Software Development Practical

Computer Vision & Deep Learning

Self-Introduction

Any programming experience?

What is your Python experience?

Any Machine Learning experience?

Any Deep Learning experience?

Overview

Introduction

- Human Perception & Computer Vision
- Math Basics

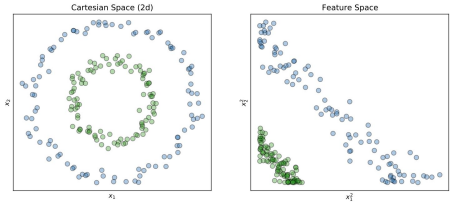
Learning from Data

- Unsupervised
- Supervised

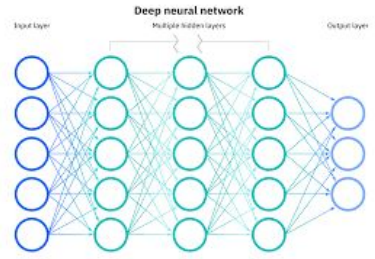
Python Fundamentals

Deep Learning

- Building Blocks
- (Convolutional) Neural Networks



https://sthalles.github.io/a-few-words-on-representation-learning/



https://www.ibm.com/topics/neural-networks

General comments

- Sometimes the math might be overwhelming on a first glance, so if you need help with understanding the concepts please don’t hesitate to contact us. We will then do our best to help you.

- Contact details:

Ming Gui
ming.gui@lmu.de

Johannes Fischer
joh.fischer@lmu.de

Organization

Teaching:

- Graded homework starting from next week

Final project:

- Group in teams of 4
- Focus on a given computer vision task using deep learning methods
- Submit a codebase and a group report before semester ends
- more details will follow…

Human Perception & Computer Vision

What makes vision hard?

What do you see in this picture?

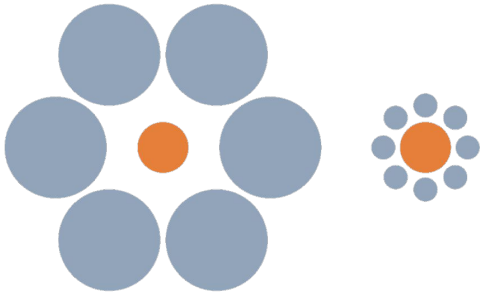
What makes vision hard?

Brain interprets this figure and made objects from the spots.

Is it always correct?

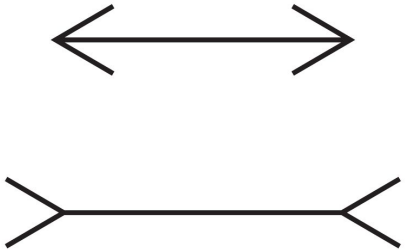
Size

Ebbinghaus illusion



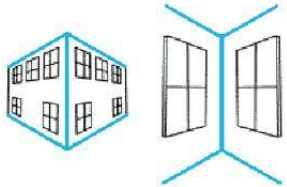
https://nebula.org/blog/de/ebbinghaus-illusion-overestimation-zhu-2020/

Müller-Lyer illusion



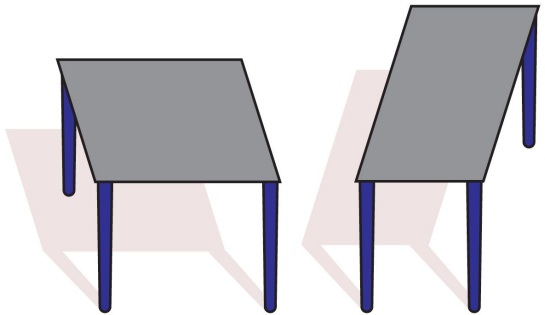
Adapted from Goldstein (2010)

…in real life



Shape

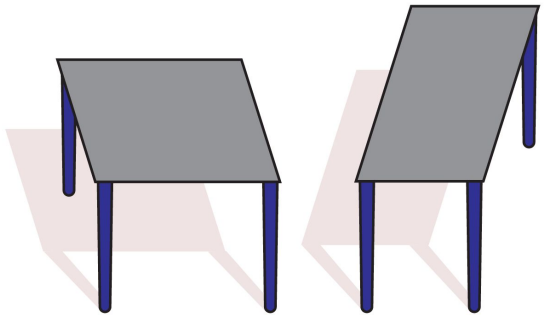
Shepard Tables



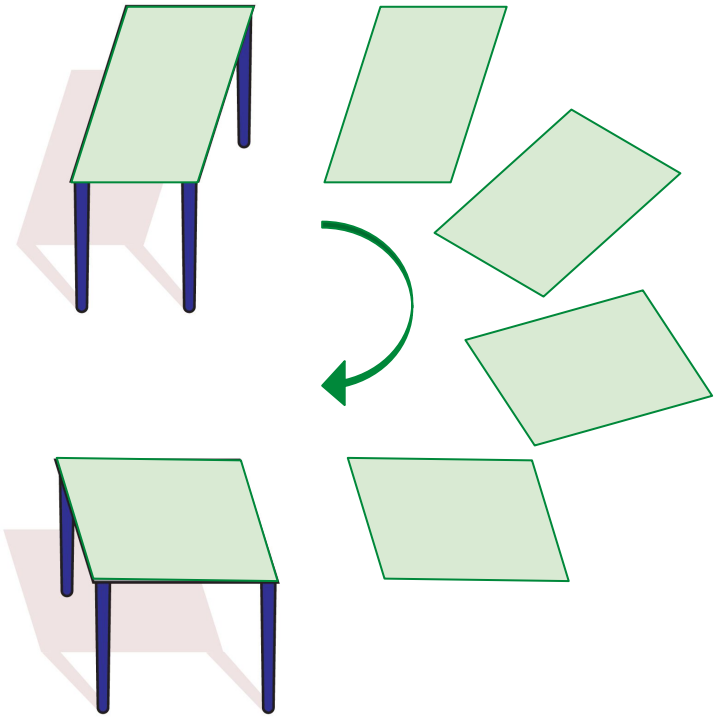
Adapted from Snowden, Thompson, & Troscianko (2011)

Shape

Shepard Tables

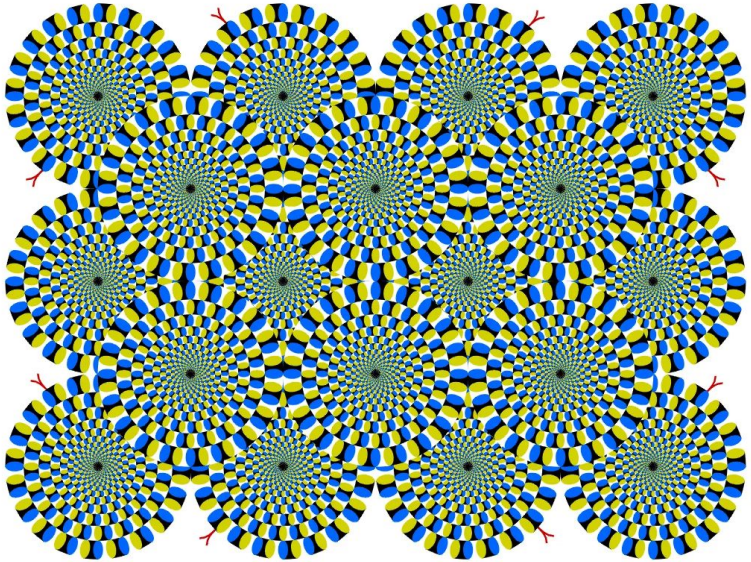


Adapted from Snowden, Thompson, & Troscianko (2011)



