PROJECT HUB

WEEK 1

MACHINE LEARNING PROJECT

ICICS 014

3:30PM | FEB 2 | BIWEEKLY

UBC LAUNCH PAD
SOFTWARE ENGINEERING TEAM

UBC CSSS

Slides: https://slides.ubclaunchpad.com/workshops/ml-intro.pdf

UBC Project Hub

- Joint initiative between UBC Launch Pad and CSSS.
- Goal: to create a learning environment at UBC that nurtures a culture of design, innovation, and community amongst the future hackers and entrepreneurs of the tech industry.
- Biweekly meetings with talks & workshops.
- Pizza will be ordered after head count. 🍕

Introduction to Machine Learning

Kevin Yap & Sherry Yuan

Slides: https://slides.ubclaunchpad.com/workshops/ml-intro.pdf

Today's Agenda

- Talk: **An Overview of Machine Learning** (Kevin)
    - Motivations, successes, and limitations of ML.
- Workshop: **Predicting Credit Card Defaults** (Sherry)
    - Interactive dive into ML with real-world data.

About

- Kevin Yap (@iKevinY)
- 5th Year Honours Computer Science
- Experimented with NLP at Axiom Zen
- Built neural network for nwHacks 2018 project
- Took CPSC 340 two years ago
- Finishing up thesis on machine learning & StarCraft II
- Former Launch Pad ML tech lead

About

- Sherry Yuan (@frostyshadows)
- 5th Year Computer Science
- Took CPSC 340 one year ago
- Launch Pad Co-President

Goals for this Talk

- Discuss motivations for machine learning.
- Short overview of the history of the field.
- Briefly touch on various techniques.
- Introduce jargon and other terminology.
- Show that machine learning is approachable!

What is "machine learning"?

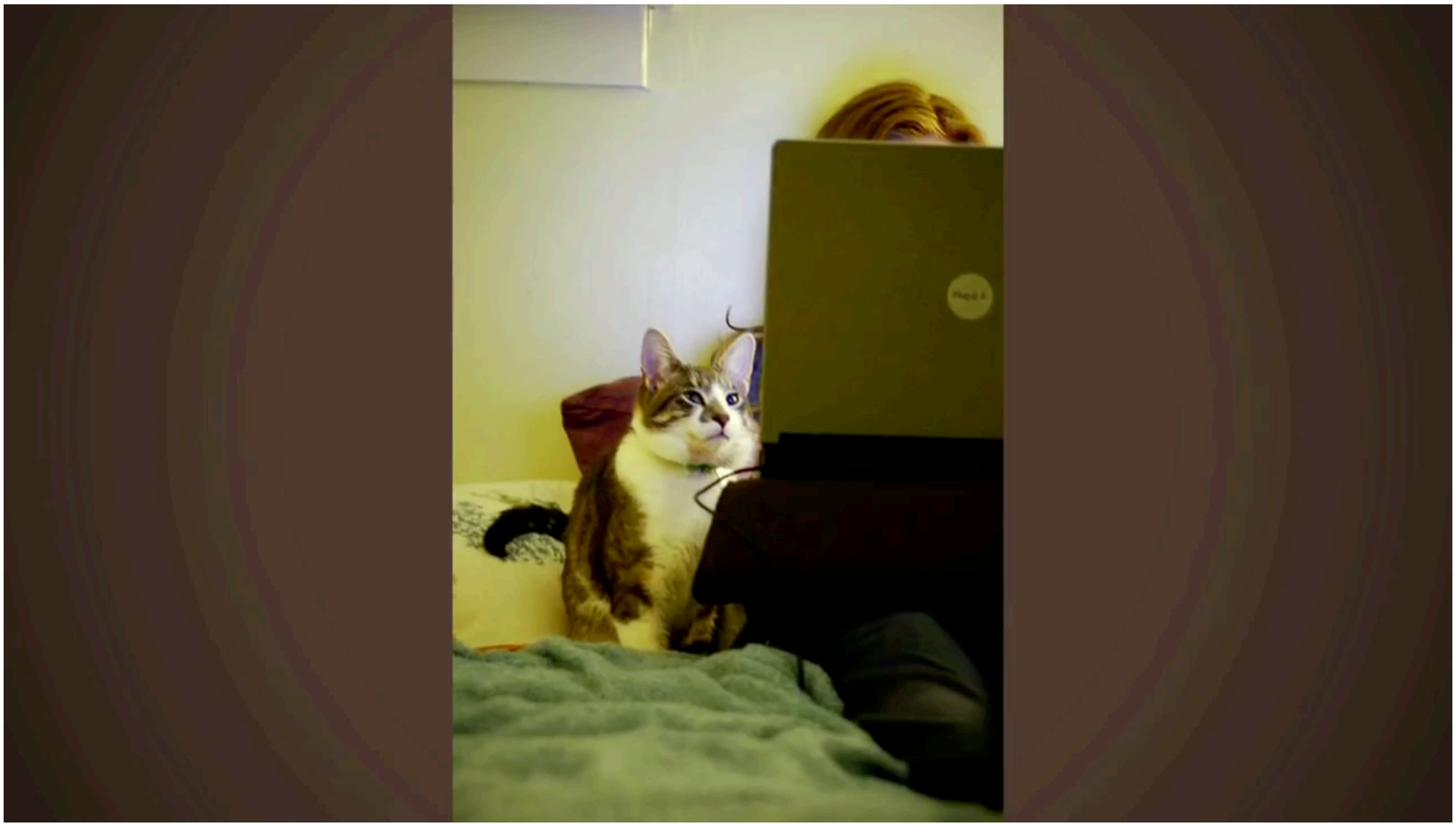
- Machine learning (ML) is the study using **algorithms** and **statistical models** to allow computer systems to effectively perform a specific task **without using explicit instructions**, relying on **models** and **inference** instead.

- Subfield of AI (artificial intelligence).

Applications of Machine Learning

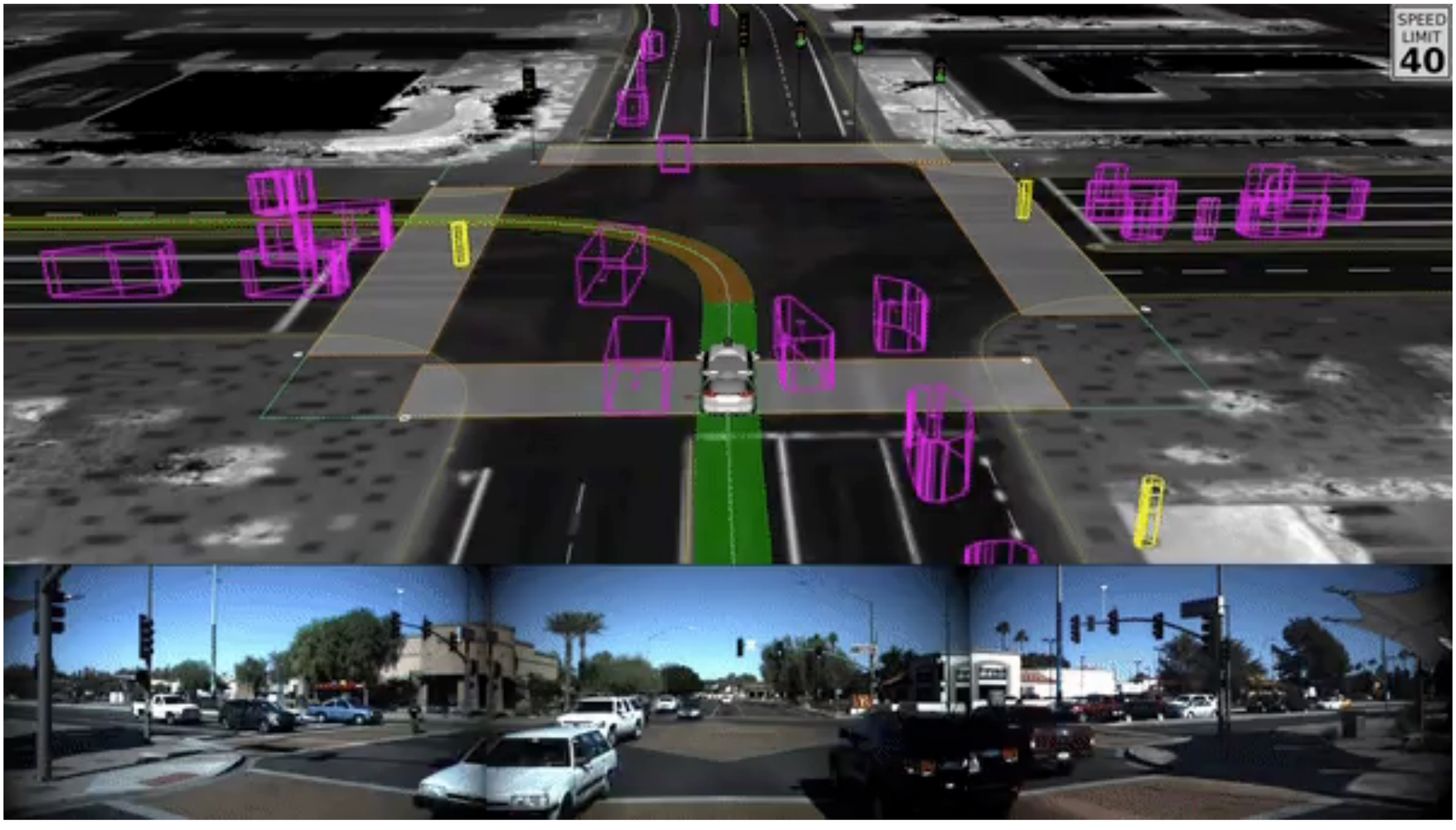
- Artificial Intelligence (game agents)
- Computer Vision (self-driving cars)
- Natural Language Processing (machine translation)
- Recommendation Systems (Netflix/Amazon suggestions)

Computer Vision



TED Talk: How we teach computers to understand pictures (Fei Fei Li)

