearning.math.rs/Cugurovic-Style.pdf



Example 3: Style Transfer

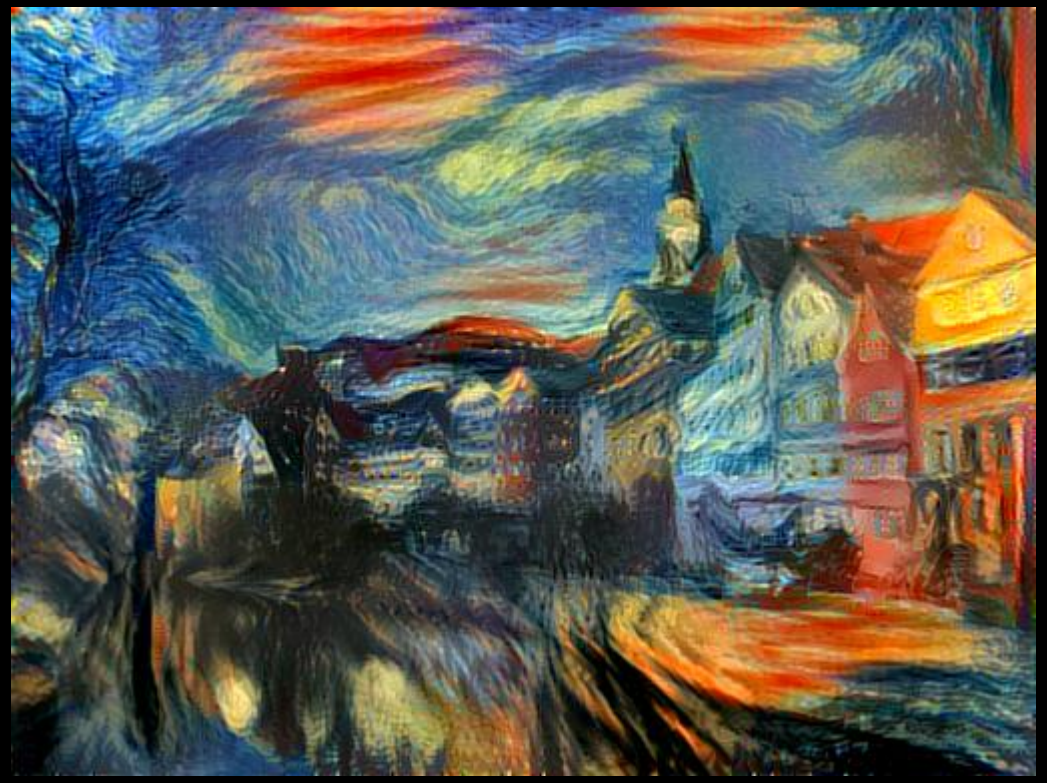
- ***The Starry Night, 1889***
- Oil on canvas
- By the Dutch post-impressionist painter Vincent van Gogh
- The Stanford campus photo

Example 3: Style Transfer

- ***The Scream, 1893***
- A composition created by Norwegian Expressionist artist
- Edvard Munch
- Tubingen, Germany

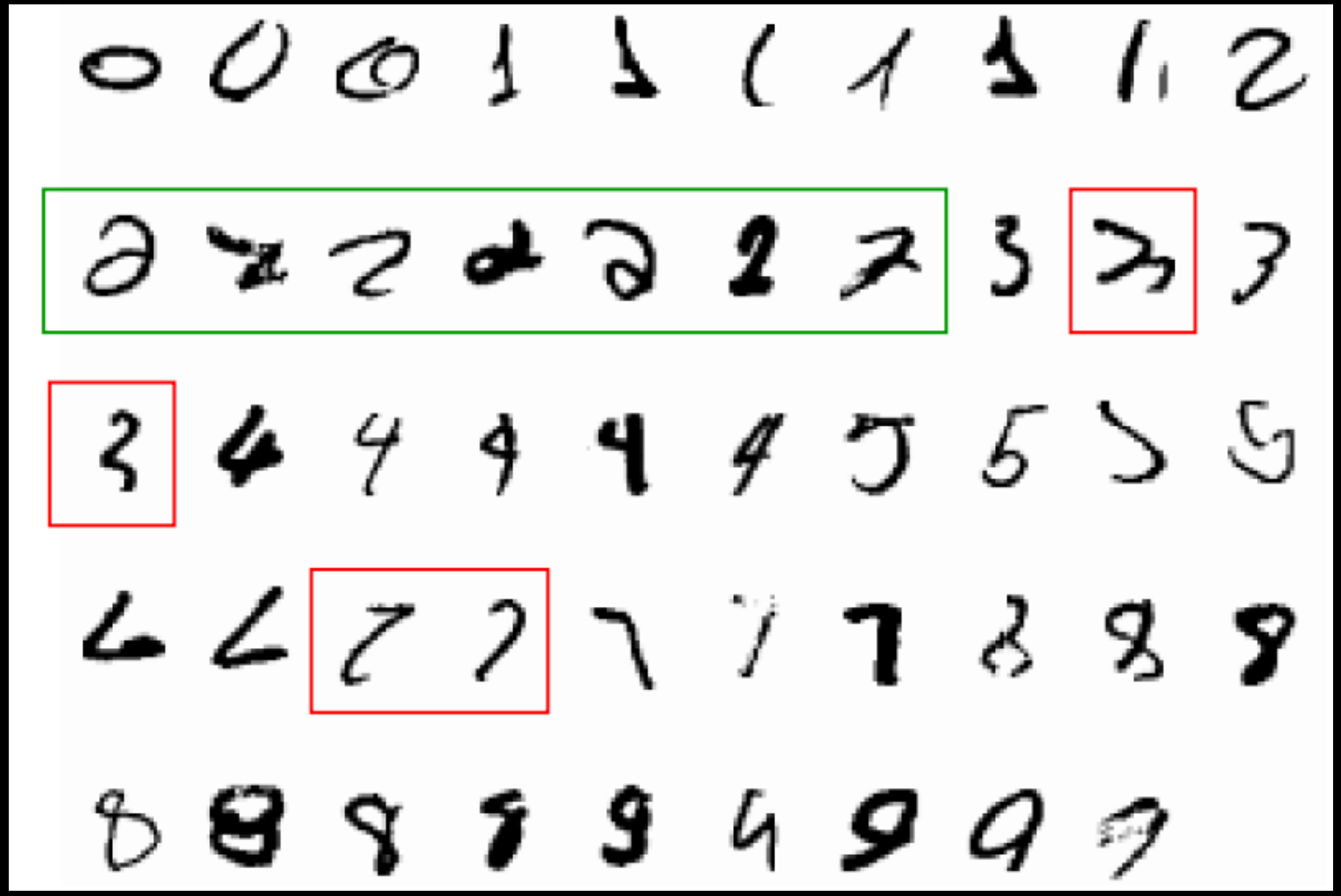
Example 3: Style Transfer

- Combine them
- The style of Picasso's 1907 self-portrait onto Brad Pitt, etc.



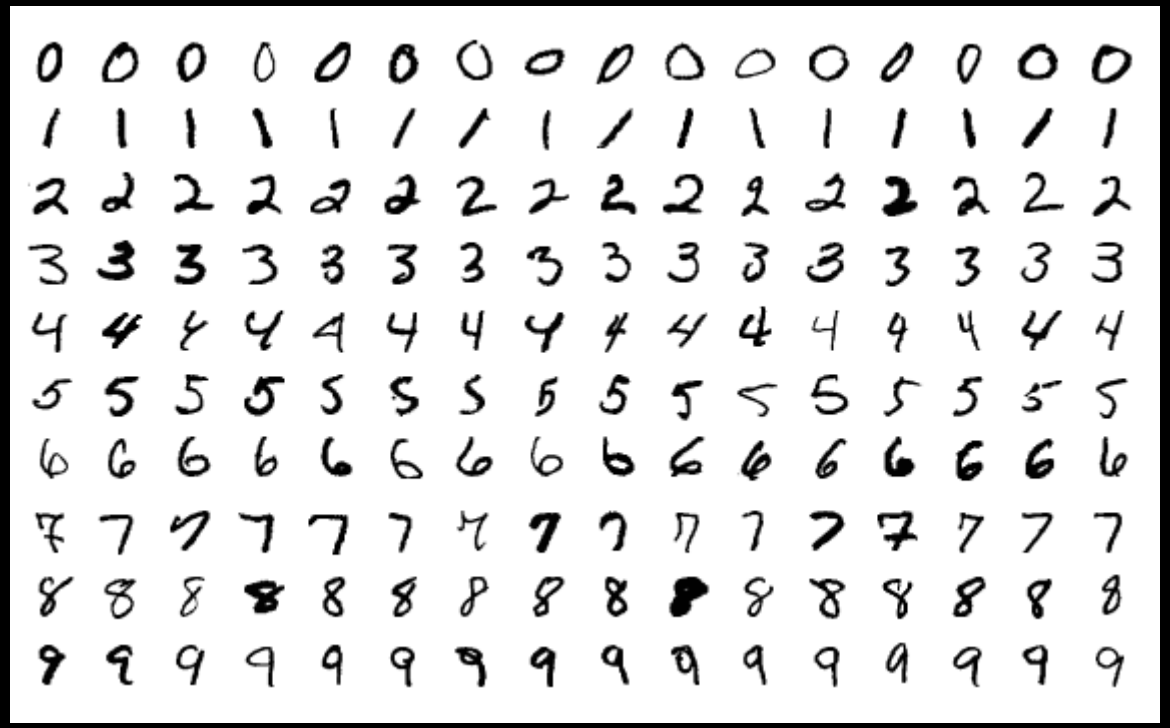
Example 4: Handwriting Recognition

- It is very hard to say what makes a 2
- There were “manual” methods for recognizing handwritten characters before
- ML (DNN) a few orders of magnitude better



Example 4: Handwriting Recognition

- We currently working on it
- Try to further improve deep CNN



Learning the Way We Write: A Lightweight Automatic Adaptation of Neural Classifier of Handwritten Text to Individual Users