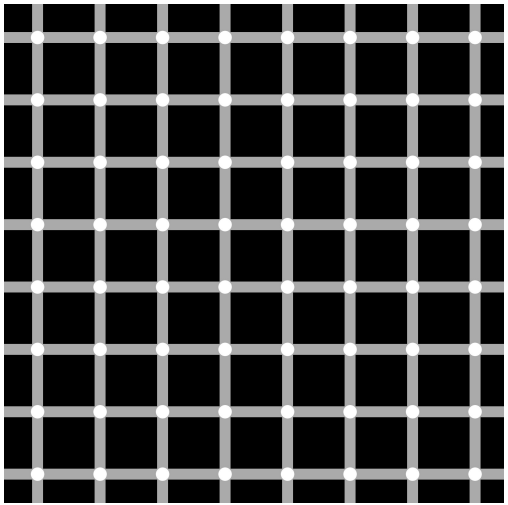
(Motion) Hallucination

Rotating Snakes



https://www.npr.org/sections/13.7/2014/03/24/293740555/the-rotating-snakes-are-all-in-your-mind

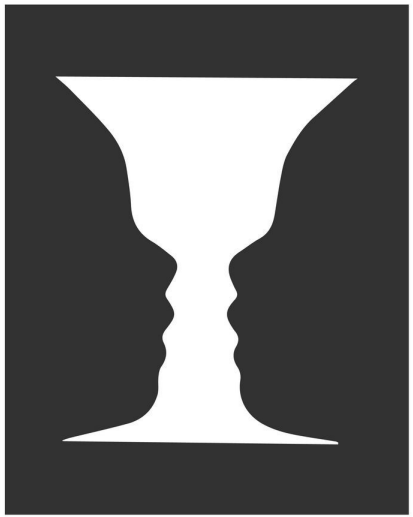
Hermann grid illusion



https://en.wikipedia.org/wiki/Grid_illusion

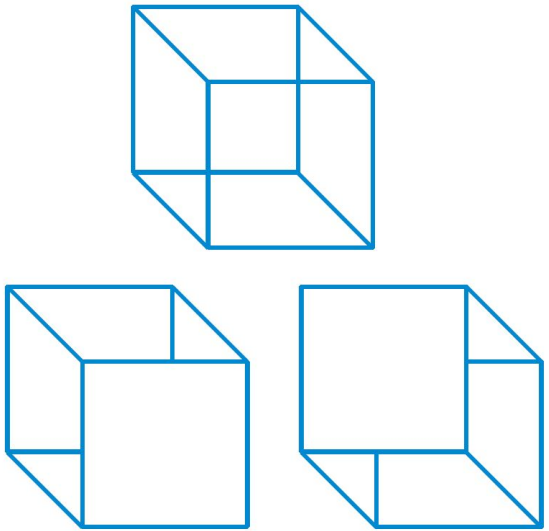
Ambiguity

Rubins vase



https://dorsch.hogrefe.com/stichwort/rubin-vase

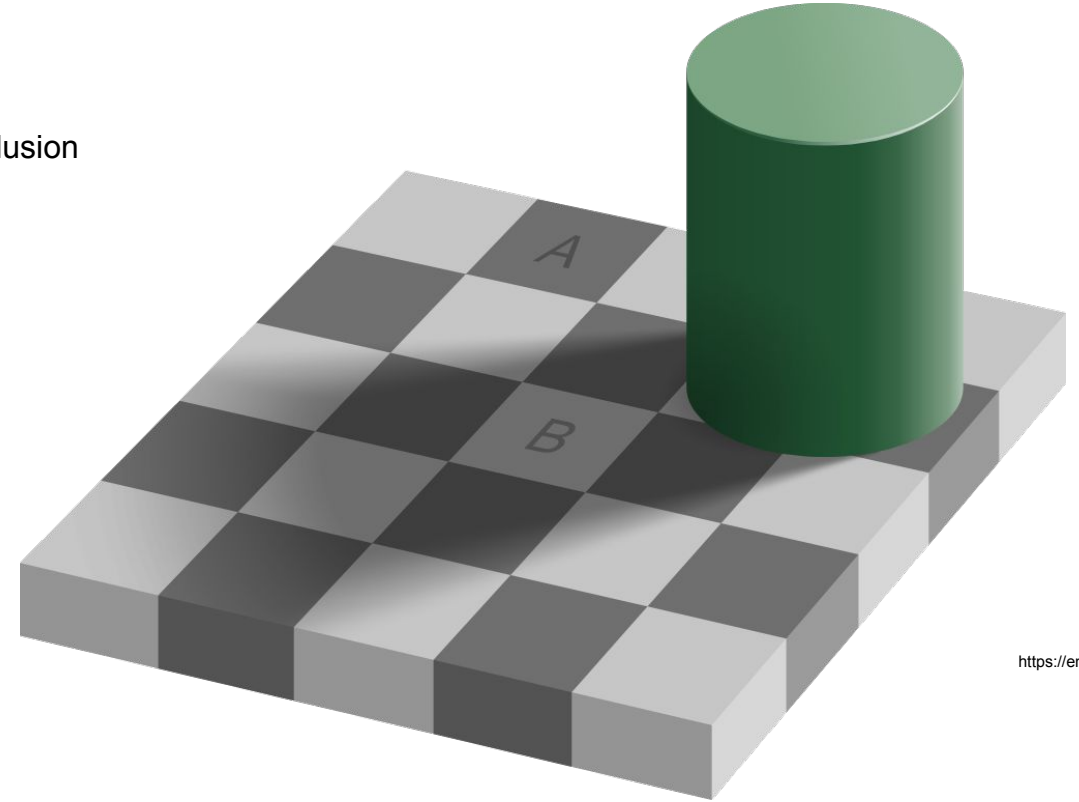
Necker Cube



Adapted from Snowden, Thompson, & Troscianko (2011)

Color

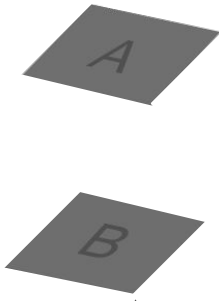
Checker shadow illusion
(E. Adelson)



https://en.wikipedia.org/wiki/Checker_shadow_illusion

Color

Checker shadow illusion
(E. Adelson)



https://en.wikipedia.org/wiki/Checker_shadow_illusion

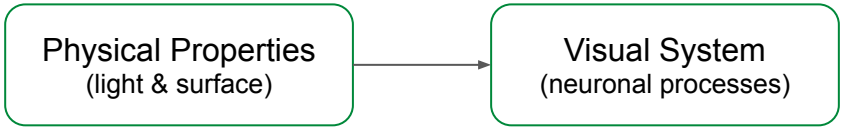
Perception vs. Measurement

Brain makes **assumptions** about the world (e.g. depth)

⇒ Perception does not represent the physical world correctly

⇒ Includes psychological interpretations

From Goldstein (2010)



Center of vision

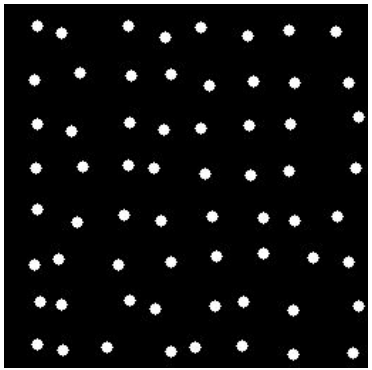
⇒ Further away from our central vision, our vision loses the ability to see fine detail

Adapted from Snowden, Thompson, & Troscianko (2011)

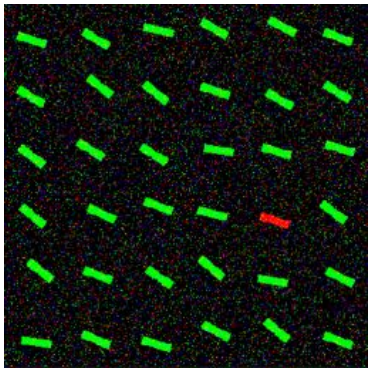
⇒ We need to move our eyes. How we decide where to look next?