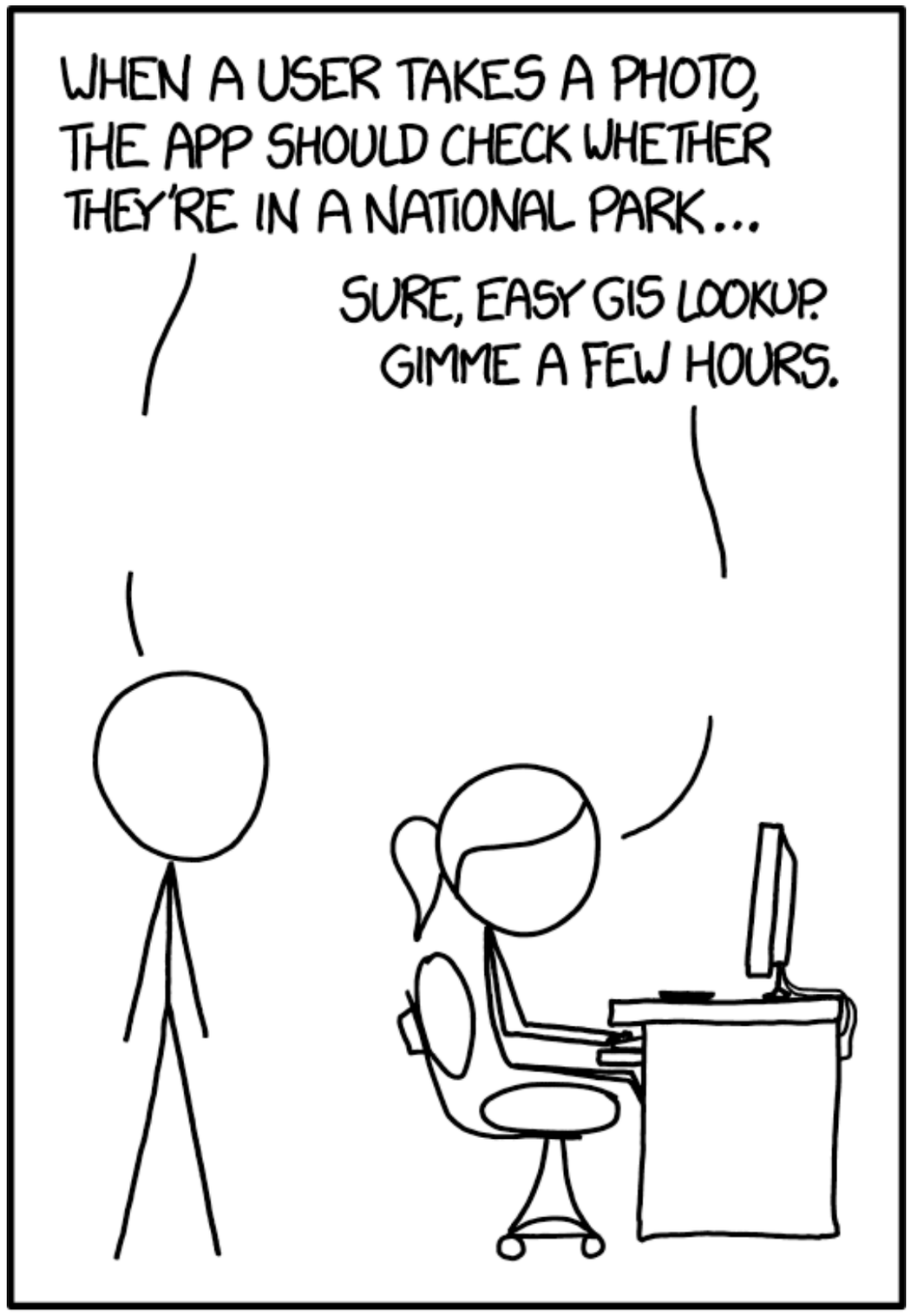
Waymo's Self-Driving Car



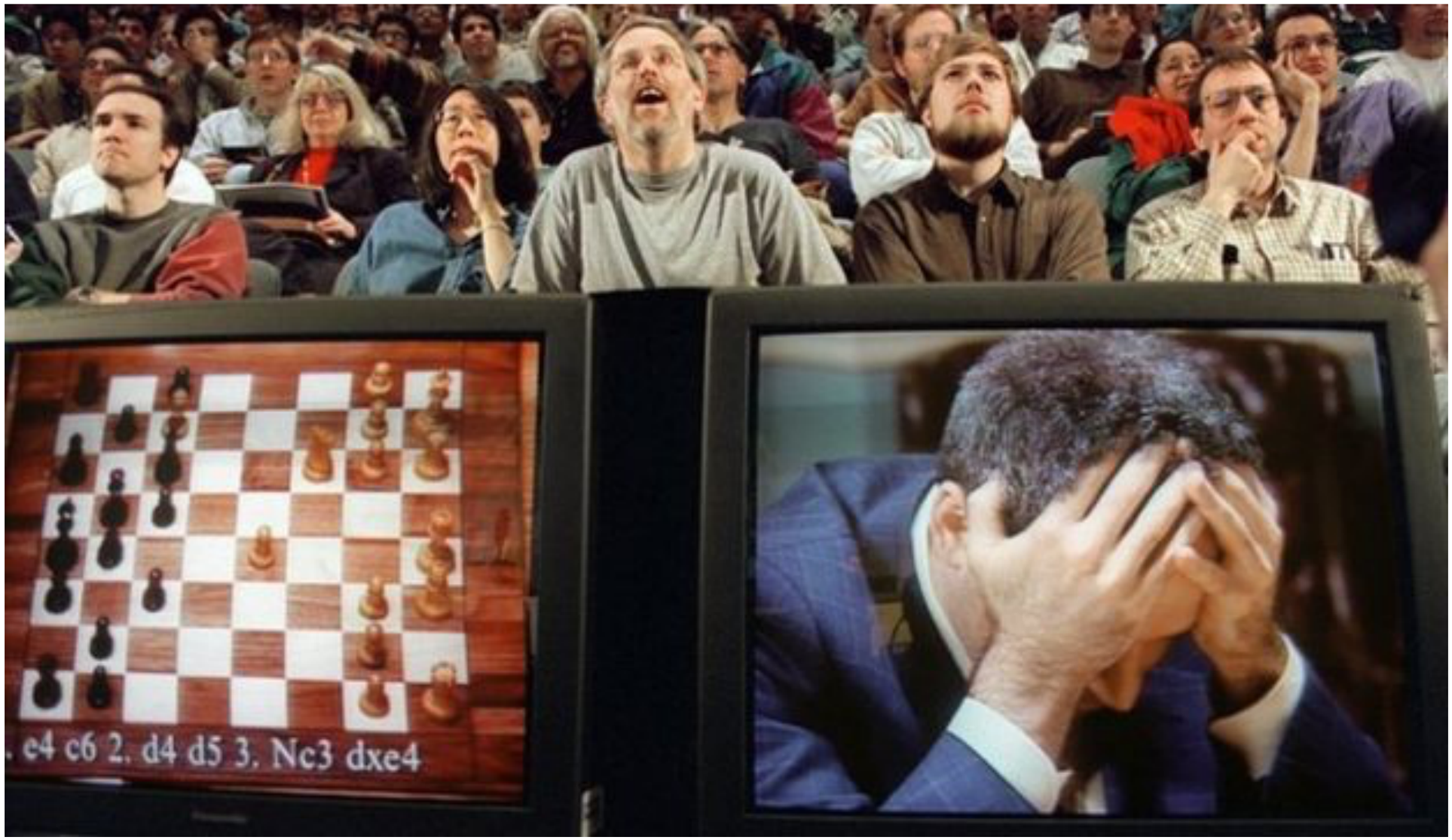
https://www.recode.net/2018/2/28/17059184/alphabet-google-waymo-self-driving-consumer-trust