Milan M. Čugurović
Department of Computer Science
Faculty of Mathematics
University of Belgrade
Belgrade, 11000
Serbia
milan_cugurovic@math.rs

Mladen Nikolić
Department of Computer Science
Faculty of Mathematics
University of Belgrade
Belgrade, 11000
Serbia
nikolic@math.rs

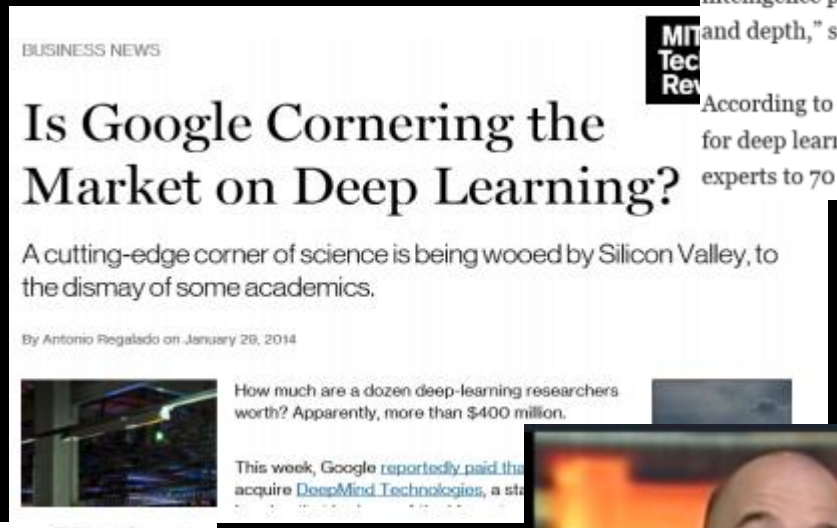
Novak Novaković
Microsoft Development Center Serbia
Belgrade, 11000
Serbia
novakn@microsoft.com

Abstract

NN basics

NN == DL

- Can see it everywhere
- The idea of mimicking the human brain



Neural Networks - First Look

- Most popular and most appropriate machine learning methods
- Five main types:
  - Feed Forward neural network (basic)
  - Convolutional neural network (images)
  - Reccurent neural network (memory)
  - Recursive neural network (tree)
  - Graph neural network (graph)
- As we said before:
  - Medical diagnostic
  - Image recognition and Object detection
  - Autonomous driving
  - NLP (Natural Language Processing)
  - Go, FlappyBird …
- The reason for their re-popularity (computational power – GPU!)

Before we dive in...

- Biological motivation and connection
  - Neuron - the basic computational unit of the brain
  - Around 86 billion neurons can be found in the human nervous system
  - They are connected with approximately $10^{14} - 10^{15}$ synapses
  - Human 'training process' - a couple of years :) (AlphaGo 40 days right now)
- Mathematical strength
  - Universal approximation theorem:

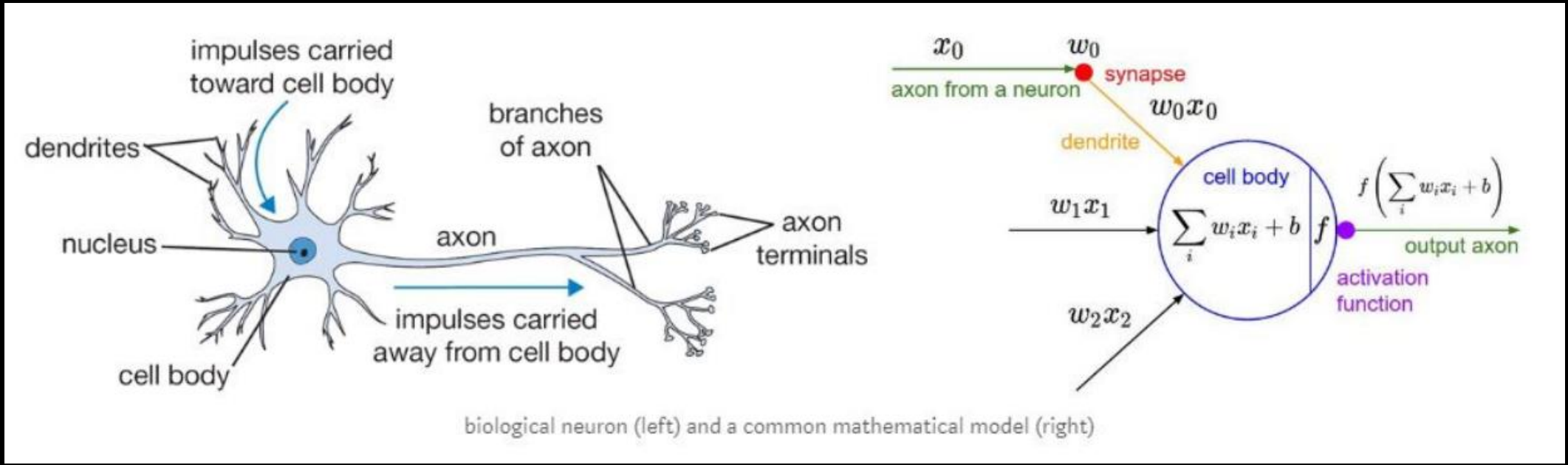
Let $\phi(\cdot)$ be a nonconstant, bounded, and monotonically-increasing continuous function. Let I_m denote he m-dimensional unit hypercube $[0,1]^m$.The space of continuous functions on I_m is denoted by $C(I_m)$. Then, given any $\epsilon > 0$ and any function $f \in C(I_m)$, there exists an integer N, real constants v_i, b_i in and real vectors $w_i \in {}^m$, where $i = 1, 2, ..., N$, such that we may define:

$$F(x) = \sum_{i=1}^{N} v_i \phi(w_i^T x + b_i)$$

as an approximate realization of the function f, where f is independent of ϕ; That is $|F(x) - f(x)| < \epsilon$, for all $x \in I_m$.

Main parts of Feed Forward Neural Network

- Neuron:



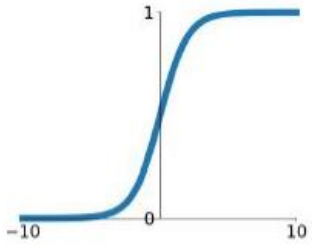
biological neuron (left) and a common mathematical model (right)

Main parts of Feed Forward Neural Network

- Activation function:

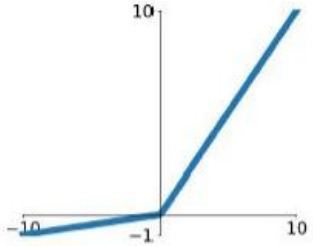
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



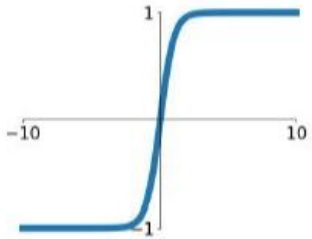
Leaky ReLU

$$\max(0.1x, x)$$



tanh

$$\tanh(x)$$

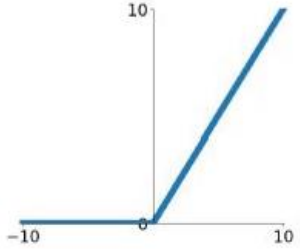


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

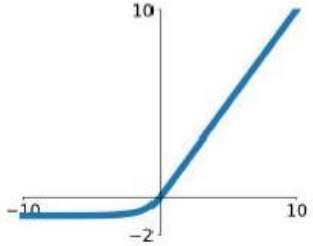
ReLU

$$\max(0, x)$$



ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



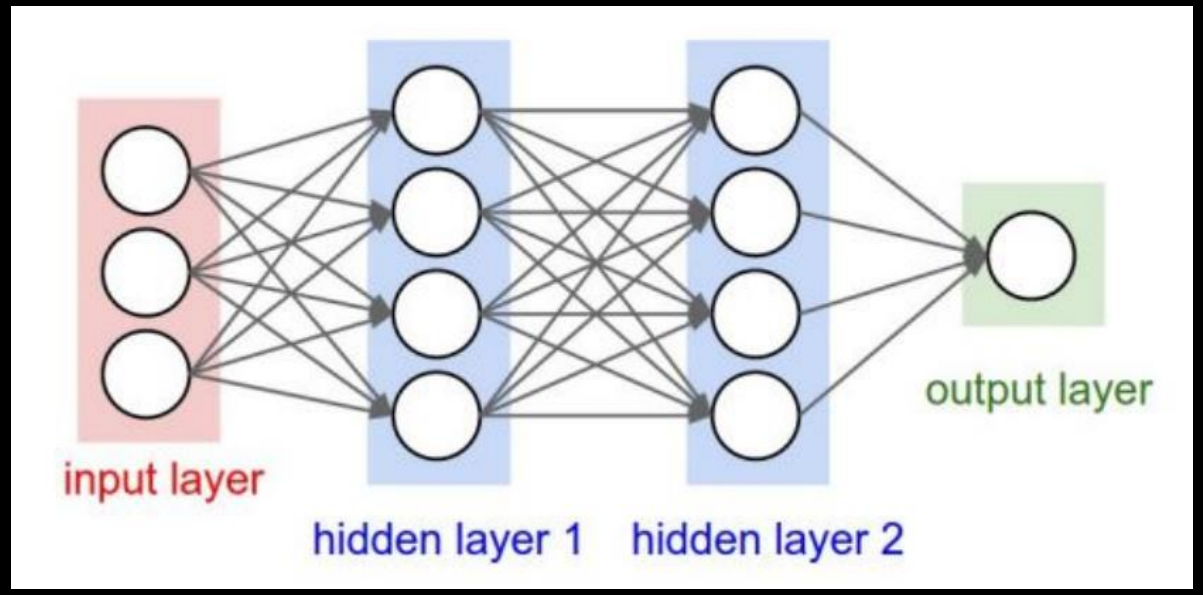
Main parts of Feed Forward Neural Network