Guiding our Perception

Bottom up with salient features



http://www.scholarpedia.org/article/Visual_salience



http://www.scholarpedia.org/article/Visual_salience

Guiding our Perception

Top down with attention

Read every other word, starting from the 1st or 2nd word:

Visual Human search perception is plays a an type important of role perceptual in task the requiring area attention of that visualization. typically An involves understanding an of active perception scan can of significantly the improve visual both environment ...

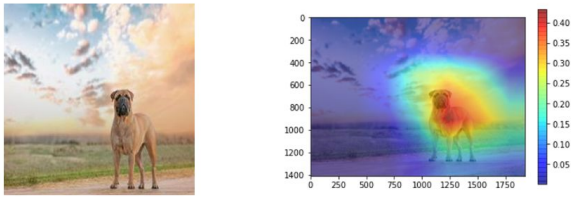
visual query: read the black text

visual search: carry out search to find patterns to resolve the query

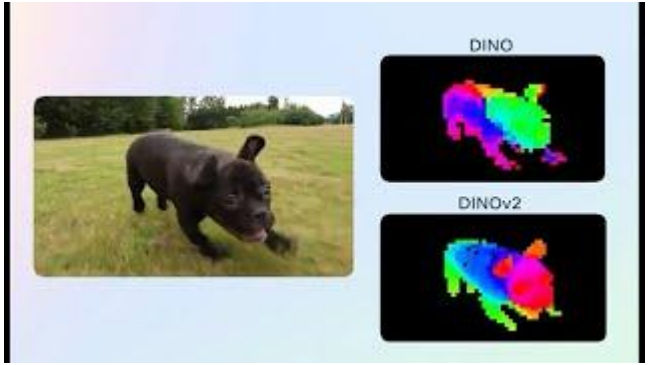
Slide credit adapted from https://mycourses.aalto.fi/pluginfile.php/925138/mod_resource/content/3/CS-E4840-2019-07.pdf

Excursus: Attention in Deep Learning

⇒ Canonical paper in 2017, introducing Attention in Deep Learning (Vaswani et al., 2017)



An, J., & Joe, I. (2022). Attention Map-Guided Visual Explanations for Deep Neural Networks. *Applied Sciences*, *12*(8), 3846. MDPI AG. Retrieved from http://dx.doi.org/10.3390/app12083846



Video from
https://ai.facebook.com/blog/dino-v2-computer-vision-self-supervised-learning/
Paper: Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., ... & Bojanowski, P. (2023). DINOv2: Learning Robust Visual Features without Supervision. *arXiv preprint arXiv:2304.07193*.

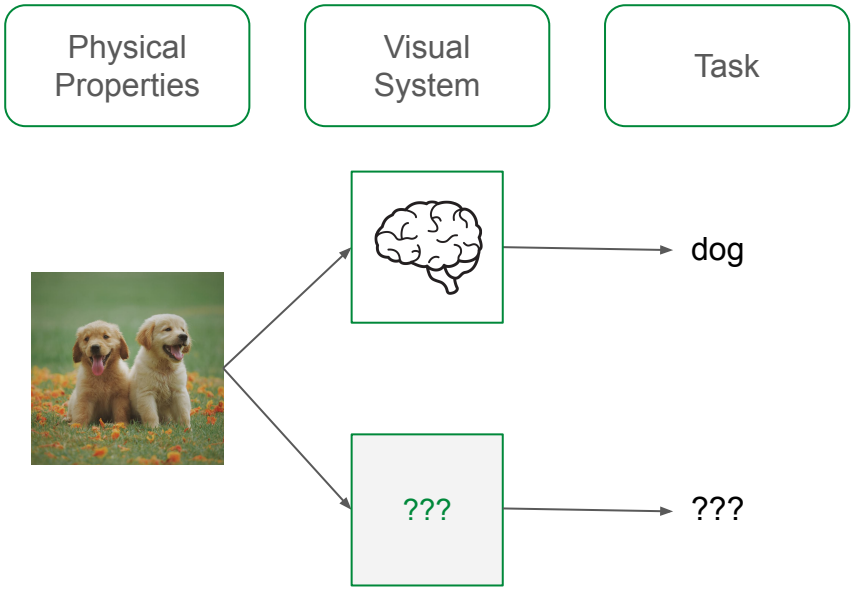
Recap

Human Vision

- not as easy as it seems
- no un-biased “window” to the world
- combination of
  - *physical* properties
  - *filter* operations
  - *psychological* interpretations

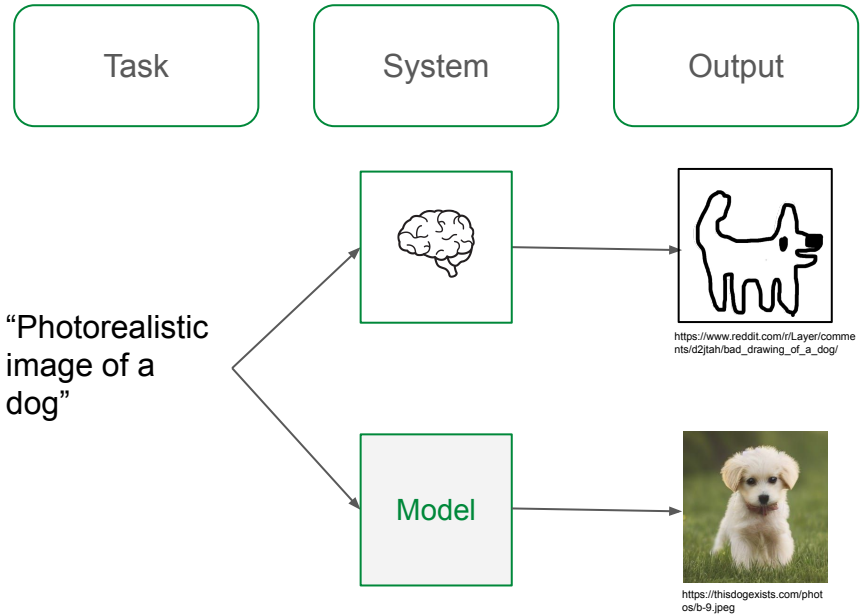
Computer Vision

A camera is a good analogy for the eye, as an image just represents physical properties. But what about the semantic information present in an image?



⇒ We will learn more about that in the upcoming lectures…

Generative Modeling



https://research.nvidia.com/labs/toronto-ai/VideoLDM/

Math Basics

Math notation

A symbolic representation of mathematical ideas and concepts using a set of symbols, characters, and mathematical operators

Why is it important?

- **Clarity and Consistency**: Notation provides a clear and concise way to express mathematical concepts and ideas.
- **Efficient communication**: With a standardized notation, individuals can quickly understand and communicate complex mathematical ideas without the need for lengthy explanations.