Chihuahua or Muffin

https://www.topbots.com/chihuahua-muffin-searching-best-computer-vision-api/

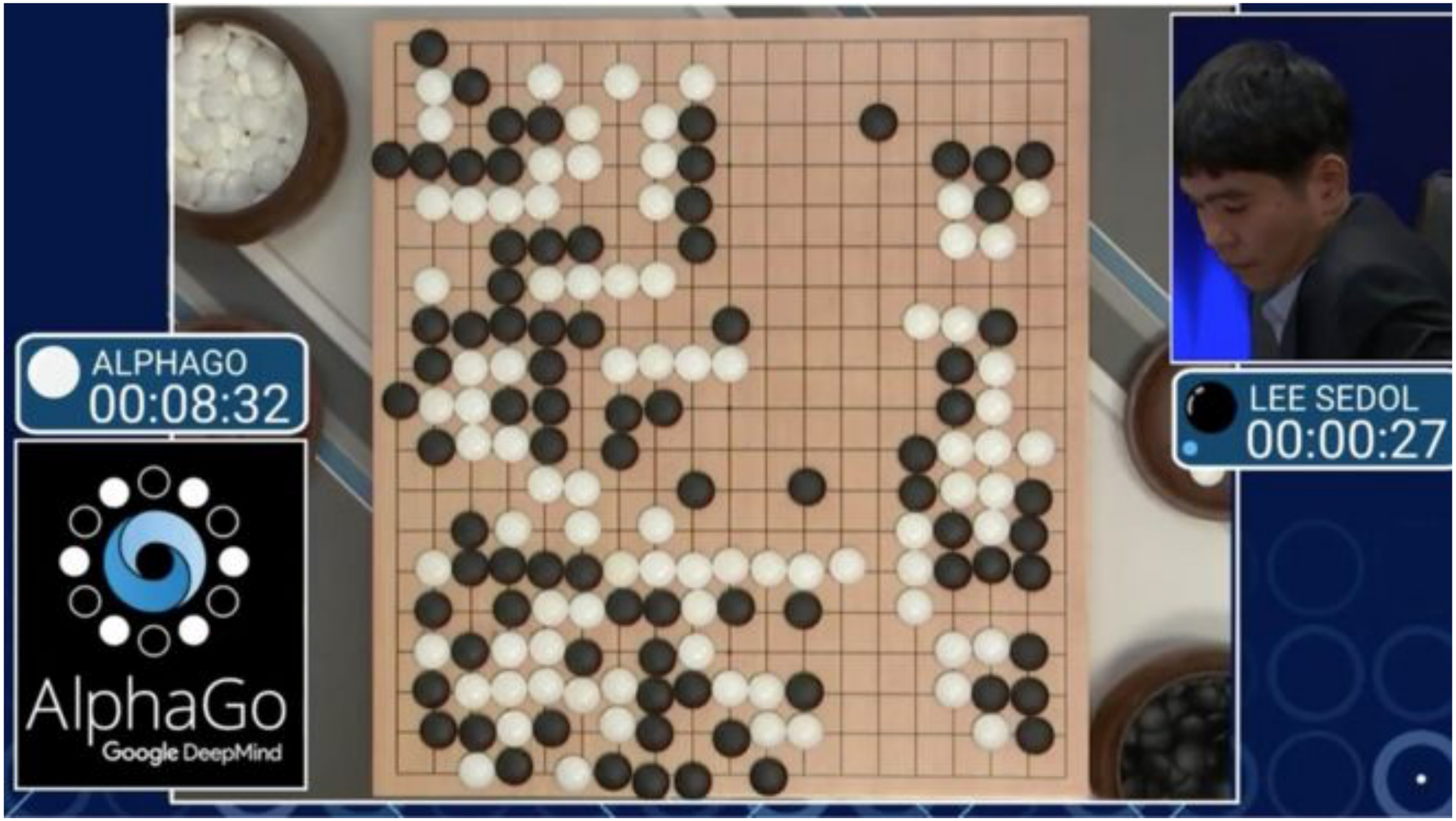


https://xkcd.com/1425/



1997: Deep Blue beats Garry Kasparov in chess

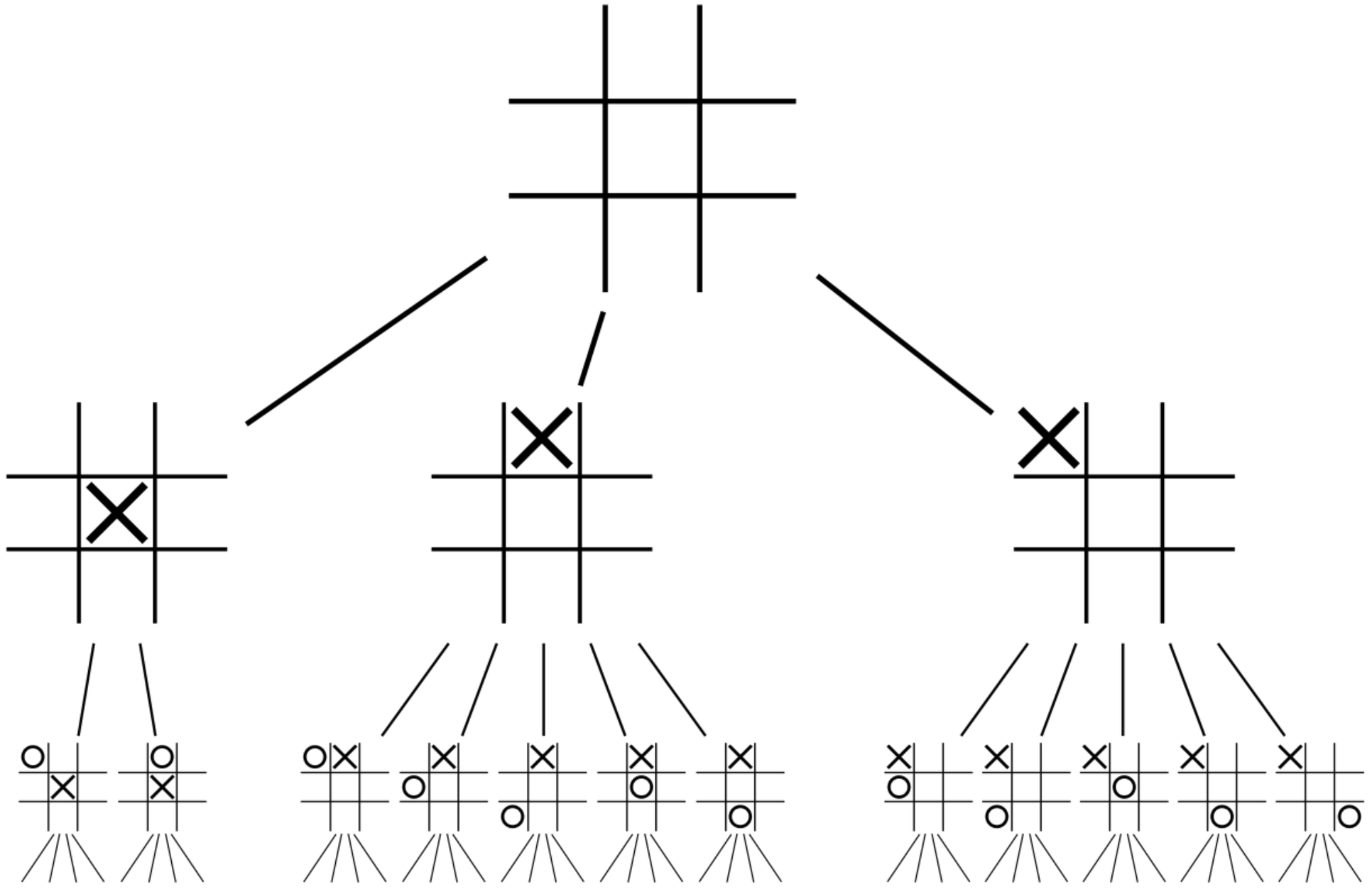
https://www.bbc.com/news/technology-35785875



2016: AlphaGo beats Lee Se-dol at Go

https://www.bbc.com/news/technology-35785875

Solving Chess vs. Go



Solving Chess vs. Go

	Board Size	Pieces	Branching Factor	Space
Tic-Tac-Toe	3×3	9	4	512
Checkers	8×8	24	2.8	$5 \cdot 10^{20}$
Chess	8×8	24	35	10^{120}
Go	19×19	361	250	**10^{360}**

Solving Chess vs. Go

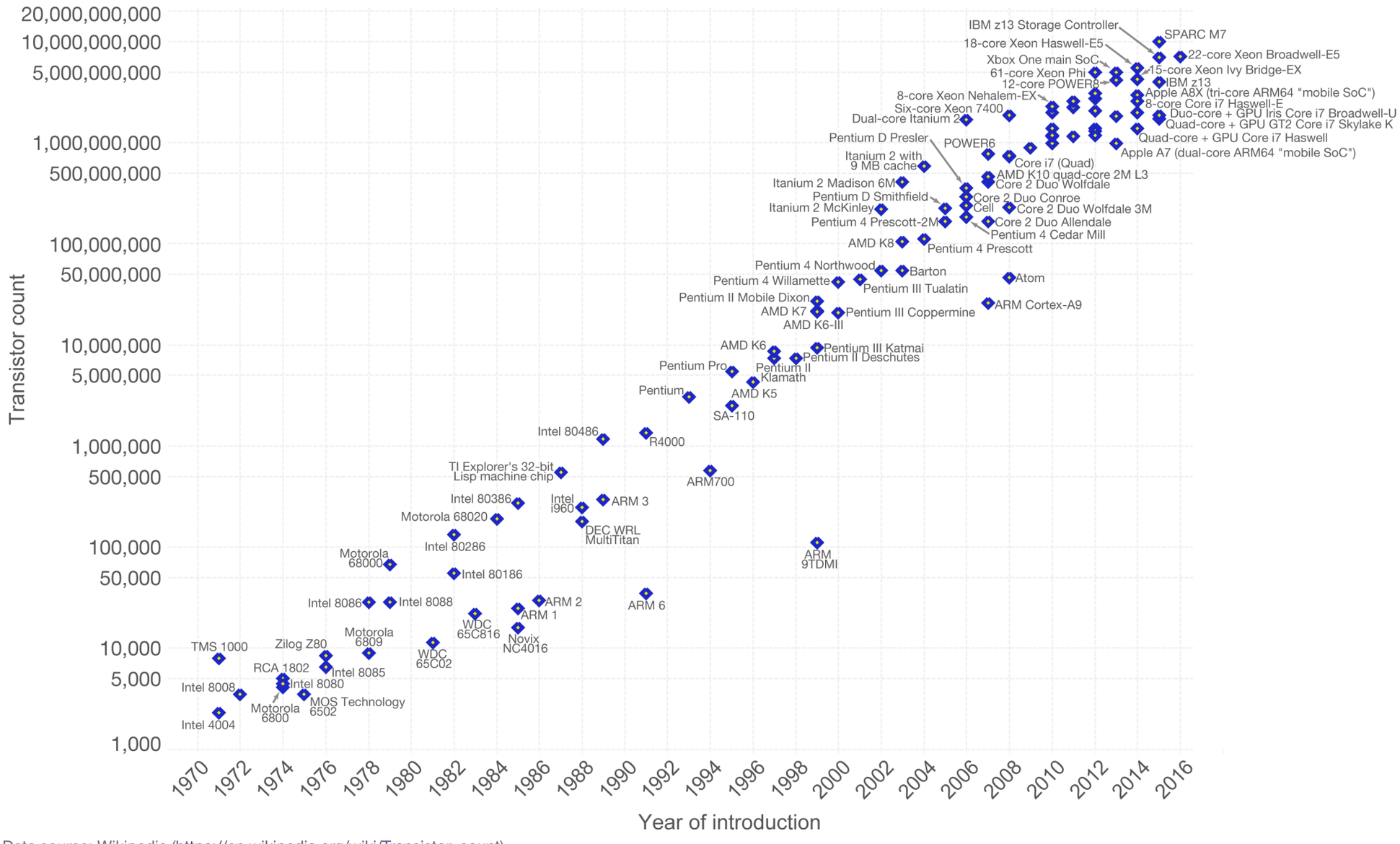
- Deep Blue: rule-based system, basic tree search
- AlphaGo: tree search + neural network

The **Big Data** Boom

Moore’s Law – The number of transistors on integrated circuit chips (1971-2016)

Our World in Data

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important as other aspects of technological progress – such as processing speed or the price of electronic products – are strongly linked to Moore's law.



Data source: Wikipedia (https://en.wikipedia.org/wiki/Transistor_count)
The data visualization is available at OurWorldinData.org. There you find more visualizations and research on this topic.