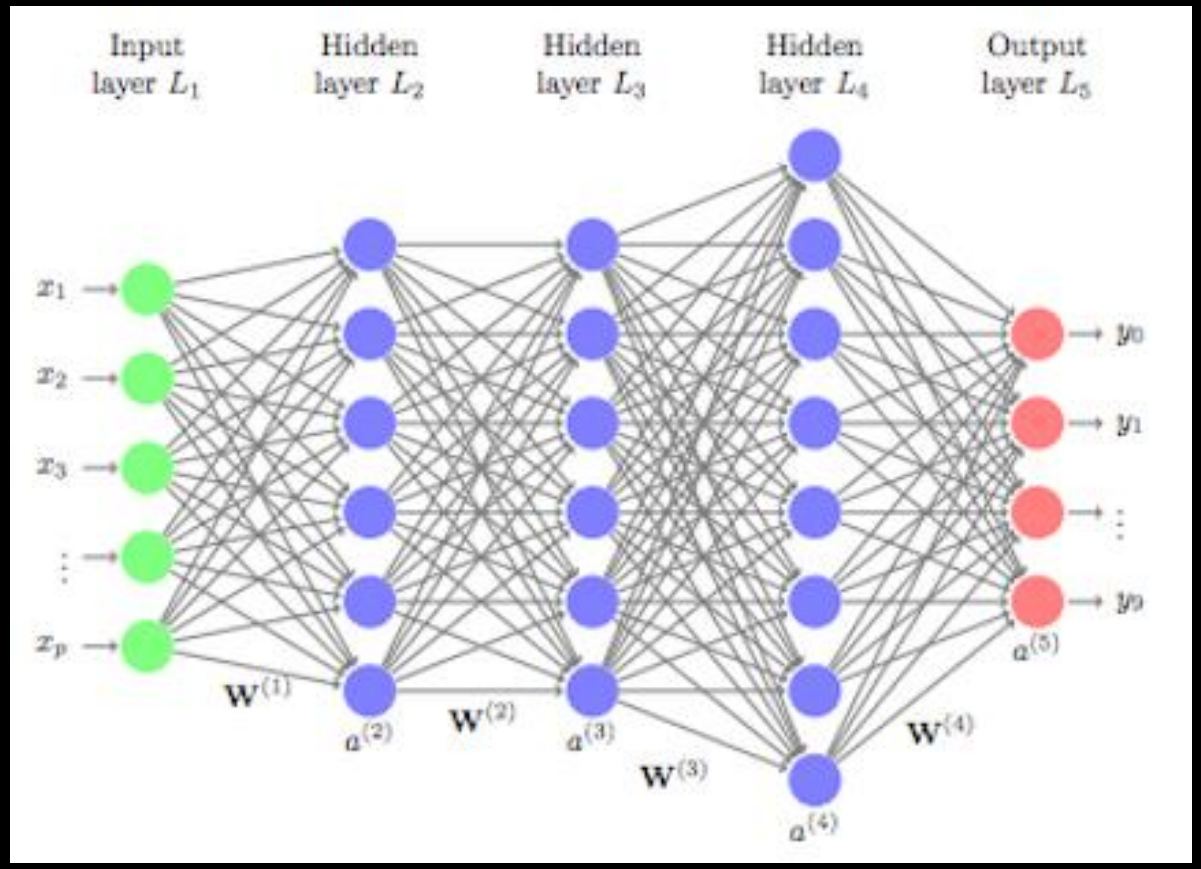
- Neurons (with its activation functions) are stored in layers
- Input layer - No computation is done here within this layer, they just pass the information to the next layer
- Hidden layers - they perform computations and then transfer the weights (signals or information) from the input layer to the following layer (another hidden layer or to the output layer)
- Output layer - Here we finally use an activation function that maps to the desired output format (e.g. Softmax for classification):

$$softmax((x_1, \ldots, x_C)) = \left(\frac{e^{x_1}}{\sum_{i=1}^{C} e^{x_i}}, \ldots, \frac{e^{x_C}}{\sum_{i=1}^{C} e^{x_i}} \right)$$

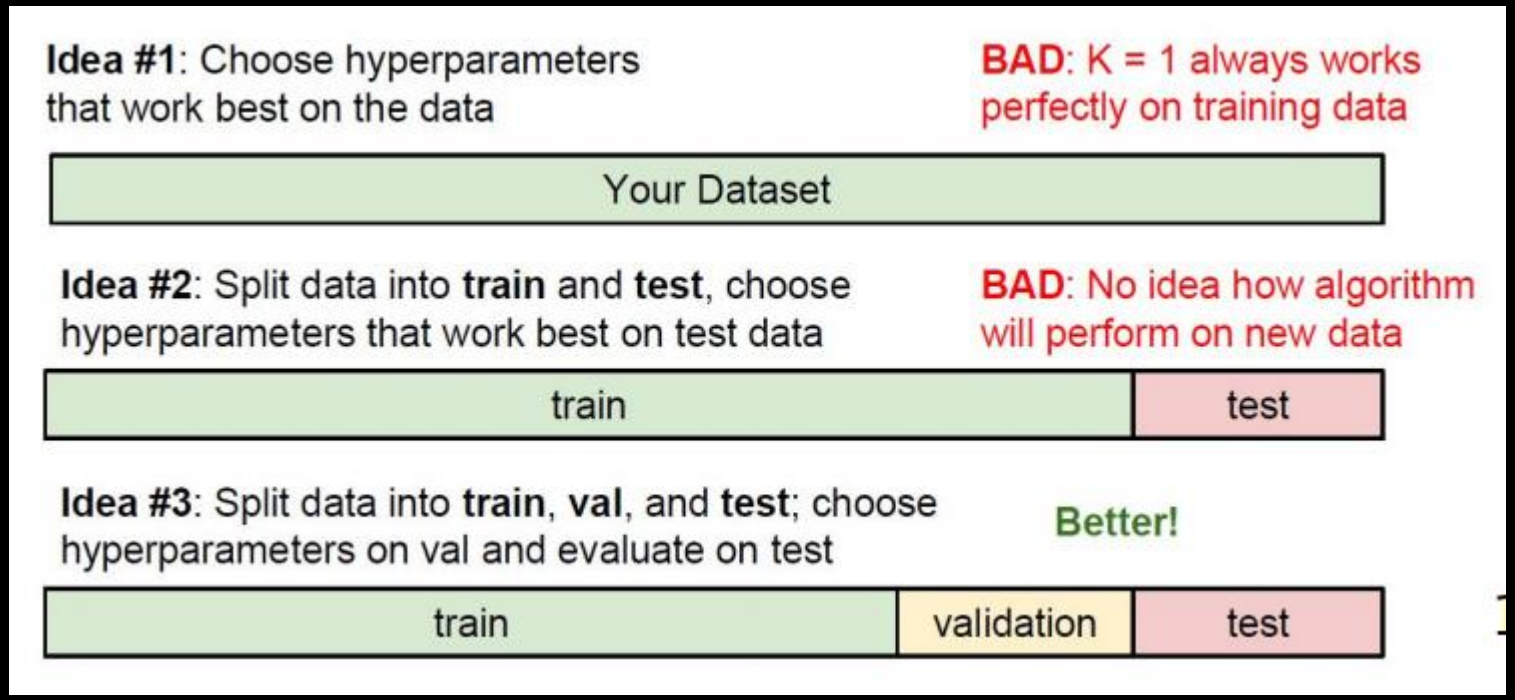
Main parts of Feed Forward Neural Network