Math notation: Sum and product

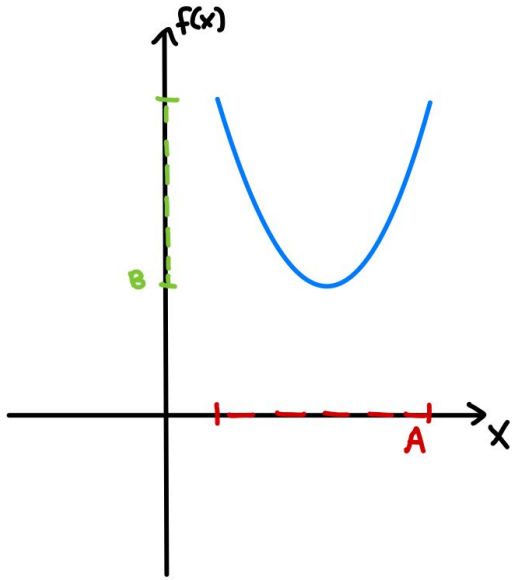
$$a_1 + a_2 + \ldots + a_n = \sum_{i=1}^{n} a_i = \sum_{i} a_i$$

$$a_1 \times a_2 \times \ldots \times a_n = \prod_{i=1}^{n} a_i = \prod_{i} a_i$$

Functions

A function ***f*** assigns to each element of its definition set ***A*** exactly one element of its target set ***B***, this is written as:

$$f\colon A \to B\,,$$
$$a \mapsto f(a)\,.$$



Run an ice-cream shop!

You want to predict how many ice-cream you sell based on the temperature and whether it rains.

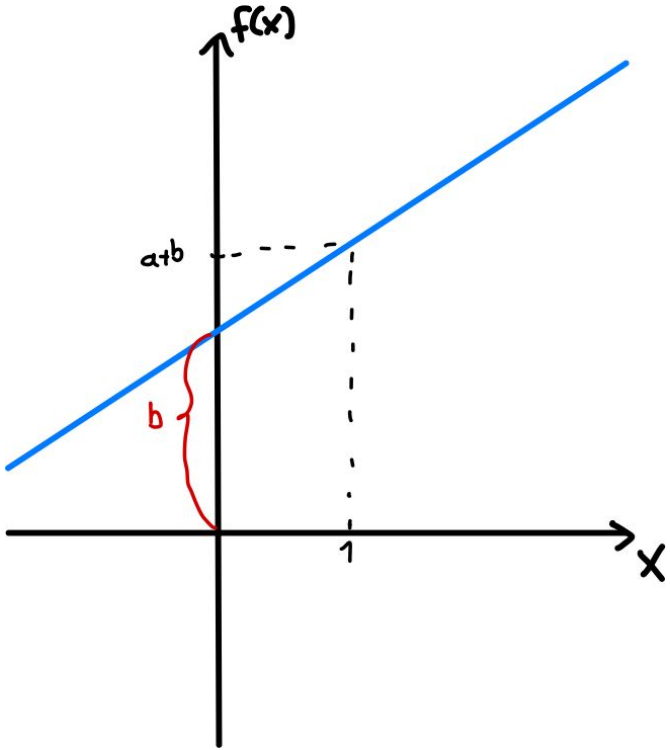
We can construct the following model using a function:

f(temperature, rain) **=** number of ice-cream

Functions in one-dimension

Linear:

$$f_{a,b}\colon \mathbb{R} \to \mathbb{R}\,,$$
$$x \mapsto f(x) = a \cdot x + b\,.$$



Functions in one-dimension

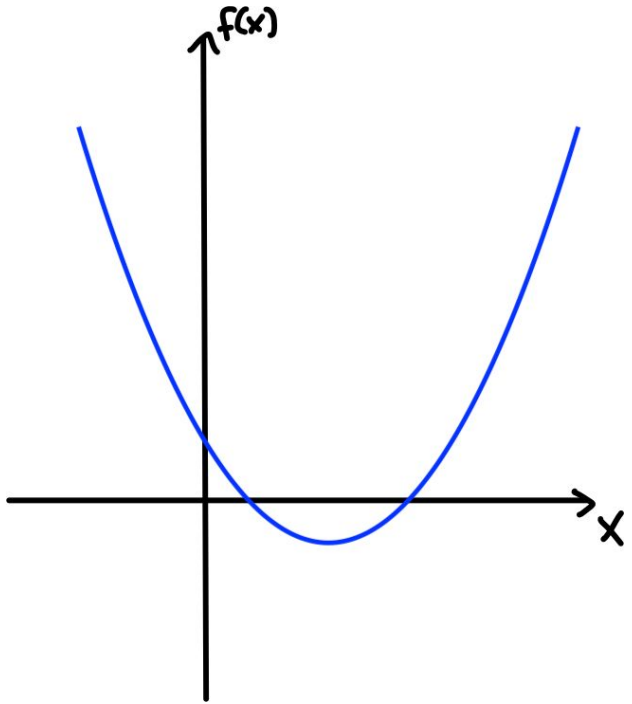
Linear:

$$f_{a,b}\colon \mathbb{R} \to \mathbb{R}\,,$$
$$x \mapsto f(x) = a \cdot x + b\,.$$

Polynomial:

$$f_{a,b}\colon \mathbb{R} \to \mathbb{R}\,,$$
$$x \mapsto f(x) = \sum_i a_i x^i + b\,.$$

Exponential, sinusodal, sigmoid …



2nd-order polynomial

Functions in one-dimension

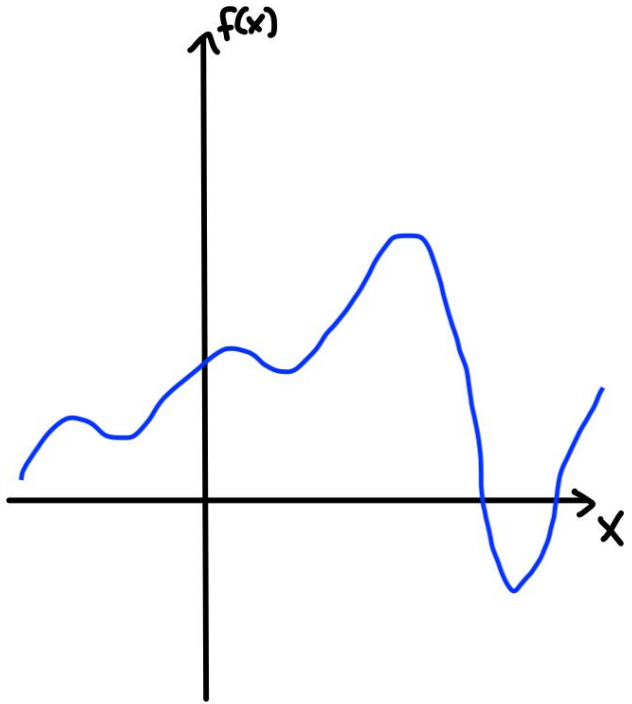
Linear:

$$f_{a,b}\colon \mathbb{R} \to \mathbb{R}\,,$$
$$x \mapsto f(x) = a \cdot x + b\,.$$

Polynomial:

$$f_{a,b}\colon \mathbb{R} \to \mathbb{R}\,,$$
$$x \mapsto f(x) = \sum_i a_i x^i + b\,.$$