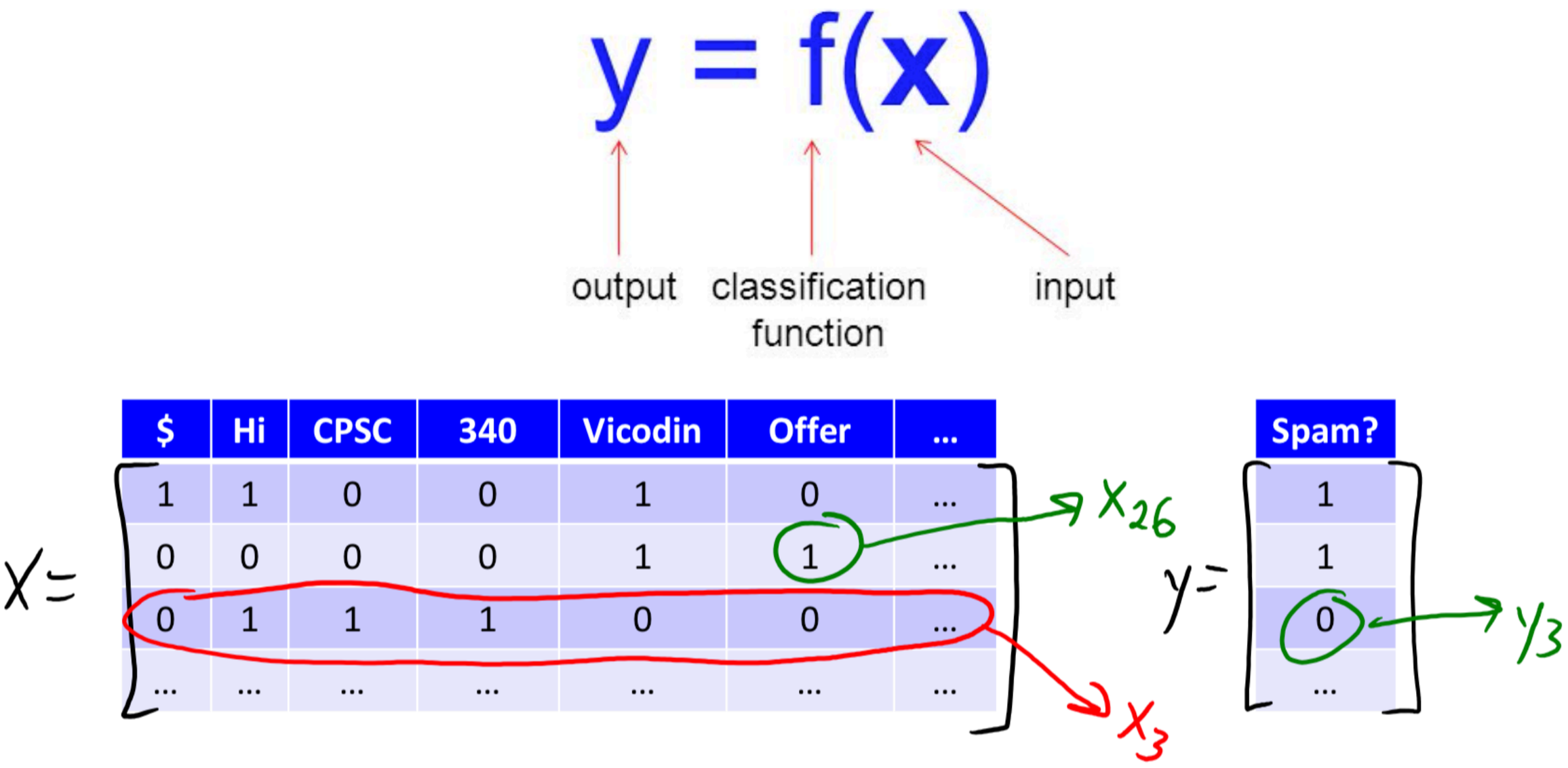
Machine Learning Basics

ML in Practice

- Python
- NumPy to interact with data (matrices)
- Uses C bindings under the hood

- We choose **hyperparameters** for the model
- Models learn **parameters** through looking at data

Predicting y from X

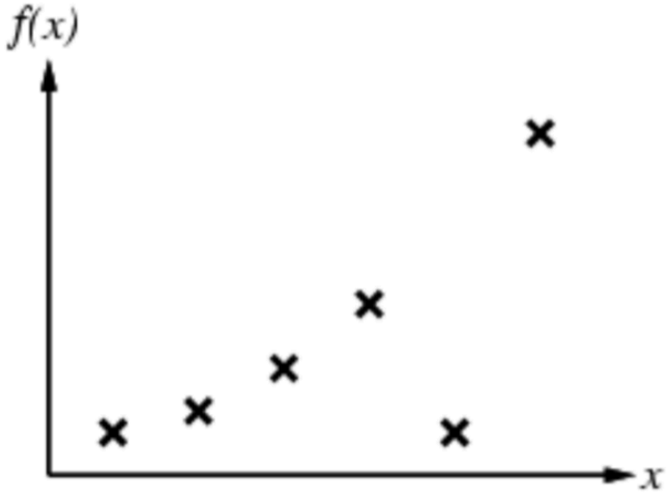


https://ubc-cs.github.io/cpsc340/lectures/L6.pdf

Supervised Learning

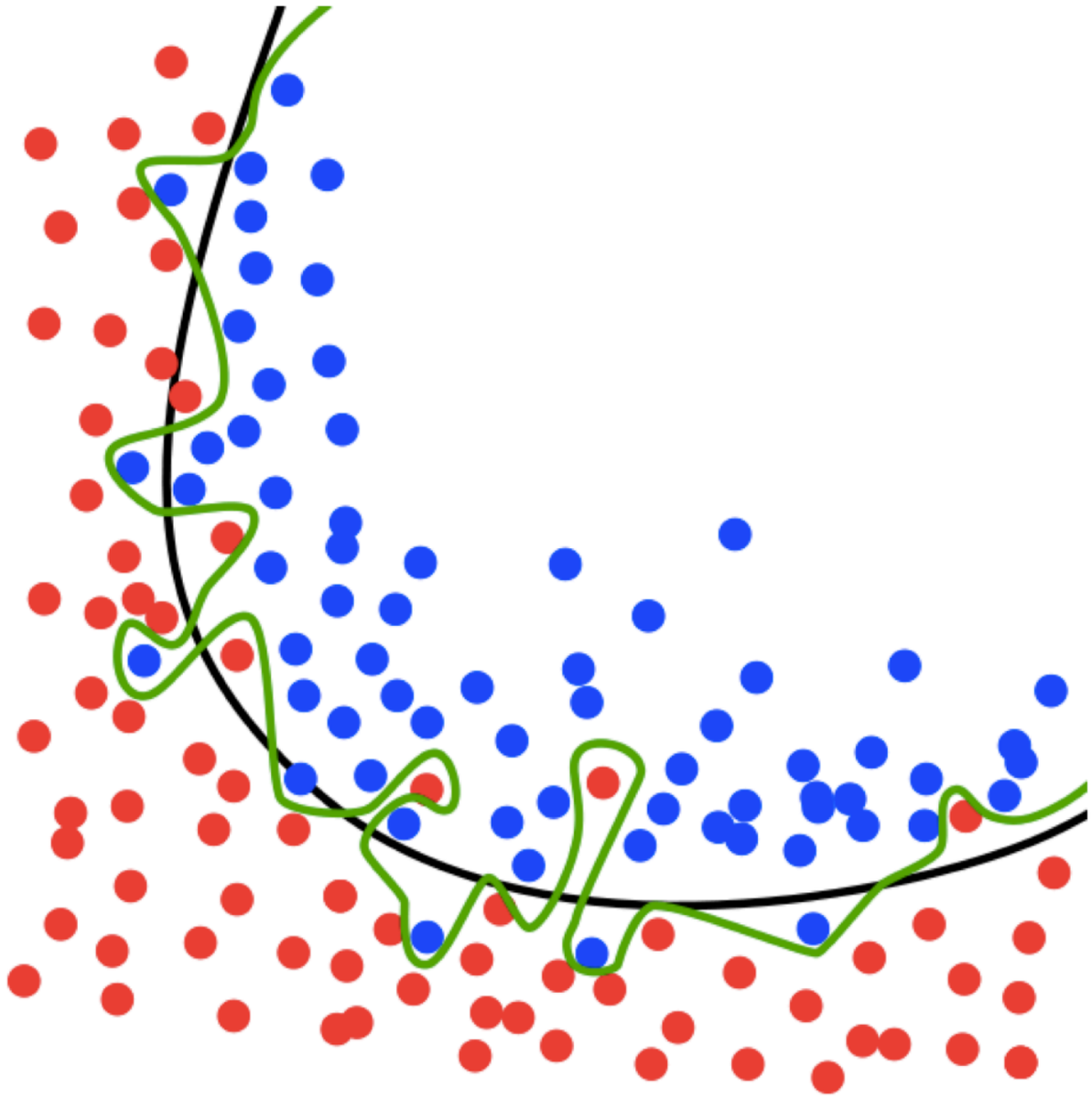
https://www.coursera.org/learn/machine-learning

Regression



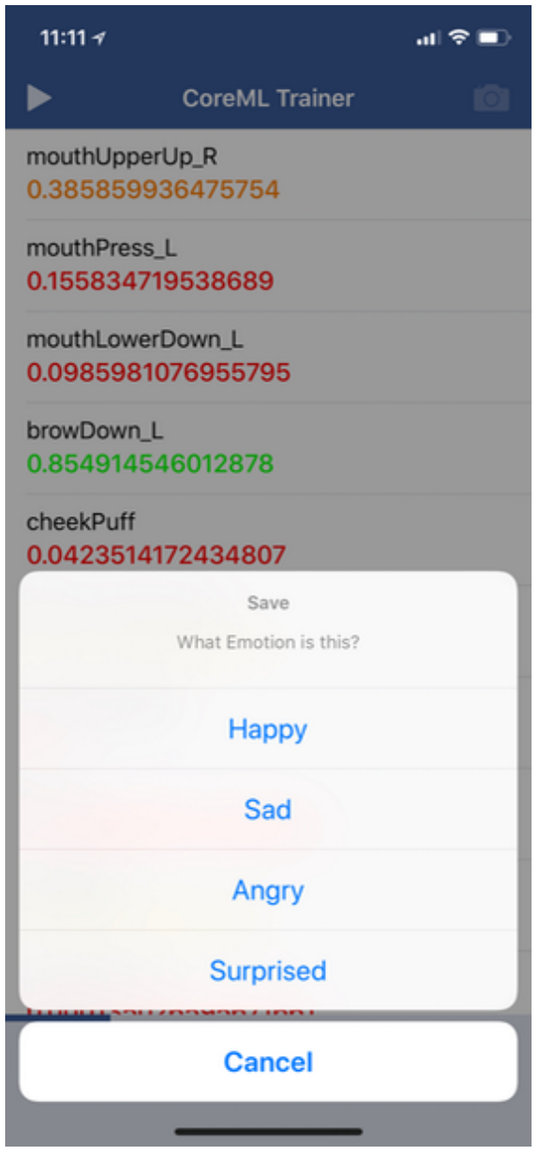
http://edithlaw.ca/teaching/cs480/w19/lectures/3-limits-and-evaluation.pdf