- Fully image:
  - FC for classification
  - FC for regression



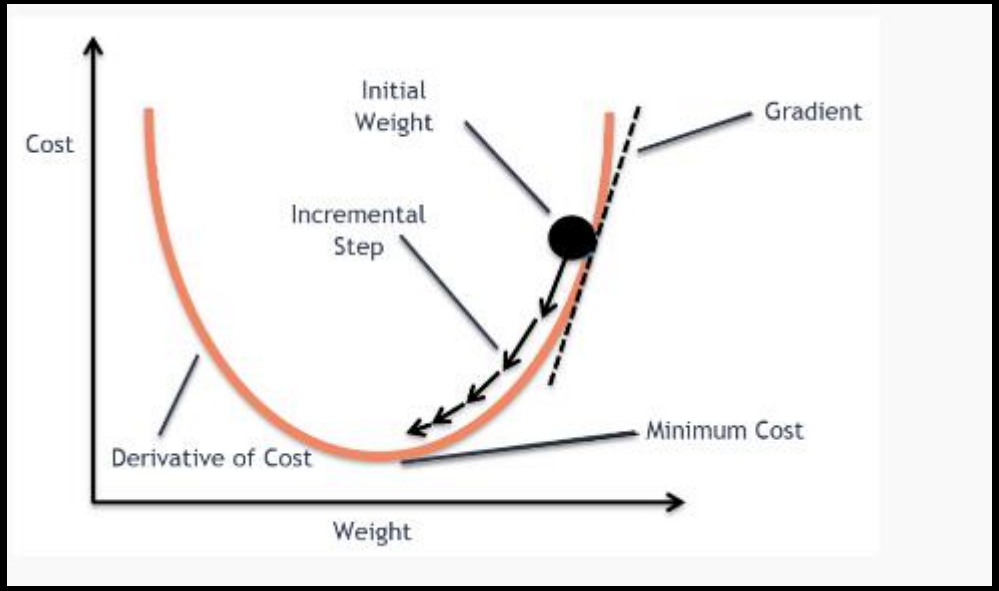


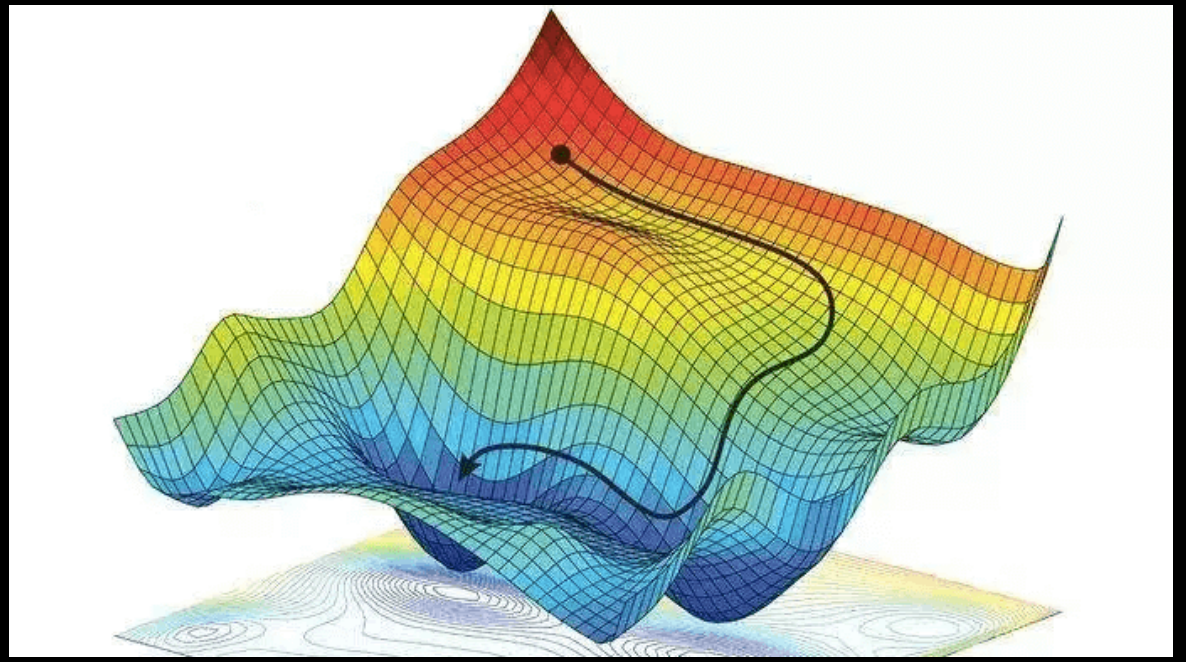
Dataset split (coding scenario)



Gradient descent

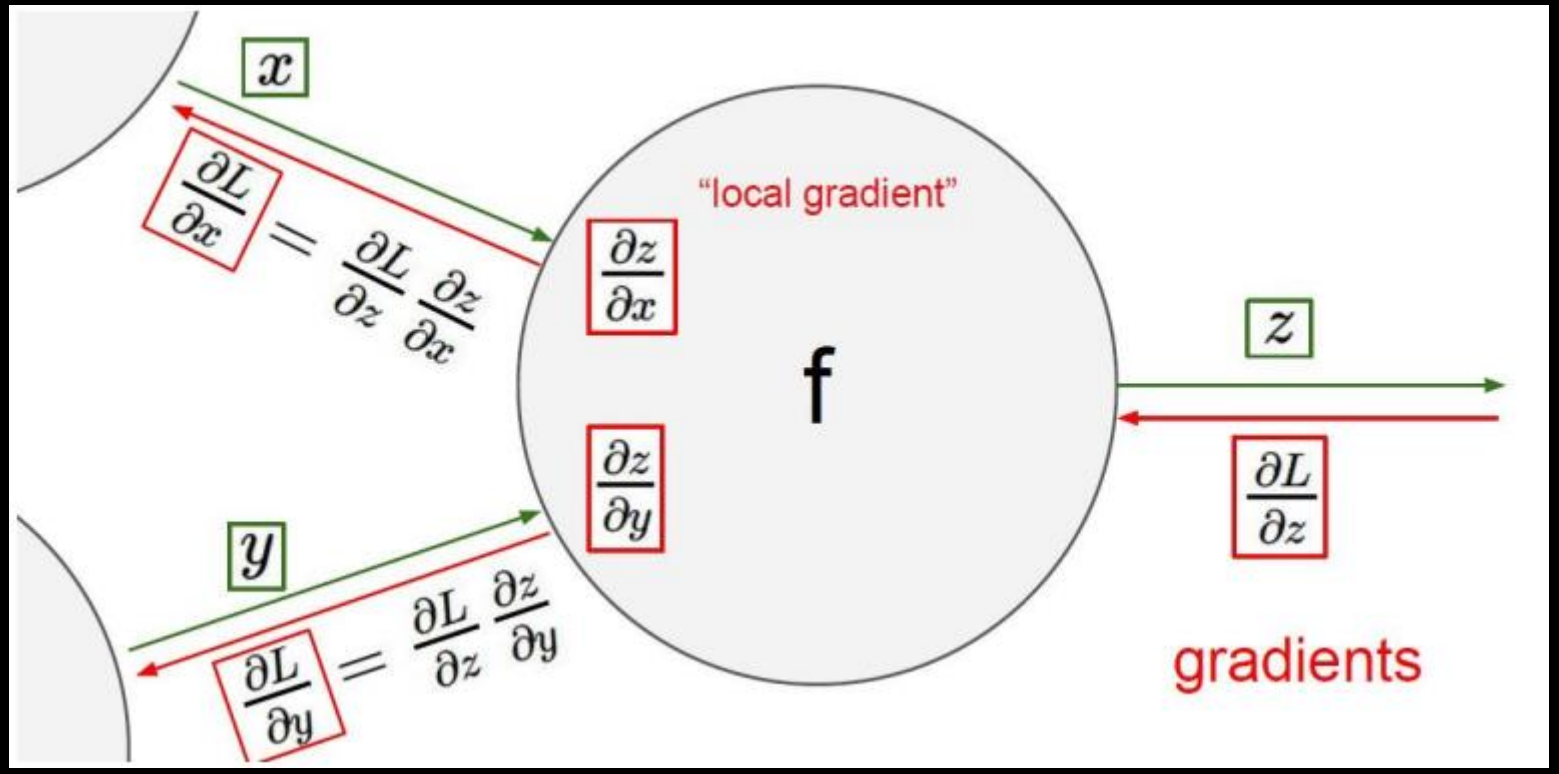
- Find the minimum of a function
- Iteratively search for a minimum
- The direction of the fastest descends