Exponential, sinusodal, sigmoid …

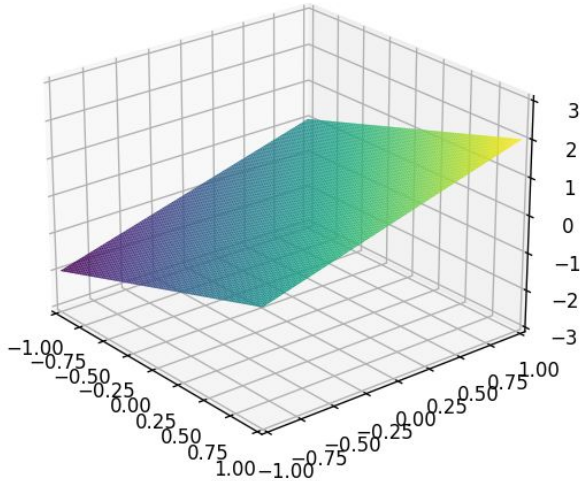


multi-order polynomial

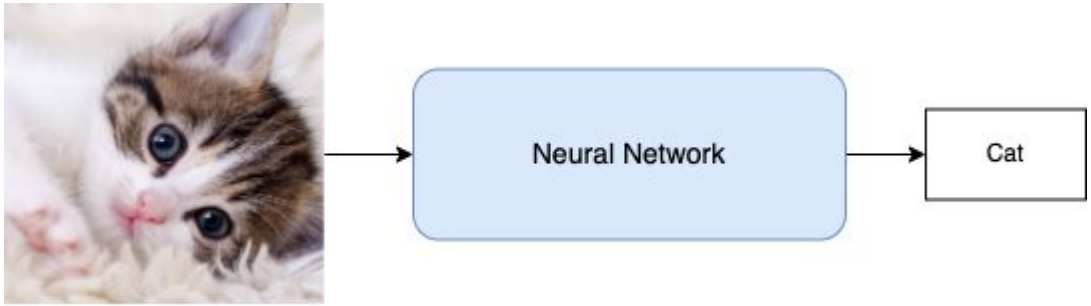
Multi-dimensional functions

2D Linear:

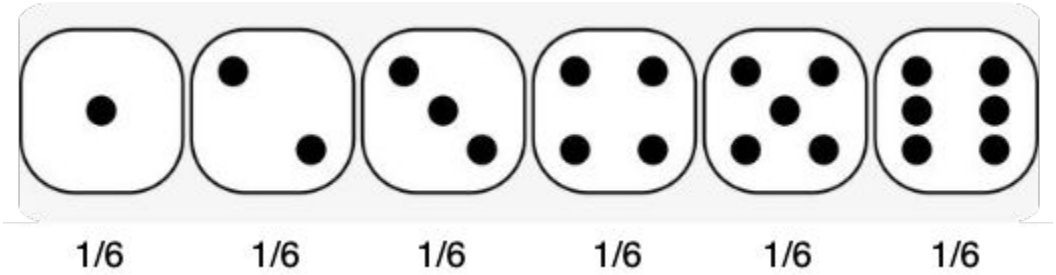
$$f(x_1, x_2) = a_1 x_1 + a_2 x_2 + b$$



Neural networks are also functions



Probability



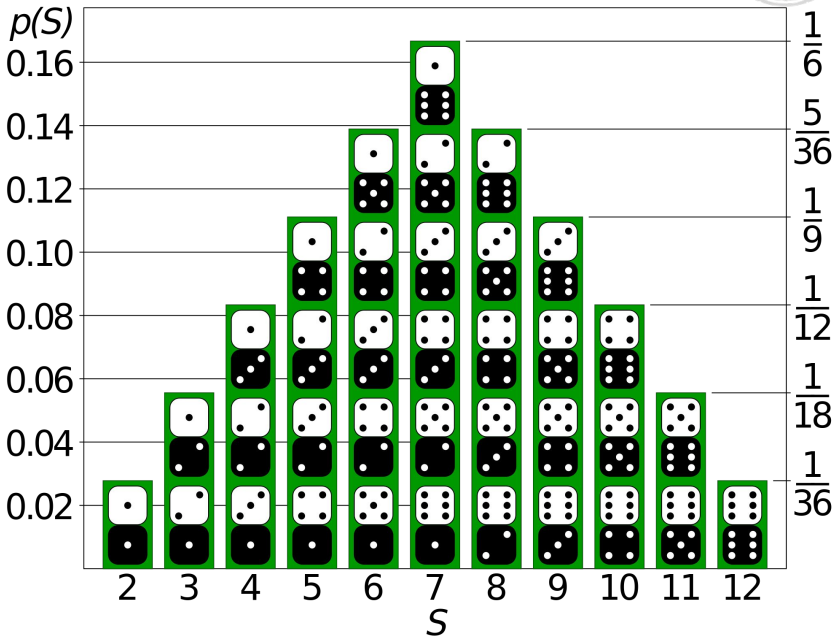
Assuming x is a random variable: $\mathbf{P}(x)$

In case of a die throw: $\mathbf{P}(x = k) = \frac{1}{6}, k = 1, \cdots, 6$

Conditional probability

Consider throwing two dice

$$\mathbf{P}(x_1 + x_2 = 5) = \frac{4}{36}$$



https://math.stackexchange.com/questions/1204396/why-is-the-sum-of-the-rolls-of-two-dices-a-binomial-distribution-what-is-define

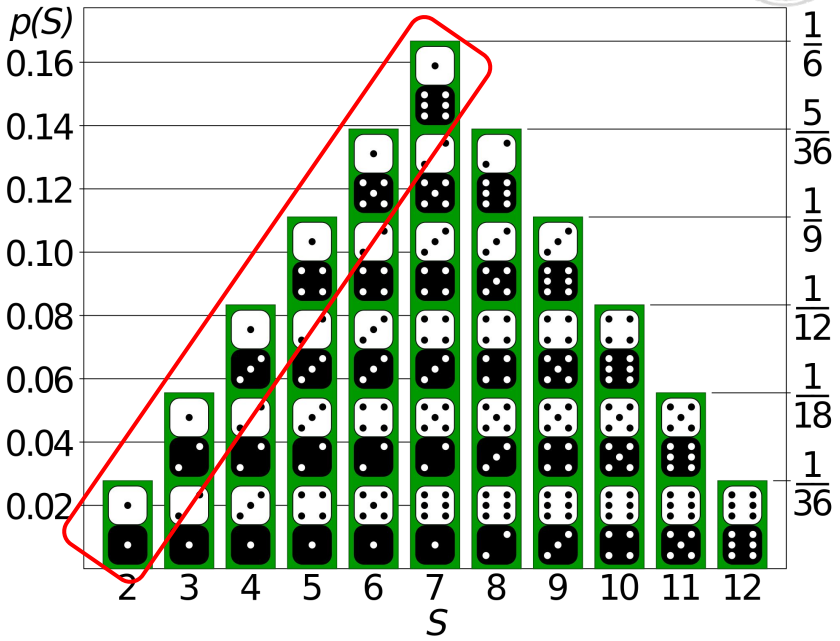
Conditional probability

Consider throwing two dice

$$\mathbf{P}(x_1 + x_2 = 5) = \frac{4}{36}$$

The conditioning changes the probability

$$\mathbf{P}(x_1 + x_2 = 5 \mid x_1 = 1) = \frac{1}{6}$$



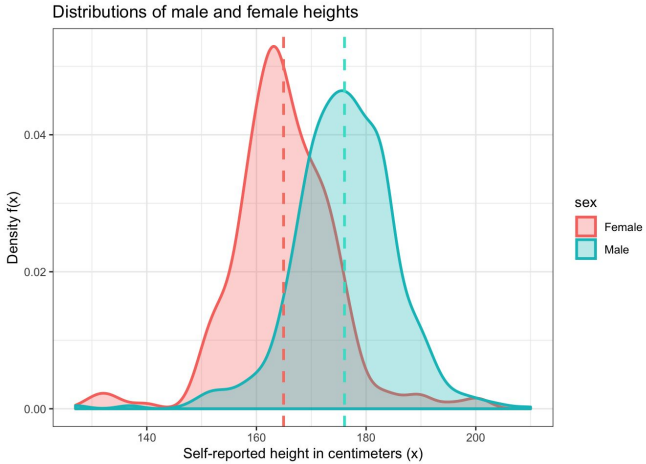
https://math.stackexchange.com/questions/1204396/why-is-the-sum-of-the-rolls-of-two-dices-a-binomial-distribution-what-is-define