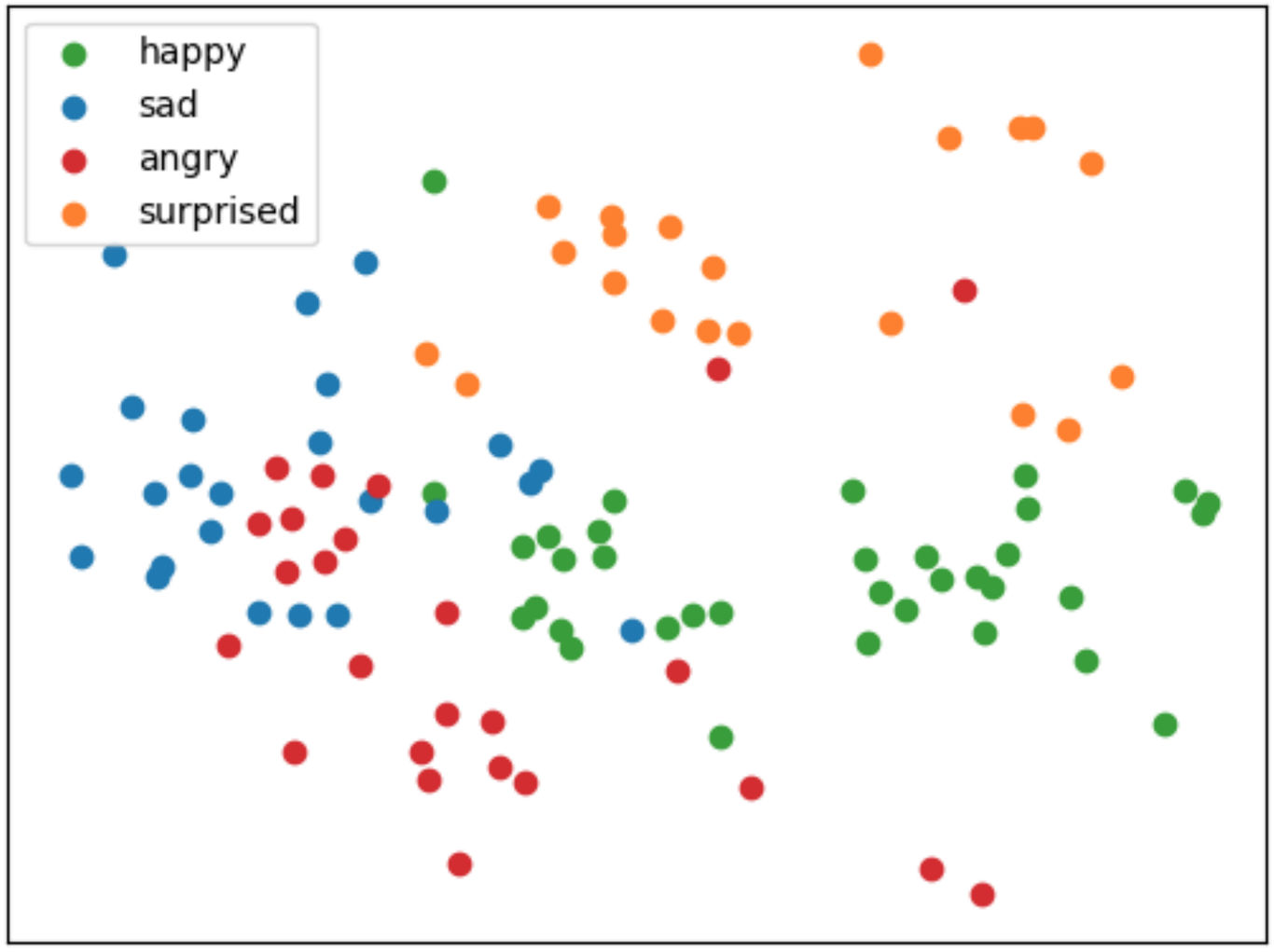
Classification



http://edithlaw.ca/teaching/cs480/w19/lectures/3-limits-and-evaluation.pdf

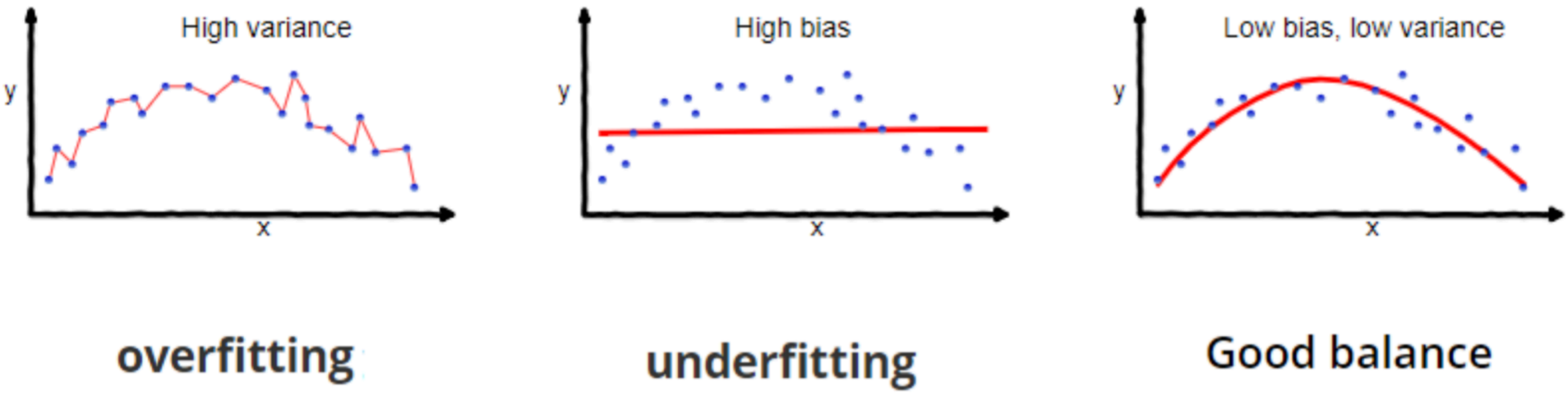
Classification





https://medium.com/nwplusubc/loki-spying-on-user-emotion-c12eafbe24bc

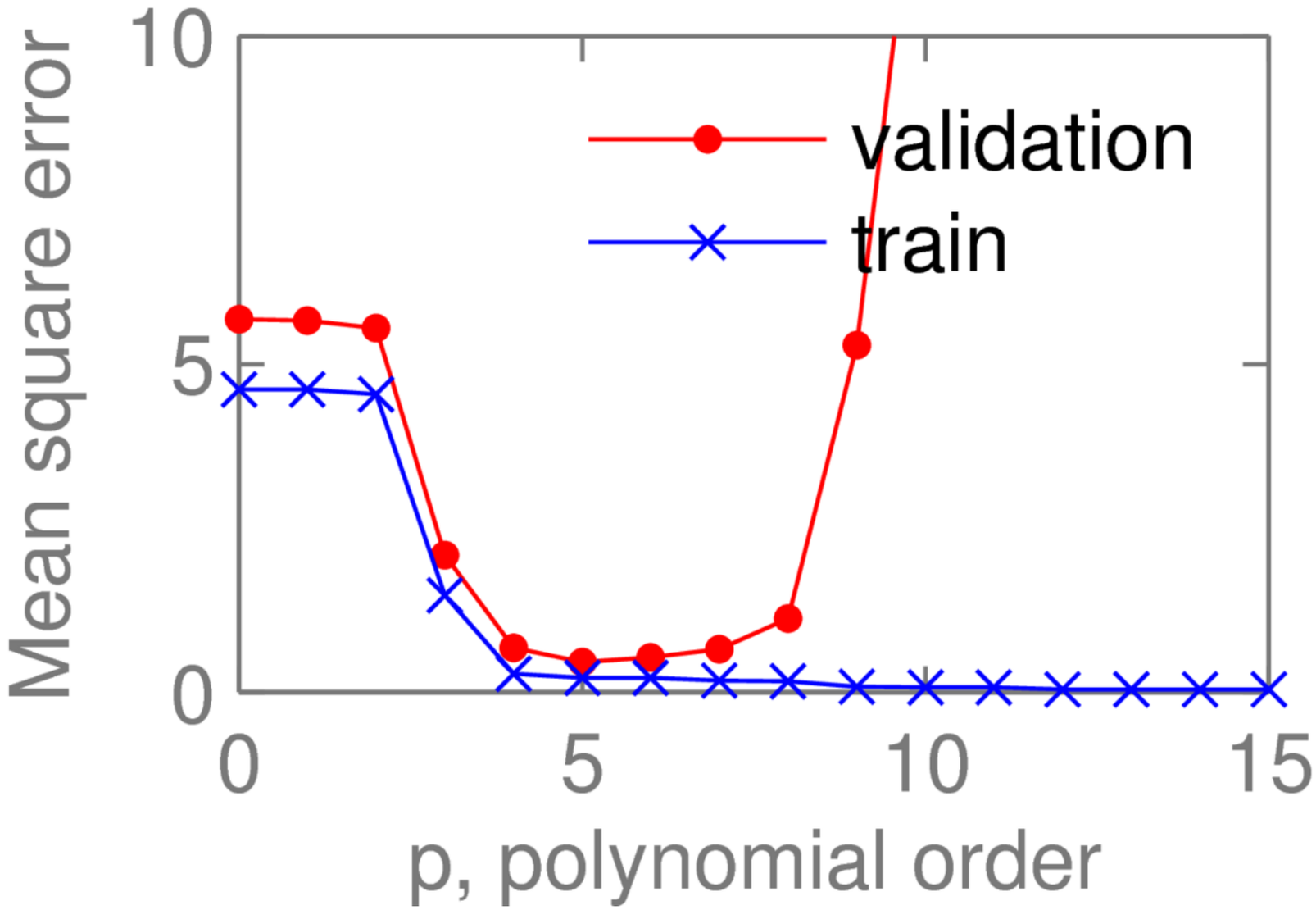
Dangers of Overfitting



overfitting **underfitting** **Good balance**