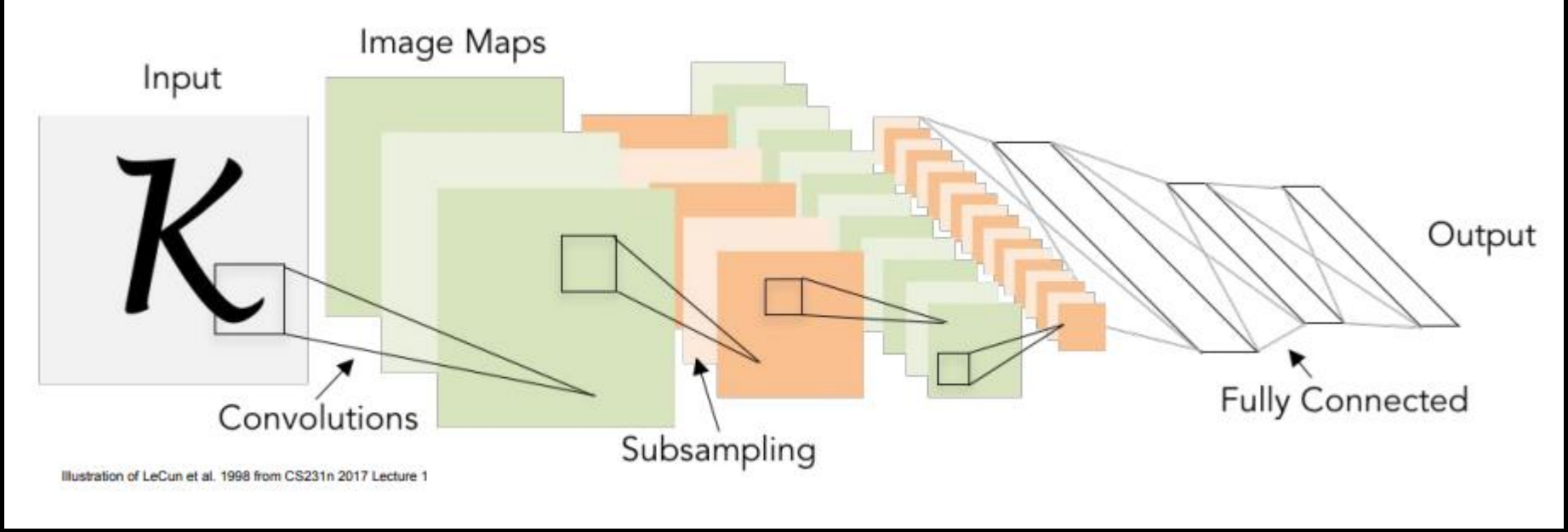
Backpropagation

- Global optimization problem
- Fine tune weights



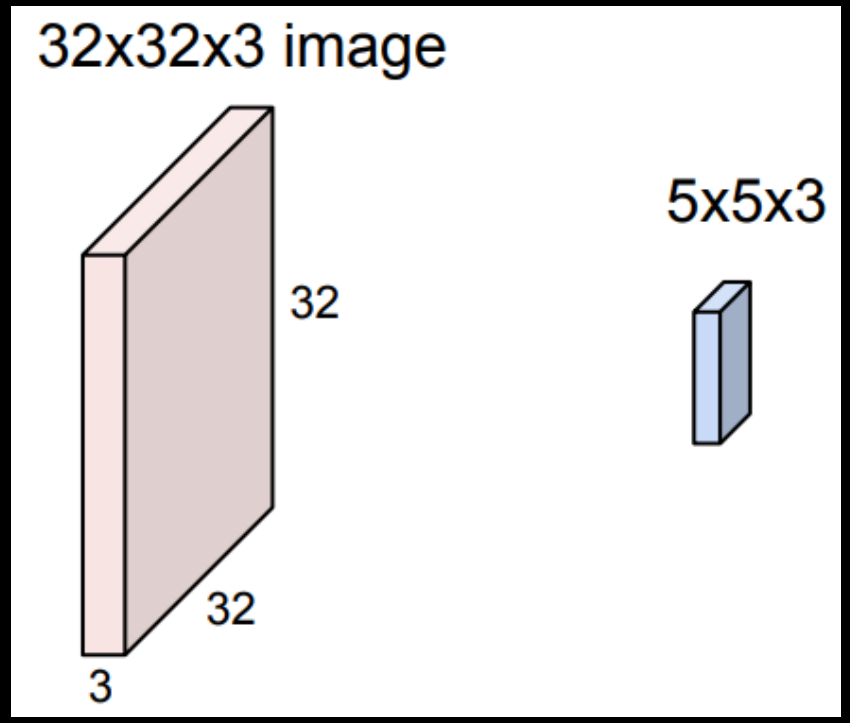
CNN

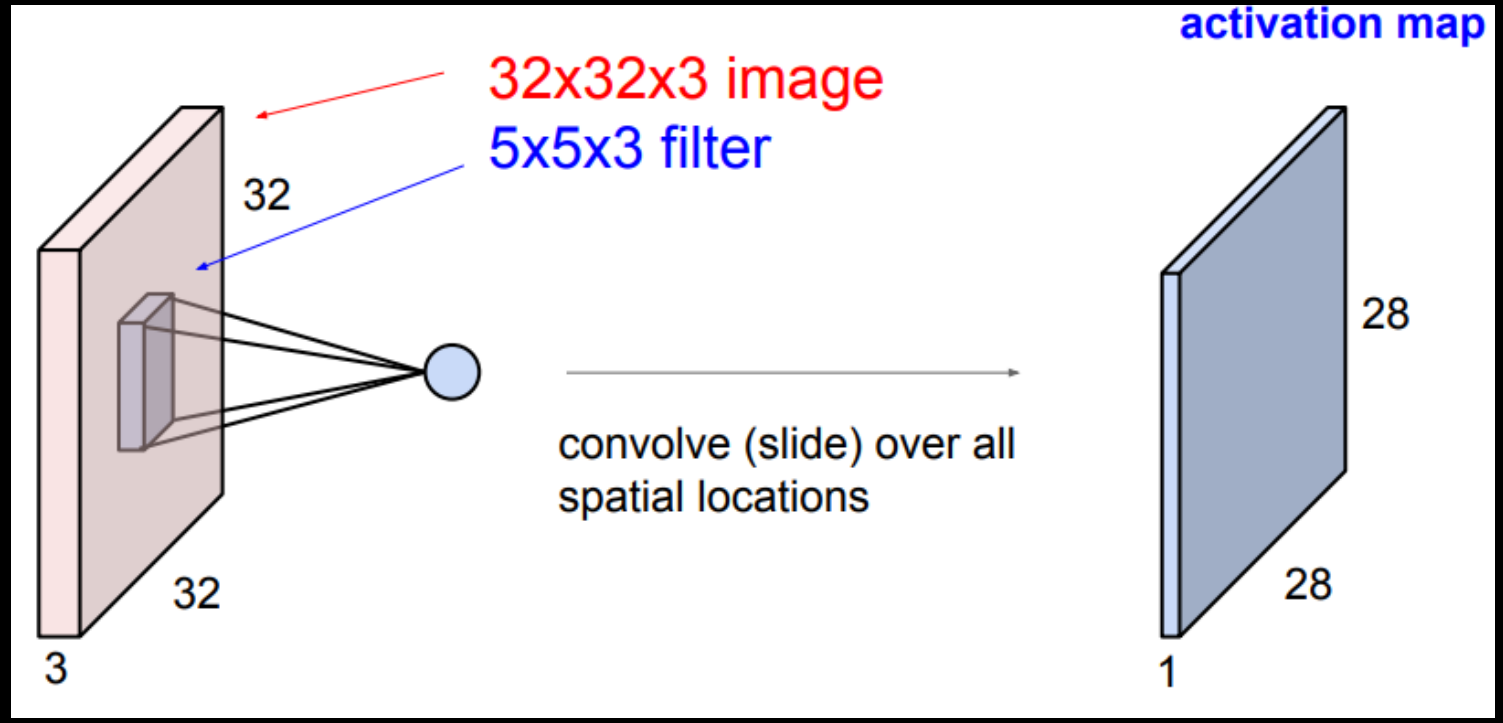
- Extensively use in Computer Vision tasks
- Raw signals (voice, images, ...)
- ConvNets stack CONV,POOL,FC layers



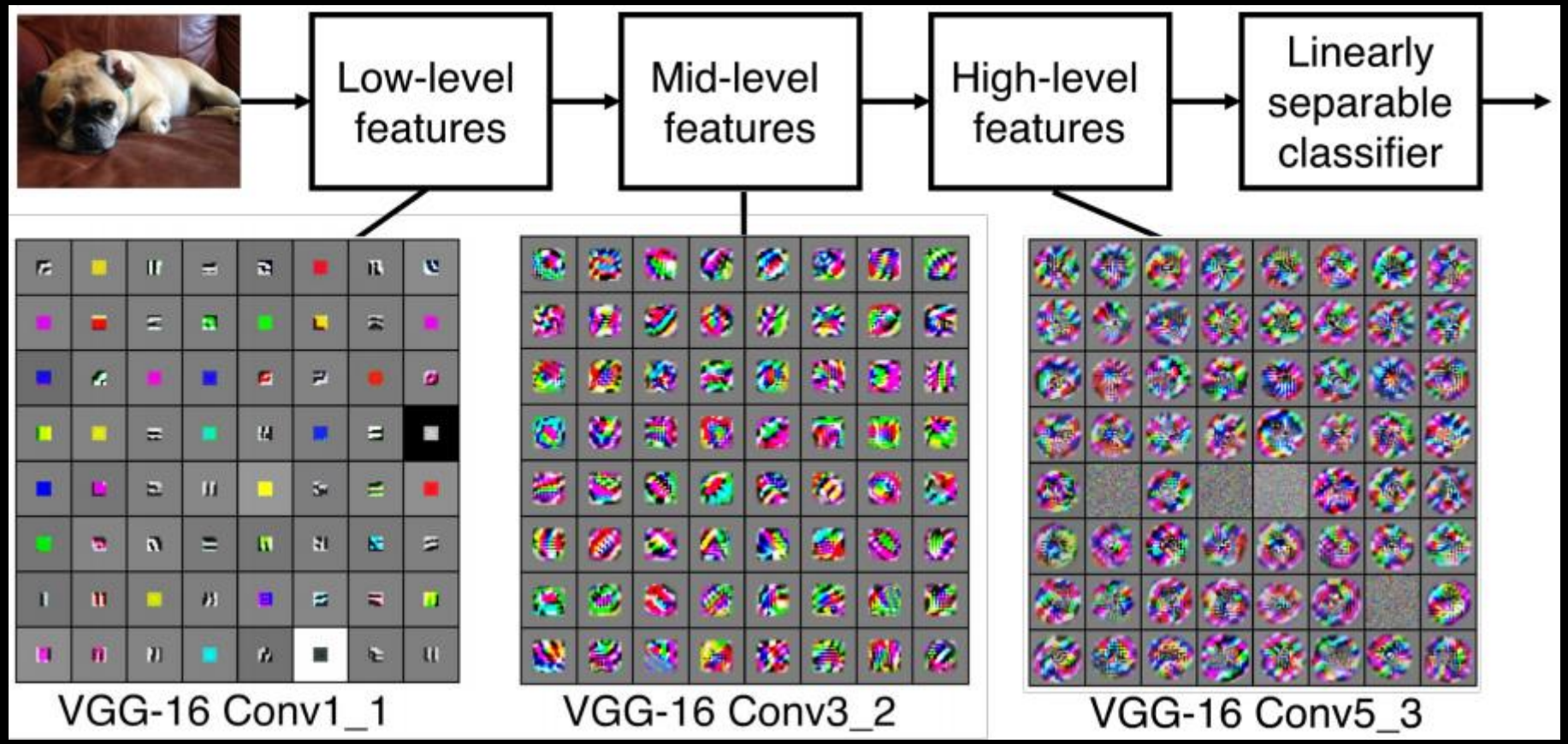
Convolution layers

- Convolve the filter with the image
- “slide over the image spatially, computing dot products”