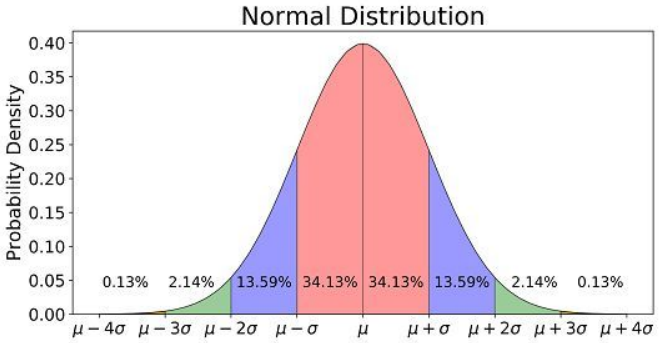
Continuous probability

Not all values are discrete (height, rainfall…)

The likelihood is defined by the probability density function

Gaussian distribution / Normal distribution $\mathcal{N}(\mu, \sigma^2)$





https://bookdown.org/gabriel_butler/ECON41Labs/tutorial-6-the-normal-distribution.html
https://www.thinkingondata.com/normal-distribution-what-means-in-numbers/

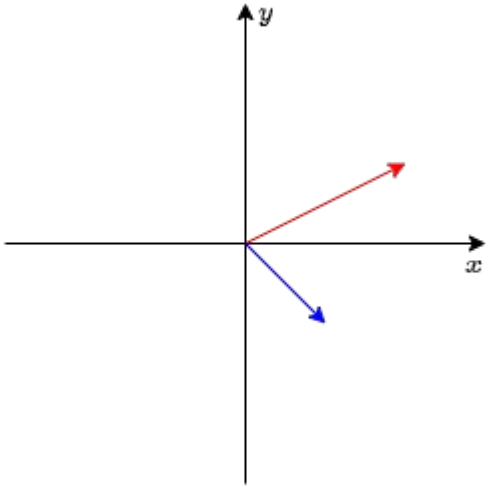
Vectors, matrices and tensors

We might need more than one number to describe the circumstance

A vector is represented as a list of numbers, where each number represents the magnitude of the vector in a particular direction.

$$a = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad b = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$\text{Norm}(a) = \sqrt{\sum_i a_i^2} = \sqrt{2^2 + 1^2} = \sqrt{5}$$

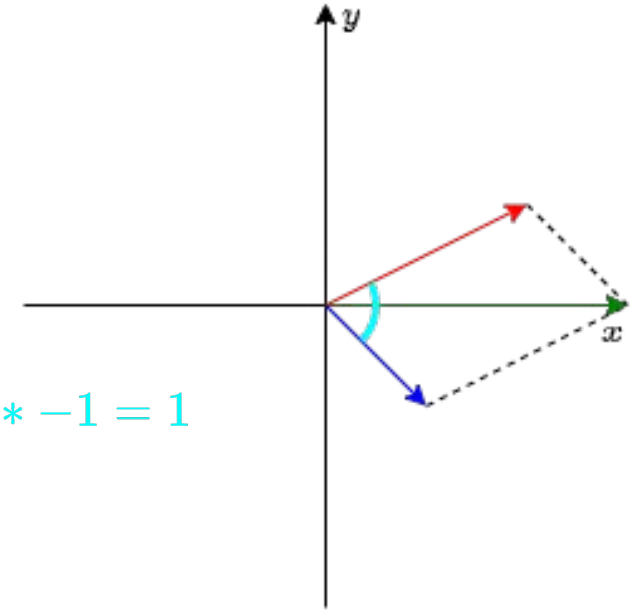


Vectors calculation

$$a = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad b = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Add $$a + b = \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

Inner product $$a \cdot b = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 2 * 1 + 1 * -1 = 1$$



Cosine similarity $$\frac{a \cdot b}{\|a\|\|b\|} = \frac{1}{\sqrt{5}\sqrt{2}} = \frac{1}{\sqrt{10}}$$

$$\sin^{-1}(1/\sqrt{10}) = 71.57°$$

Matrix

A matrix is just a table of scalars:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} \in \mathbb{R}^{n \times m}$$

And its transpose:

$$A^\top = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

Matrix

A matrix is just a table of scalars:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} \in \mathbb{R}^{n \times m}$$

And its transpose:

$$A^\top = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

$$A^\top = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$$

Matrix multiplication

For $\boldsymbol{A} \in \mathbb{R}^{k \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times m}$:

$$\boldsymbol{A} \cdot \boldsymbol{B} = \begin{pmatrix} — & \boldsymbol{a}_{1\bullet} & — \\ — & \boldsymbol{a}_{2\bullet} & — \\ & \vdots & \\ — & \boldsymbol{a}_{n\bullet} & — \end{pmatrix} \cdot \begin{pmatrix} | & | & & | \\ \boldsymbol{b}_{\bullet 1} & \boldsymbol{b}_{\bullet 2} & \dots & \boldsymbol{b}_{\bullet m} \\ | & | & & | \end{pmatrix}$$

$$= \begin{pmatrix} \langle a_{1\bullet}, b_{\bullet 1} \rangle & \dots & \langle a_{1\bullet}, b_{\bullet m} \rangle \\ \vdots & \ddots & \vdots \\ \langle a_{n\bullet}, b_{\bullet 1} \rangle & \dots & \langle a_{n\bullet}, b_{\bullet m} \rangle \end{pmatrix} \in \mathbb{R}^{k \times m}$$

Matrix multiplication

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \qquad A \cdot B = \begin{bmatrix} \varphi & \cdot \\ \cdot & \cdot \end{bmatrix}$$

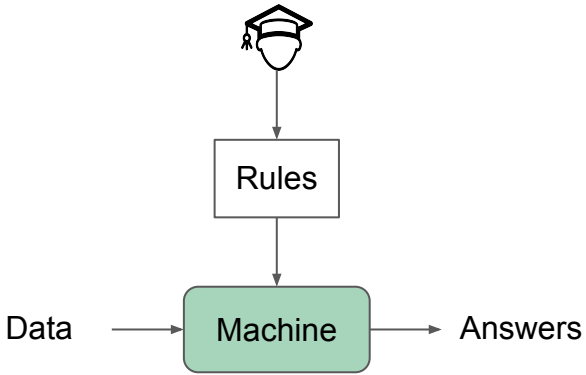
$$\varphi = \begin{bmatrix} 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 7 \end{bmatrix} = \left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 \\ 7 \end{pmatrix} \right\rangle = 1 * 5 + 2 * 7 = 19$$

Learning from Data

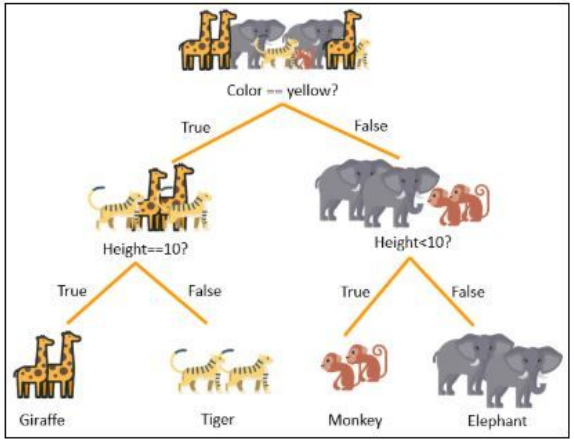
Why should we learn from data?