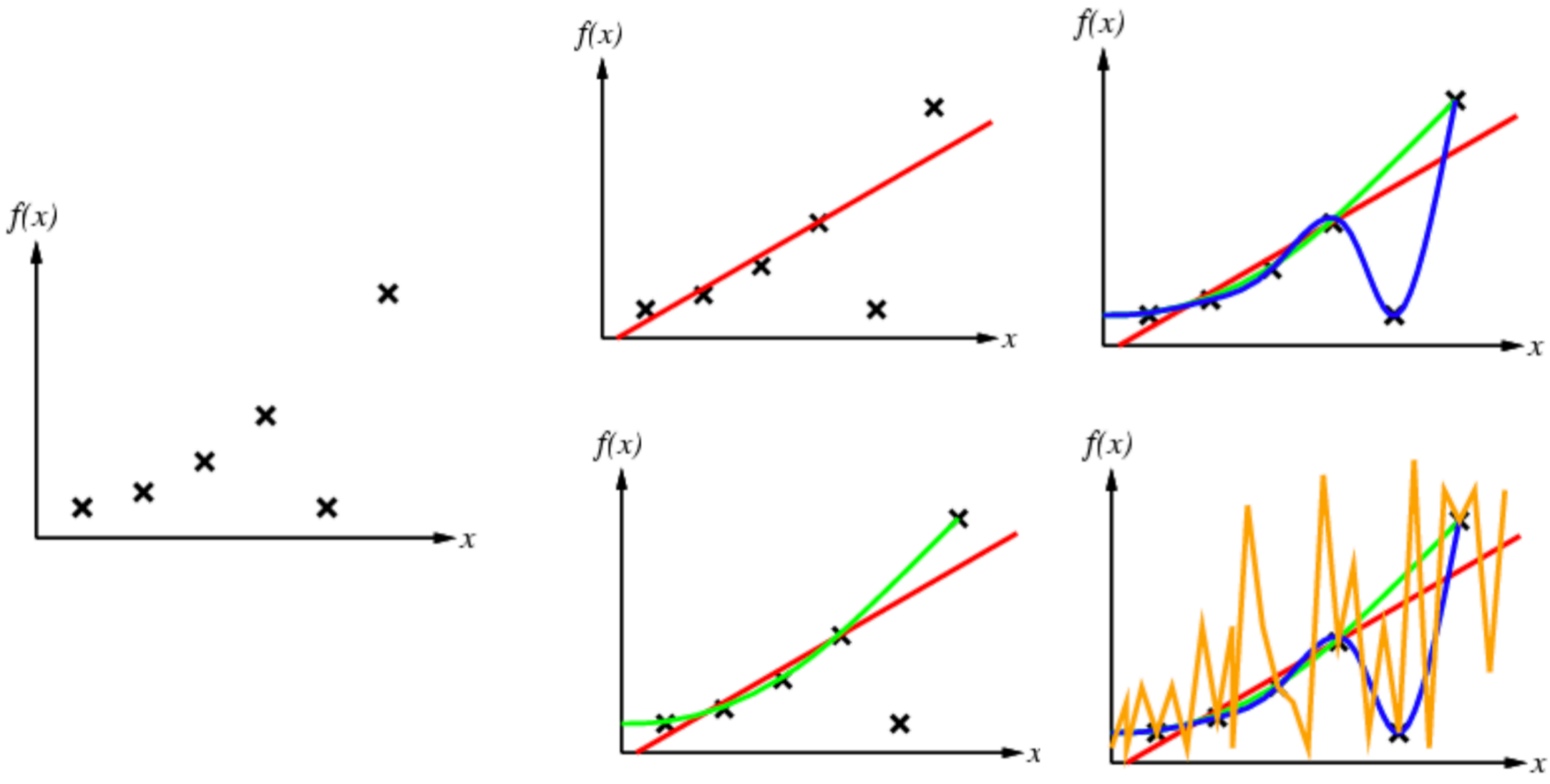
https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229

Dangers of Overfitting



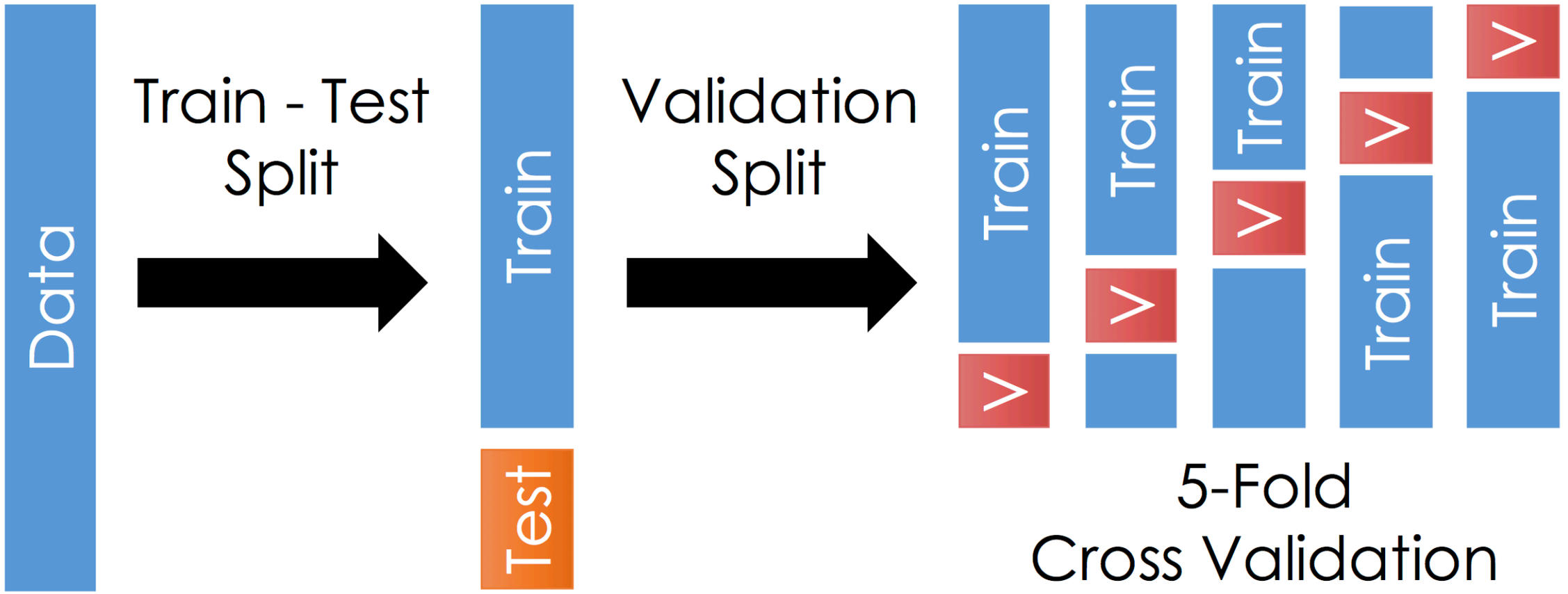
https://www.inf.ed.ac.uk/teaching/courses/mlpr/2017/notes/w2a_train_test_val.html

Dangers of Overfitting



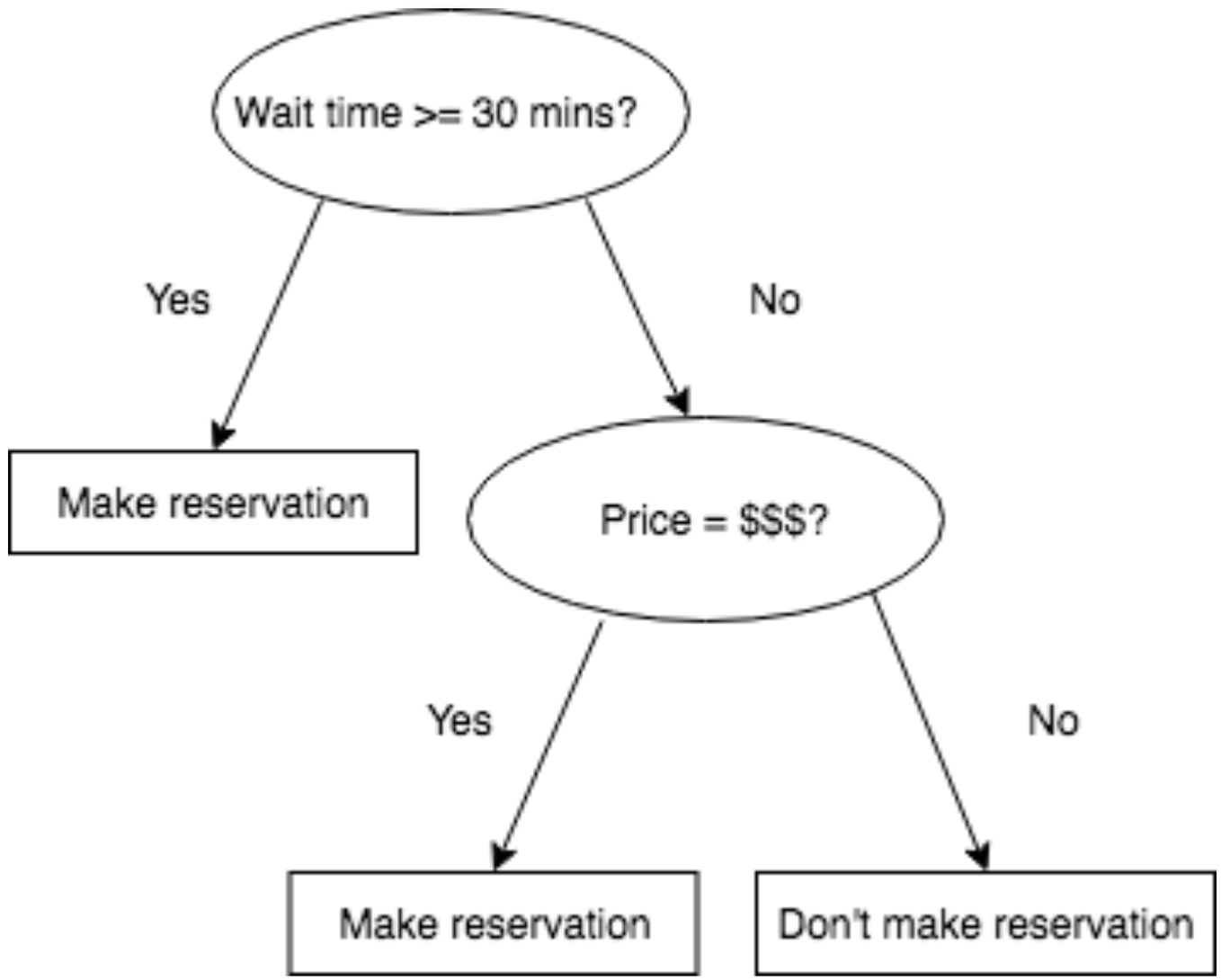
http://edithlaw.ca/teaching/cs480/w19/lectures/3-limits-and-evaluation.pdf