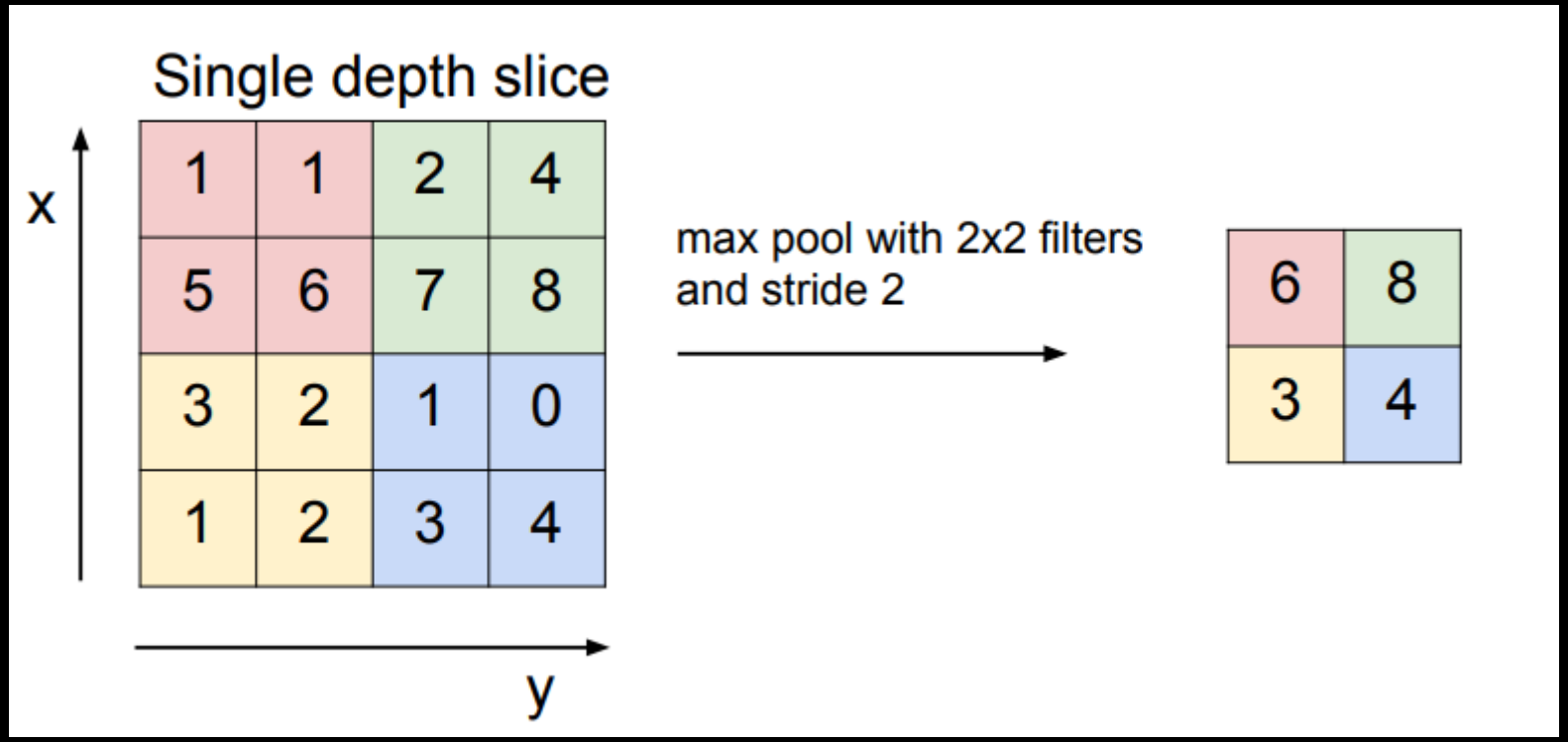


Convolution layers



Pooling layers

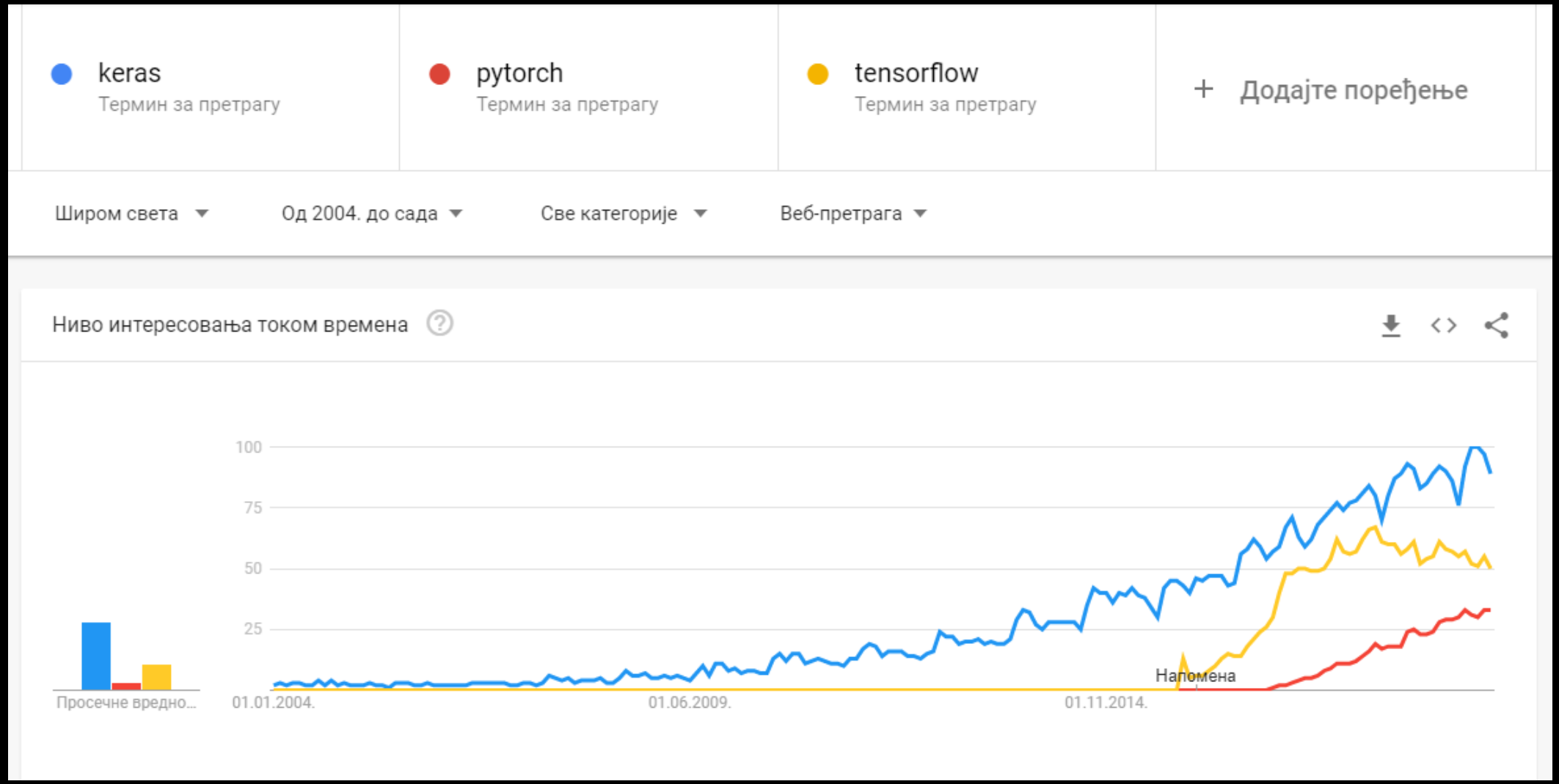
- Max Pooling, Average Pooling, etc.
- Makes the representations smaller and more manageable