Classical Programming Paradigm

- Expert defines rules
- Input data
- Output answer



Example: Classify elefant, monkey, giraffe, tiger

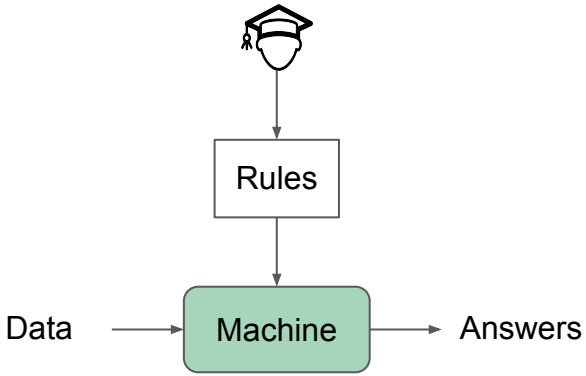


https://www.simplilearn.com/tutorials/machine-learning-tutorial/decision-tree-in-python

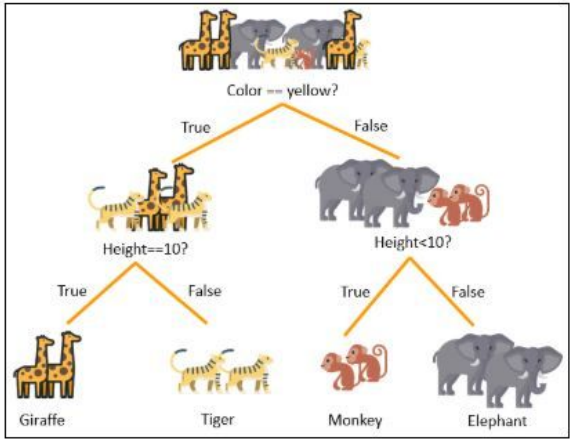
Why should we learn from data?

Classical Programming Paradigm

- Expert defines rules
- Input data
- Output answer



Example: Classify elefant, monkey, giraffe, tiger



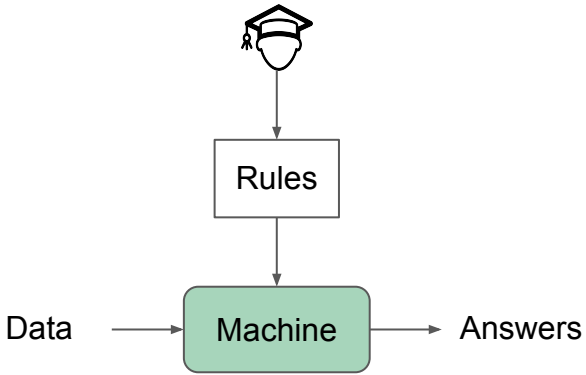
https://www.simplilearn.com/tutorials/machine-learning-tutorial/decision-tree-in-python

⇒ **What can be potential drawbacks?**

Why should we learn from data?

Classical Programming Paradigm

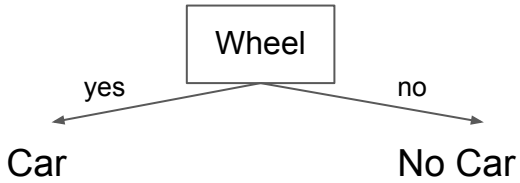
- Expert defines rules
- Input data
- Output answer



Example: Write program to detect cars in natural images.

image: Freepik.com

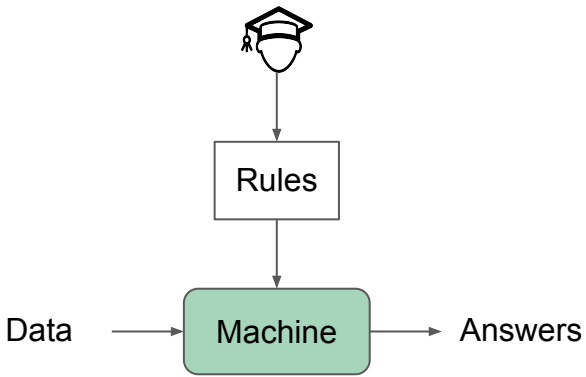
- Cars have wheels
- Simple geometric shape



Why should we learn from data?

Classical Programming Paradigm

- Expert defines rules
- Input data
- Output answer



Example: Write program to detect cars in natural images.

image: Freepik.com

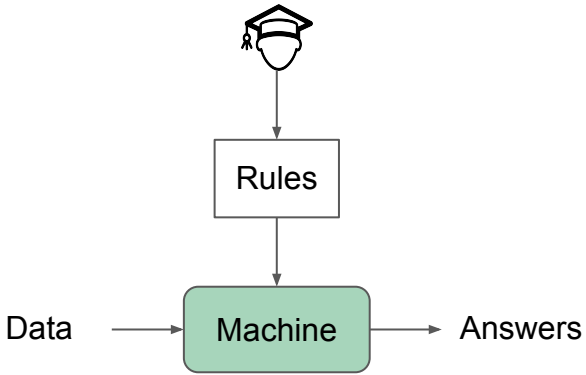
- Cars have wheels
- Simple geometric shape

⇒ How can we describe a feature in terms of pixel values?
⇒ Invariance of features w.r.t. Shadows, reflections, occlusion, etc.

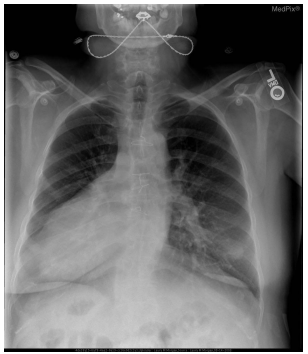
Why should we learn from data?

Classical Programming Paradigm

- Expert defines rules
- Input data
- Output answer



Example: Write program to detect anomalies in MRI imagery.



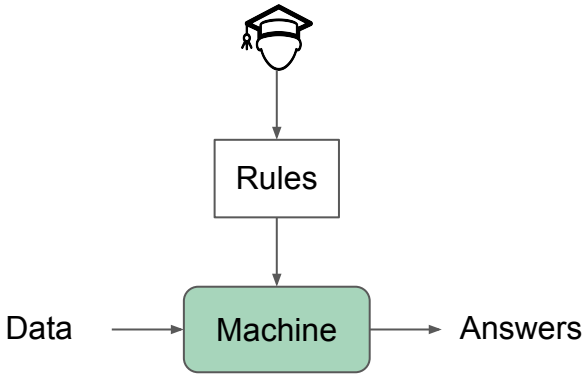
https://ecode.dev/cnn-for-medical-imaging-using-tensorflow-2/

⇒ Can you spot an anomaly?

Why should we learn from data?

Classical Programming Paradigm

- Expert defines rules
- Input data
- Output answer



Example: Chihuahua or Muffin?

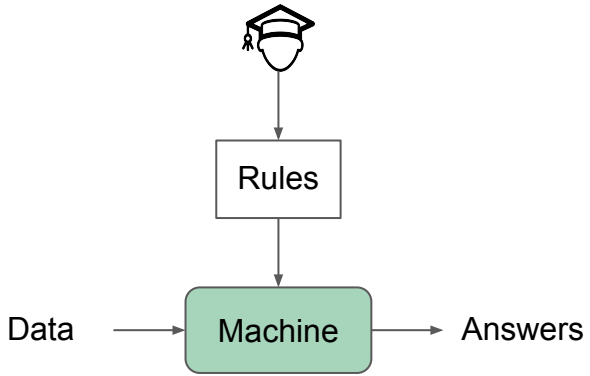
https://www.freecodecamp.org/news/chihuahua-or-muffin-my-search-for-the-best-computer-vision-api-cbda4d6b425d/

⇒ What is a discriminate feature?