Training / Test / Validation

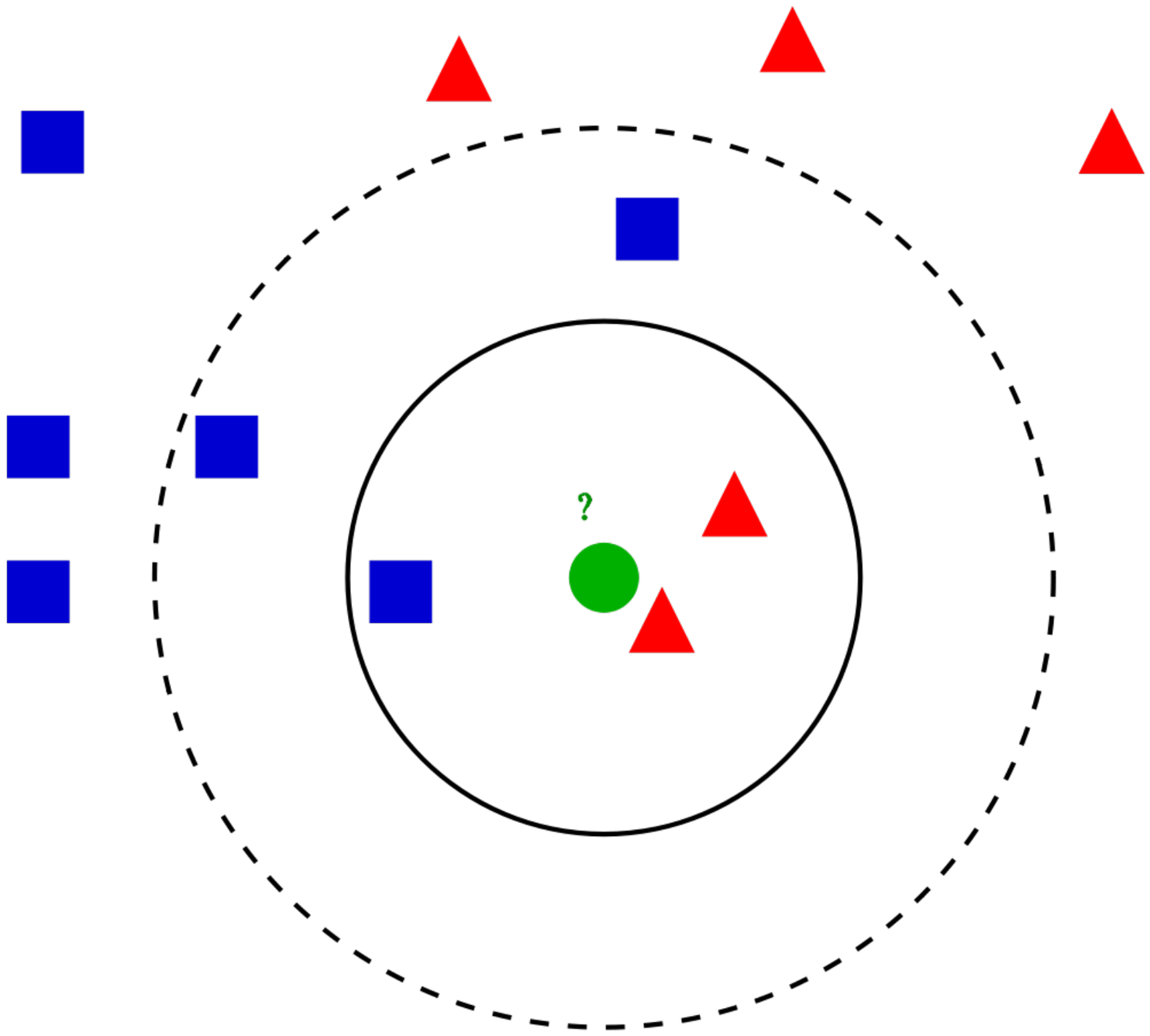


http://www.ds100.org/sp17/assets/notebooks/linear_regression/Cross_Validation_and_the_Bias_Variance_Tradeoff.html

Decision Trees (Boolean Logic)