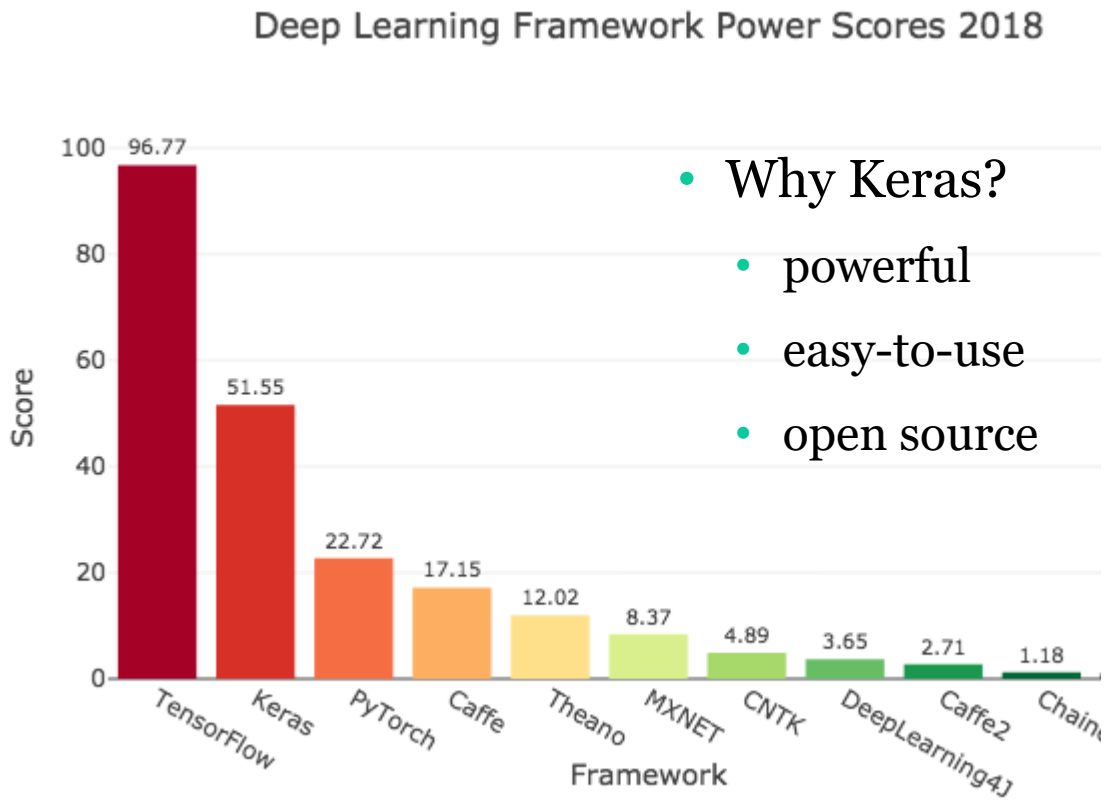
Keras Library

keras.io

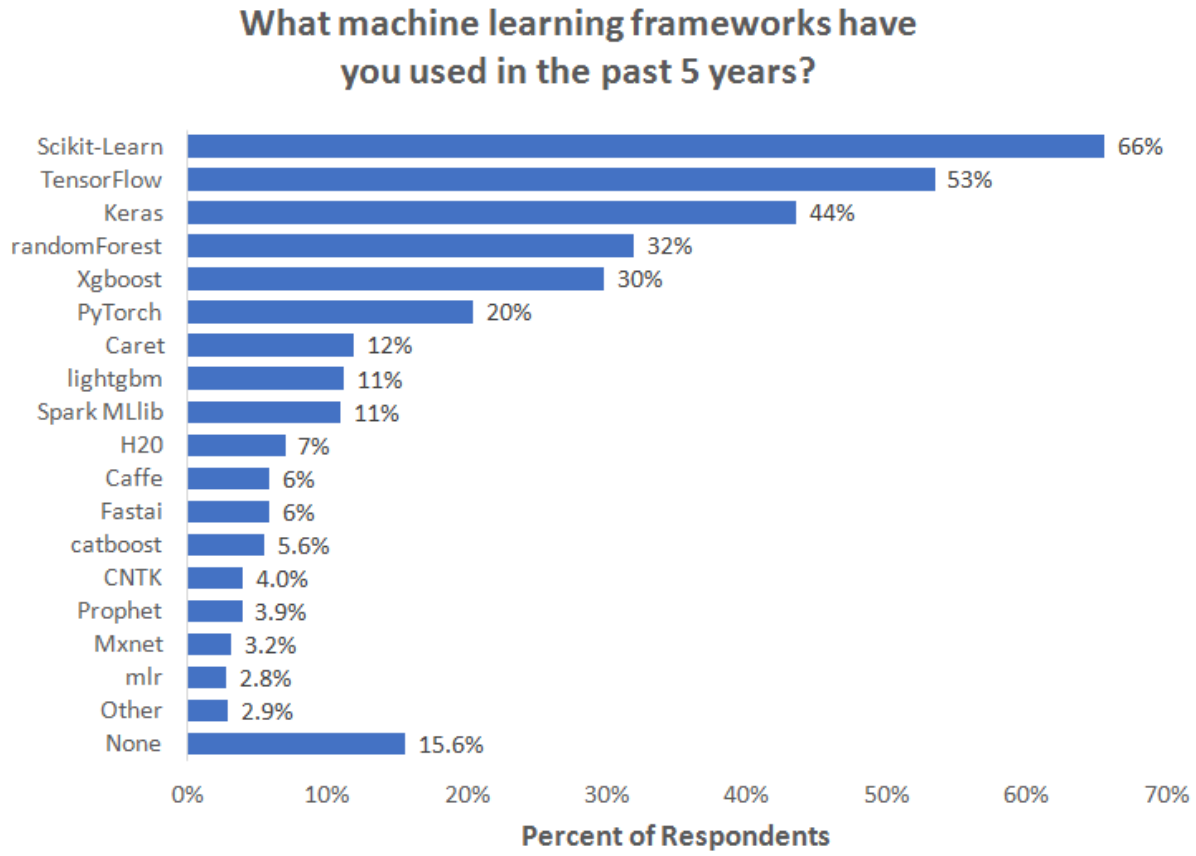
Keras (κέρας)


keras
Термин за претрагу
pytorch
Термин за претрагу
tensorflow
Термин за претрагу
+ Додајте поређење
Широм света
Од 2004. до сада
Све категорије
Веб-претрага
Ниво интересовања током времена
100
75
50
25
Напомена
Просечне вредно...
01.01.2004.
01.06.2009.
01.11.2014.

Without cheating



- Why Keras?
  - powerful
  - easy-to-use
  - open source

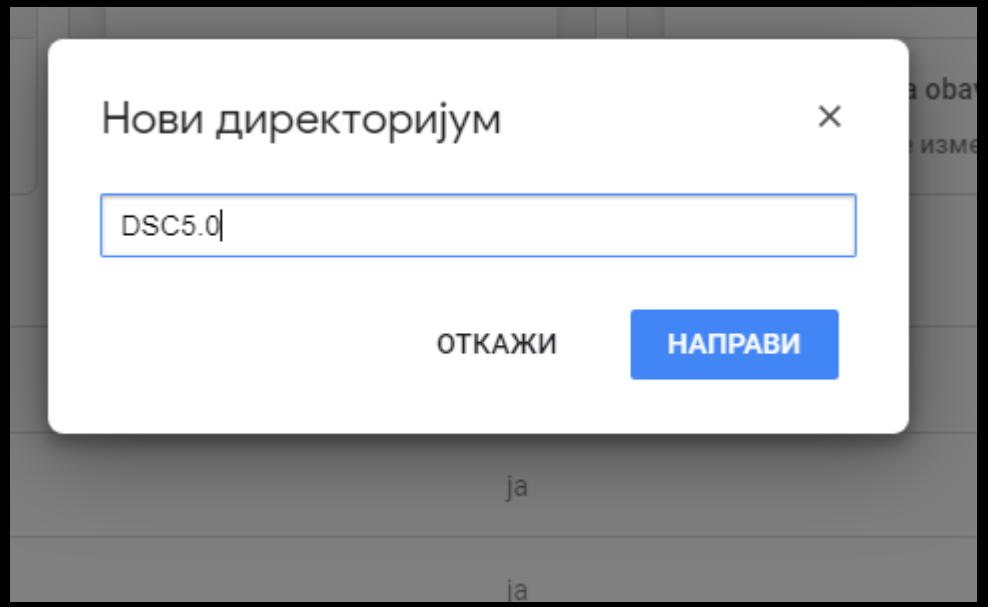


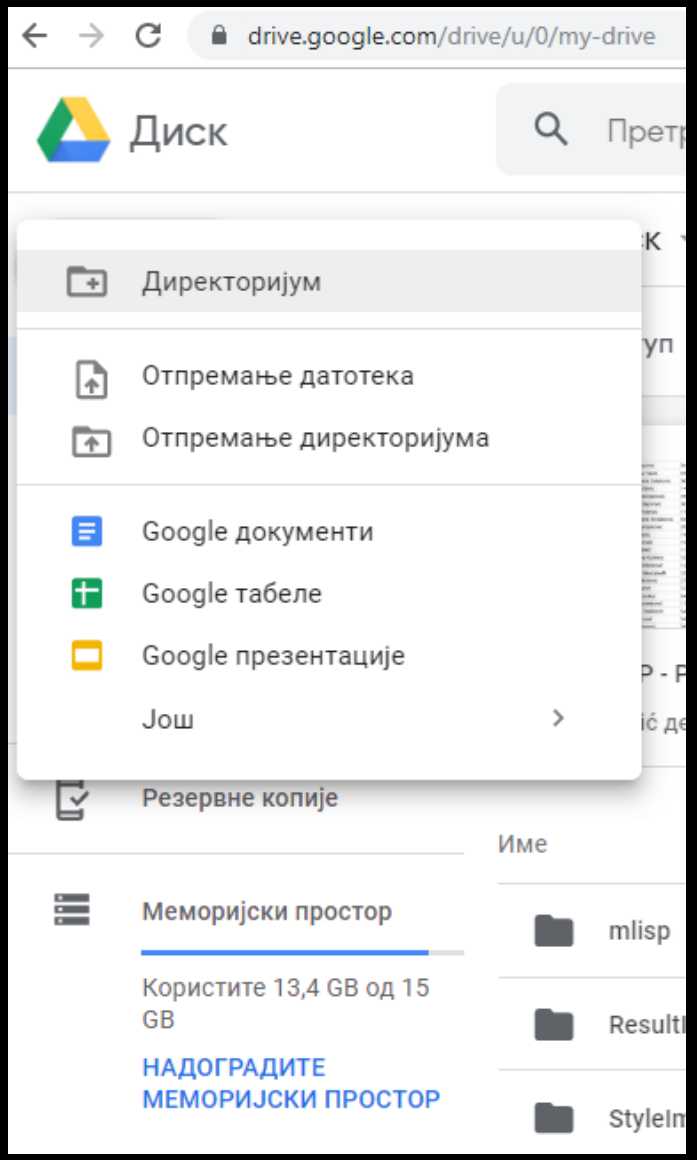
Note: Data are from the 2018 Kaggle ML and Data Science Survey. You can learn more about the study here: http://www.kaggle.com/kaggle/kaggle-survey-2018.
A total of 23859 respondents completed the survey; the percentages in the graph are based on a total of 18788 respondents who provided an answer to this question.

BUSINESS BROADWAY
DATA SCIENCE | CUSTOMER ANALYTICS | MACHINE LEARNING

Coding time

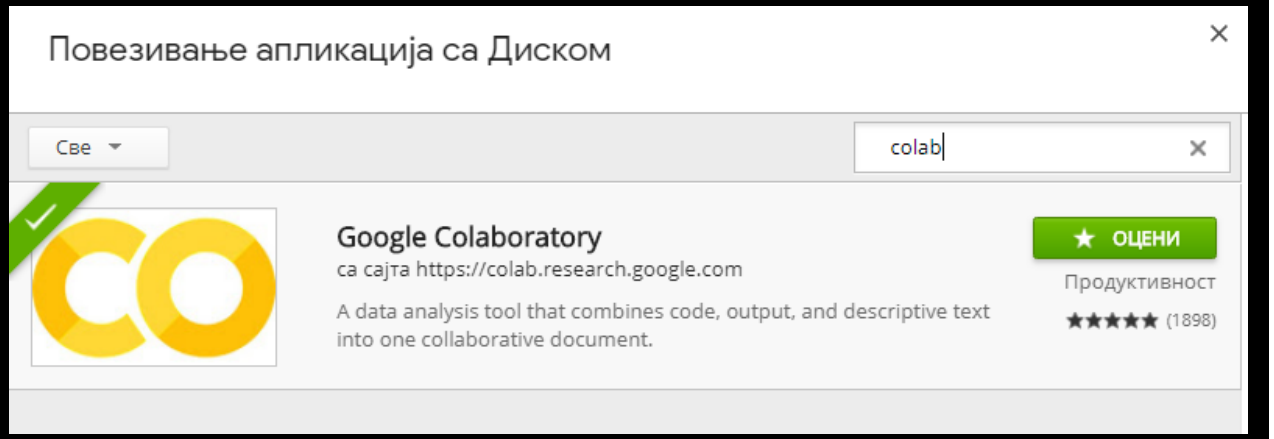
- Go to drive.google.com/
- Create new folder