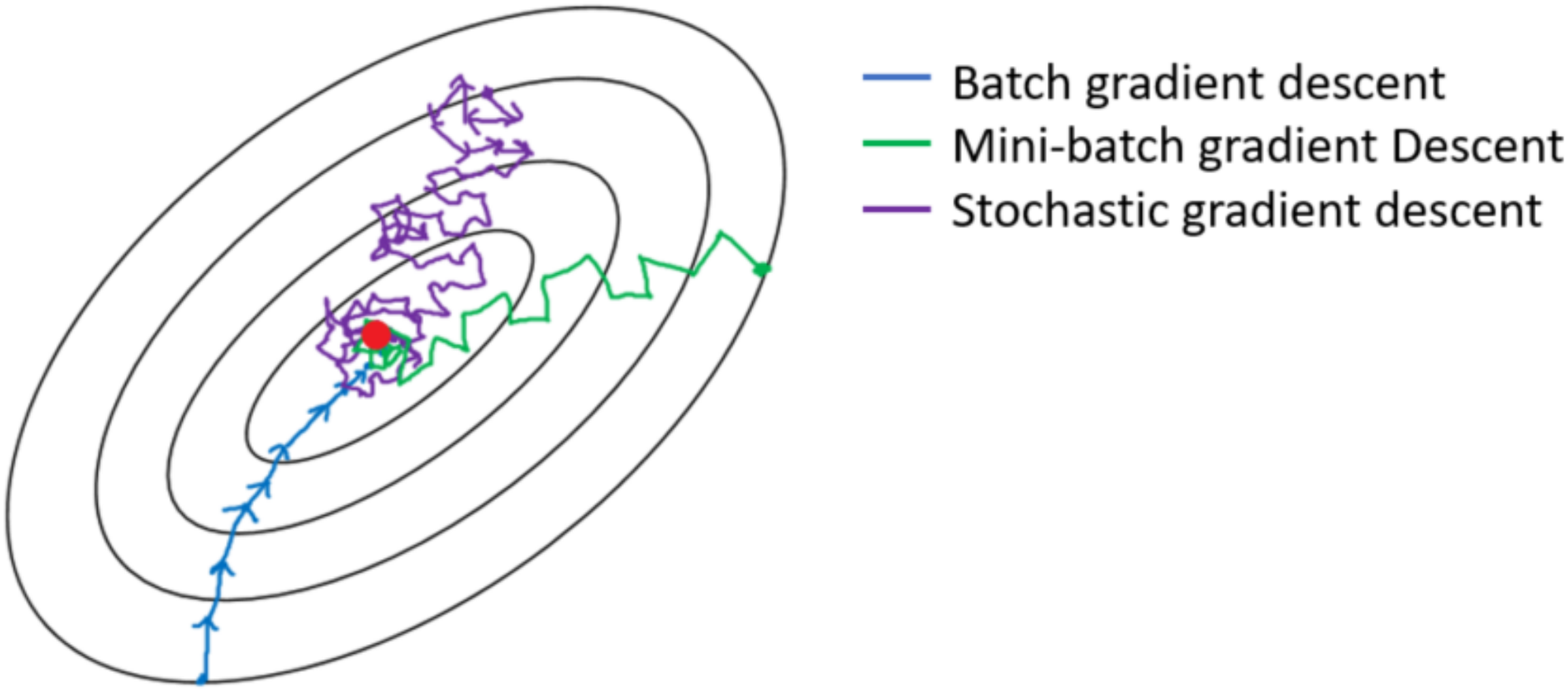


k-Nearest Neighbours



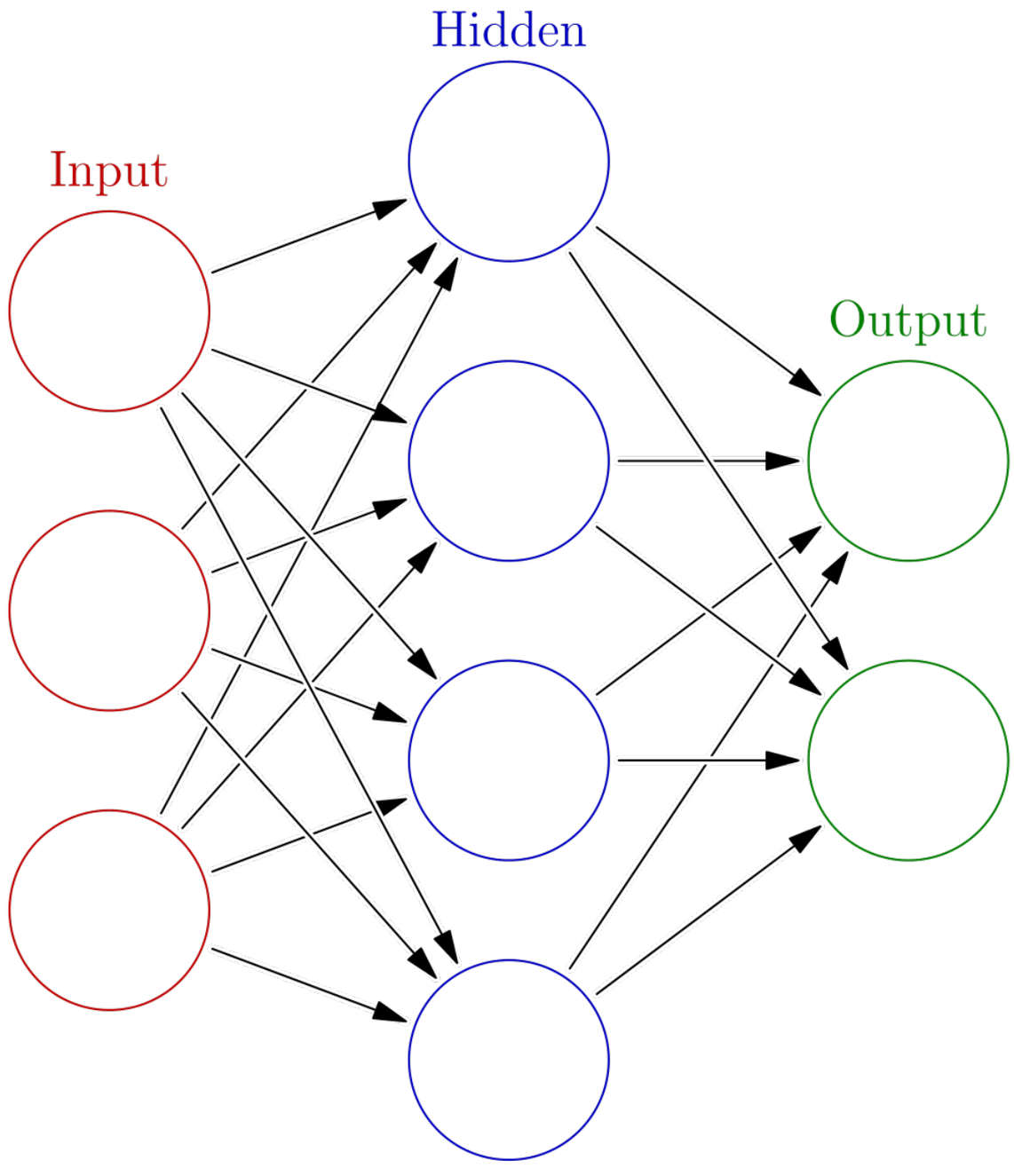
Wikipedia

Stochastic Gradient Descent



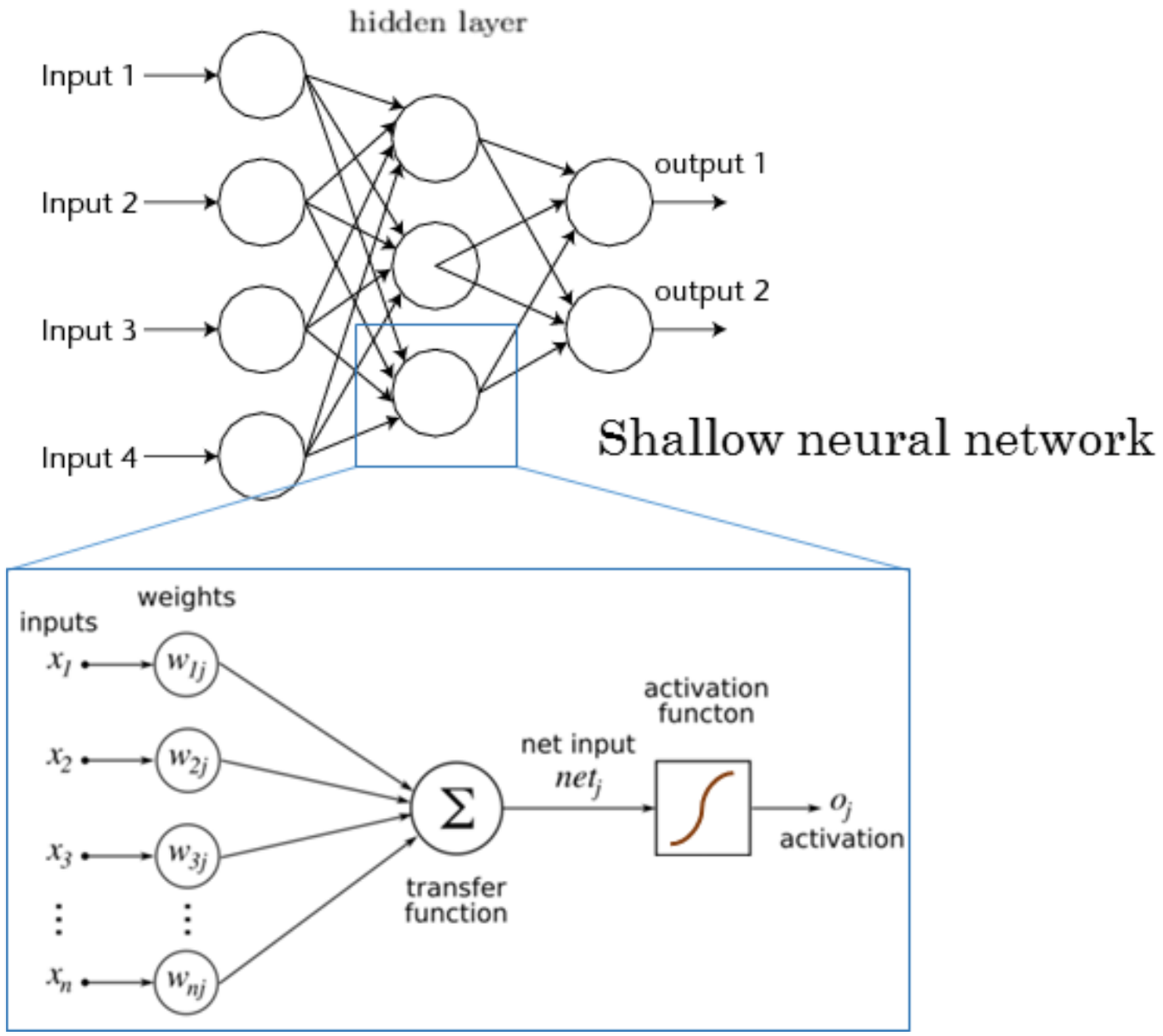
https://towardsdatascience.com/gradient-descent-algorithm-and-its-variants-10f652806a3

Neural Networks



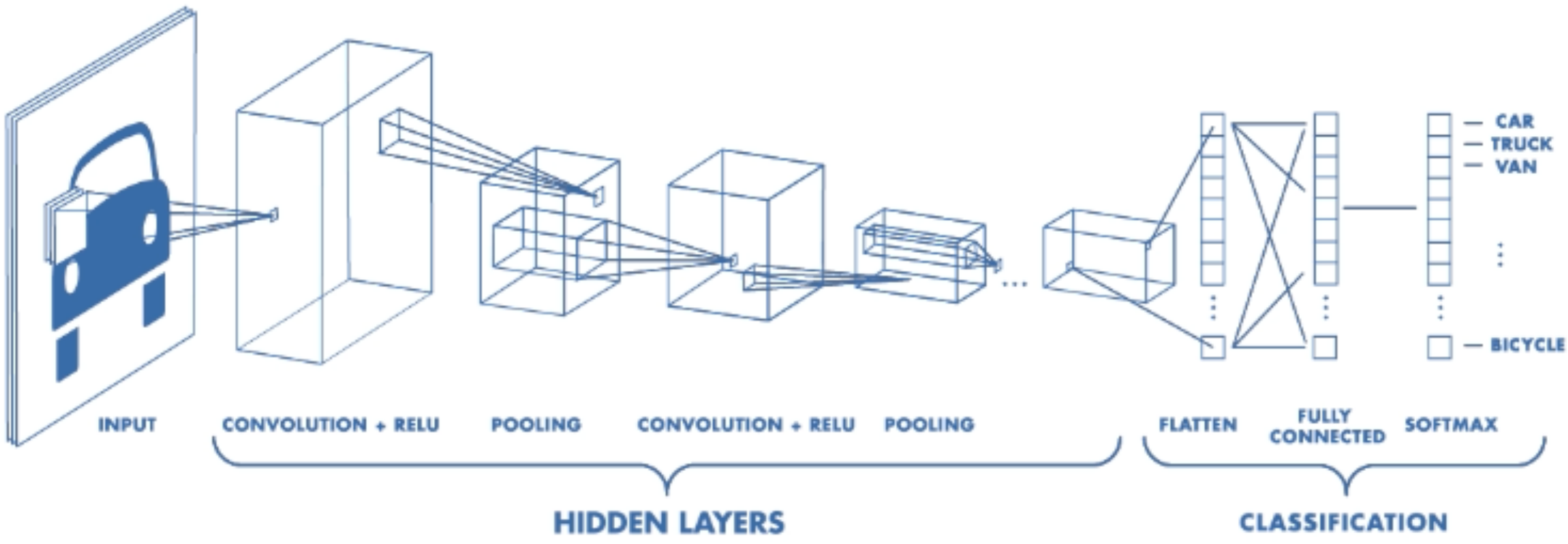
Wikipedia

Neural Networks



https://www.rsipvision.com/exploring-deep-learning/

Convolutional Neural Network



https://medium.freecodecamp.org/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050?gi=62e1aca455b9

Resources

- 3Blue1Brown on neural networks (http://3b1b.co/neural-networks)
- Welch Labs on computer vision + ML (https://youtu.be/i8D90DkCLhI)
- Google's crash course (https://developers.google.com/machine-learning/crash-course/ml-intro)
- CPSC 340 (https://ubc-cs.github.io/cpsc340/)

Questions?

Workshop Time!

- Colaboratory (Jupyter Notebooks + Google Docs)
- Notebook: https://colab.research.google.com/drive/1wRlZsW8pTz94leolVXgY1j0mpKTb_8KZ