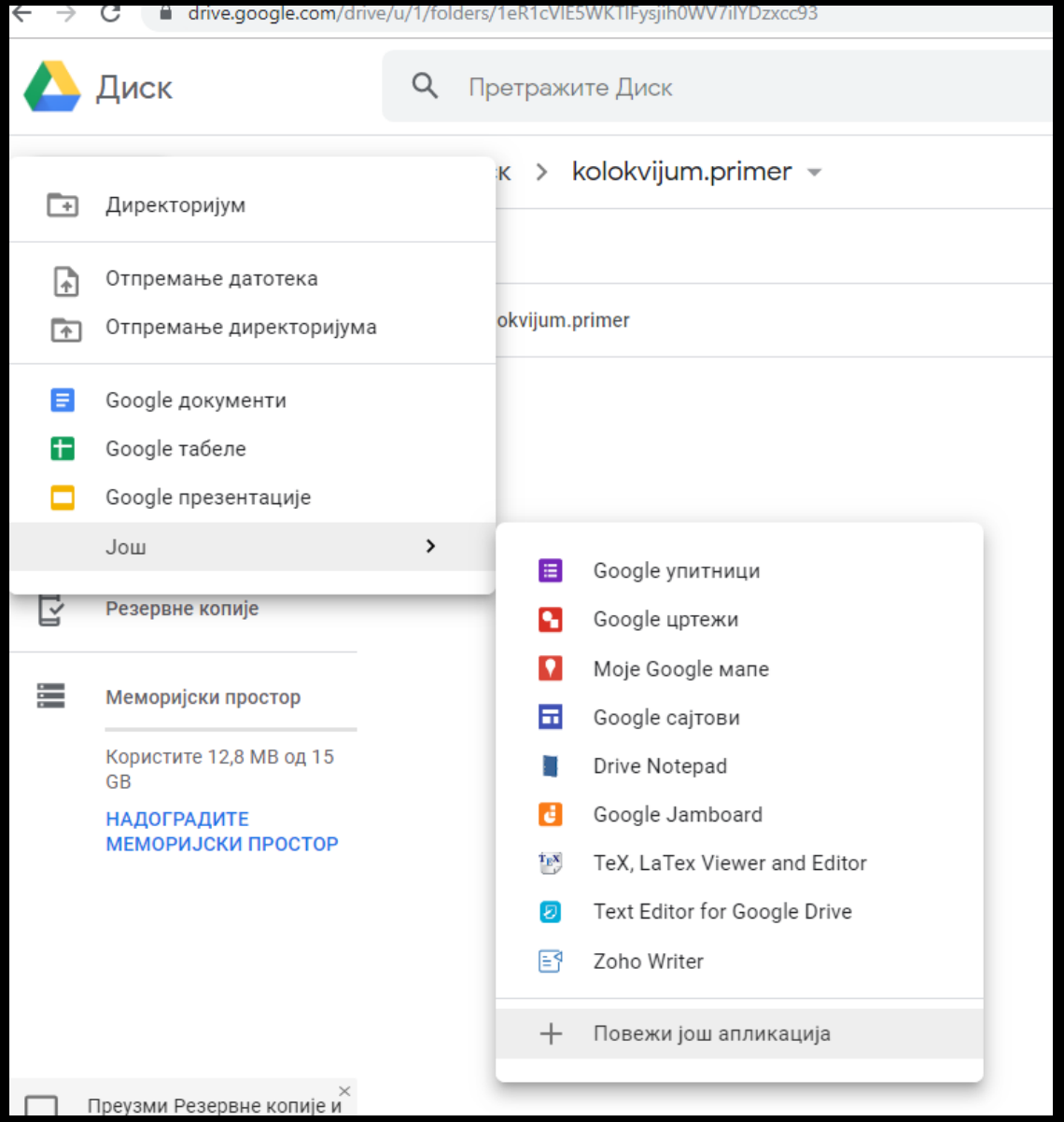




Coding time

- Connect with Google Colab
- Go to: colab.research.google.com/
- Enjoy coding!

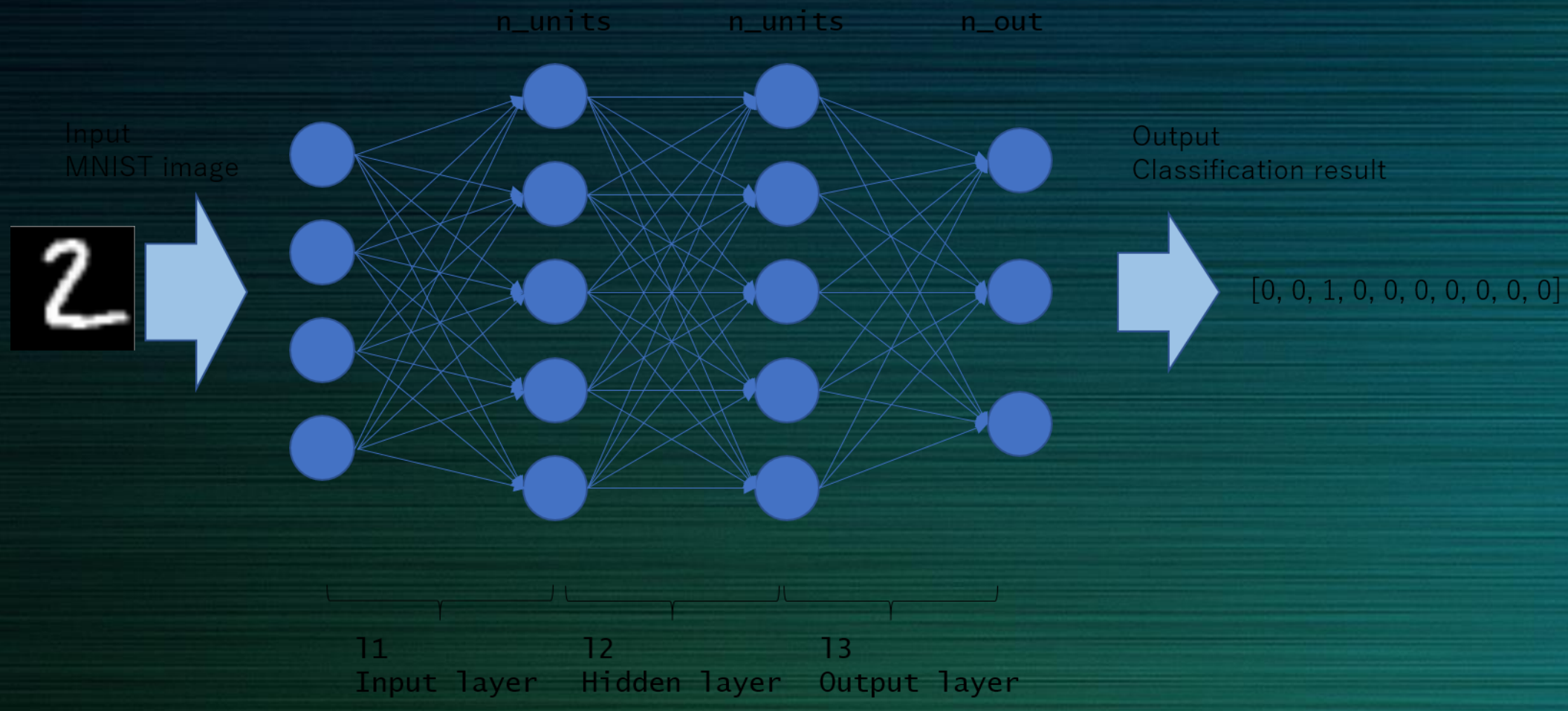




Example 1

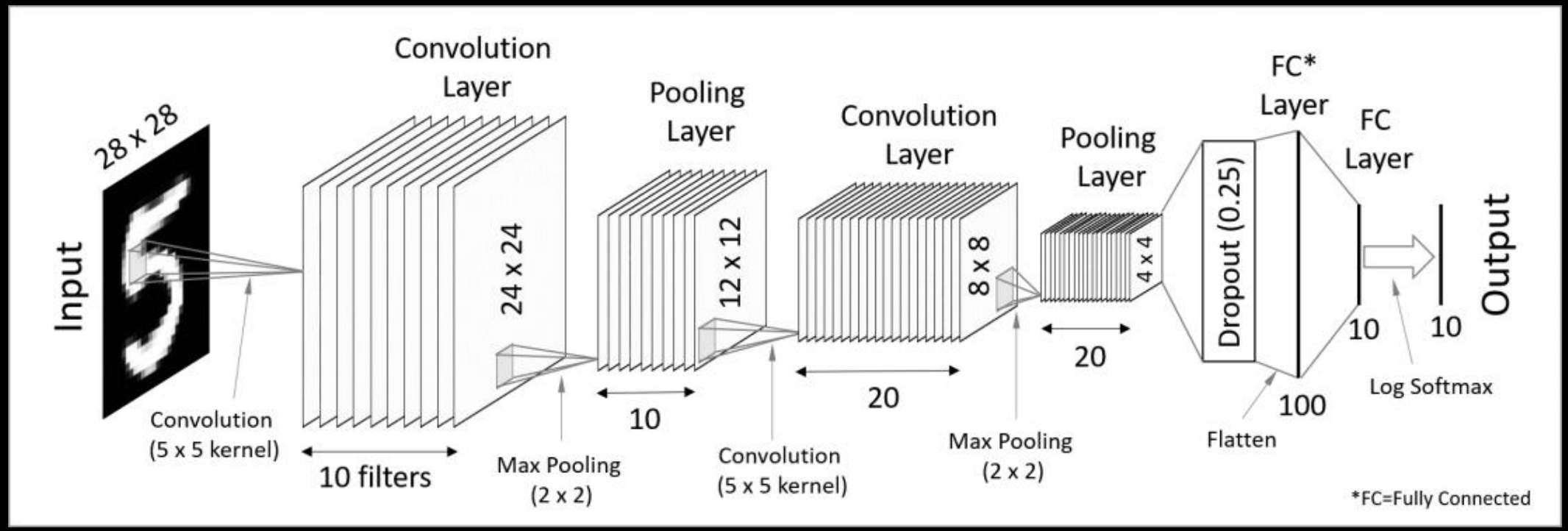
- **Pima Indians onset of diabetes dataset**
- **Predict diabetes**

Example 2


n_units
n_units
n_out
Input
MNIST image
Output
Classification result
[0, 0, 1, 0, 0, 0, 0, 0, 0, 0]
l1
Input layer
l2
Hidden layer
l3
Output layer

Example 3

- Ex2_MNIST_CNN.ipynb
- Trains a simple ConvNet on the MNIST dataset.
- Gets to 99.25% test accuracy after 12 epochs



Thank you

- milan_cugurovic@matf.bg.ac.rs
- milancugurovic@gmail.com
- +381 65 69 24 577
- matf.bg.ac.rs/milan_cugurovic (in Serbian only)
- linkedin.com/in/milancugurovic
- github.com/MilanCugur