Why should we learn from data?

Classical Programming Paradigm

- Expert defines rules
- Input data
- Output answer



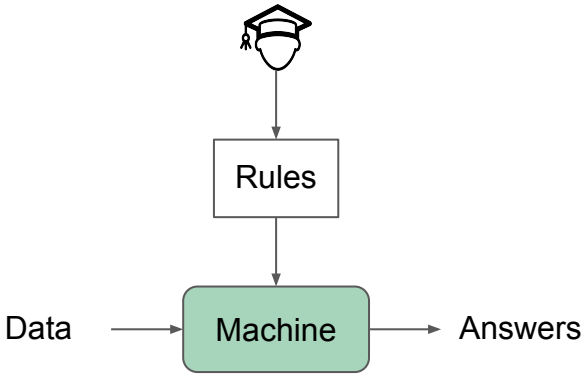
Drawbacks

- difficult to describe features and derive rules (limited to simple problems)
- requires expert knowledge (expensive)
- prone to oversight or bias

Why should we learn from data?

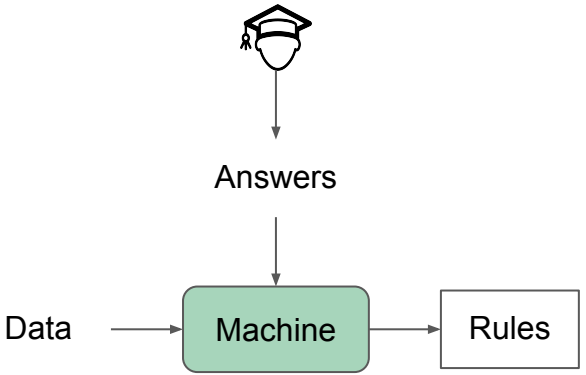
Classical Programming Paradigm

- Expert defines rules
- Input data
- Output answer



Machine Learning Paradigm

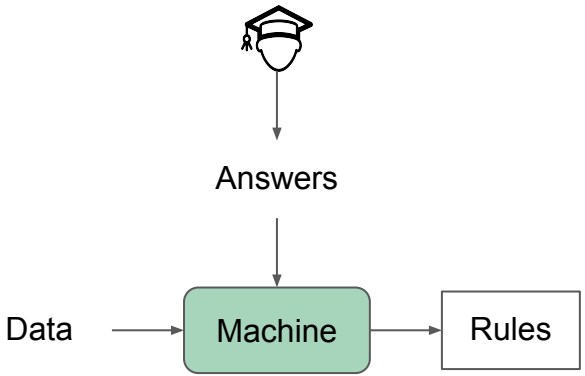
- Expert defines labels for data
- Machine infers rules based on data



Why should we learn from data?

Machine Learning Paradigm

- Expert defines labels for data
- Machine infers rules based on data



Advantages

- Leverage large data availability (Data as *fuel* for Machine Learning)

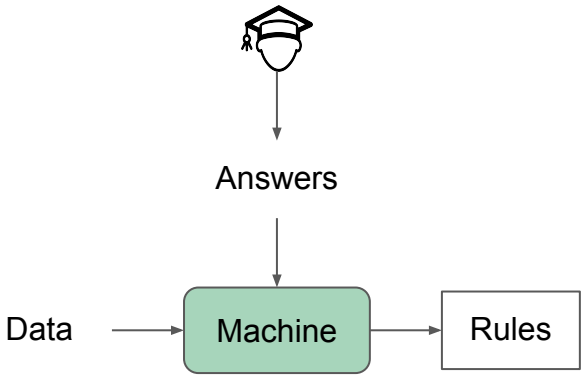
image: Freepik.com

⇒ Millions of images of cars available

Why should we learn from data?

Machine Learning Paradigm

- Expert defines labels for data
- Machine infers rules based on data



Advantages

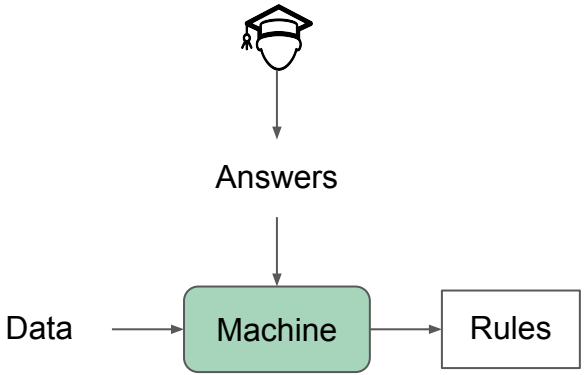
- Leverage large data availability (Data as *fuel* for Machine Learning)
- Can operate on pixel value level and learn features by itself

https://ecode.dev/cnn-for-medical-imaging-using-tensorflow-2/

Why should we learn from data?

Machine Learning Paradigm

- Expert defines labels for data
- Machine infers rules based on data



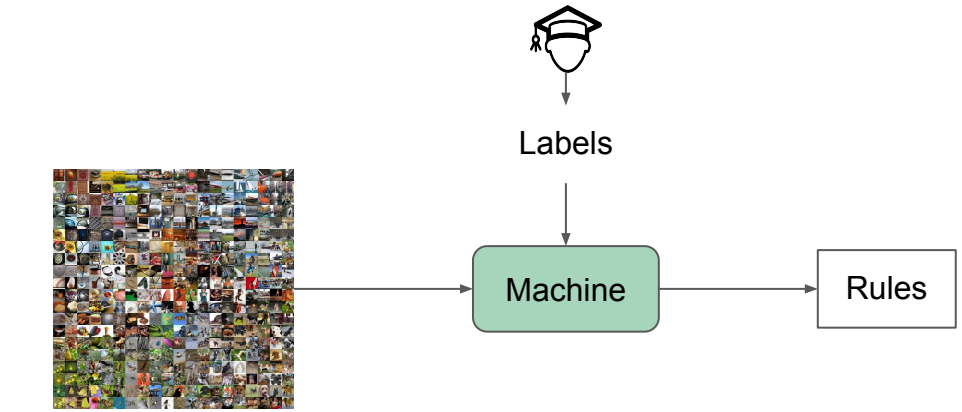
Advantages

- Leverage large data availability (Data as *fuel* for Machine Learning)
- Can operate on pixel value level
- Suitable for complex, dynamic problems

Machine Learning Paradigm

Training

- Present the algorithm many examples
- Let it find statistical structure in examples
- Allow system to come up with rules for automating the task



https://paperswithcode.com/dataset/imagenet

⇒ We will now learn the building blocks

Machine Learning Terms

Input

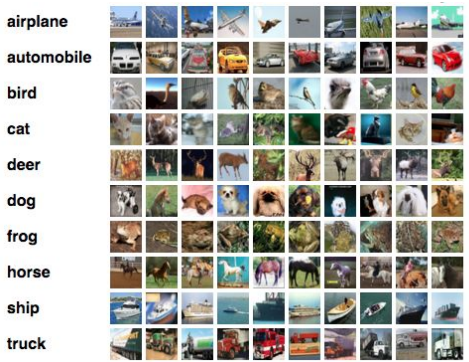
- Data points that we want to process
- e.g. Images, Audio, etc

Output / Labels

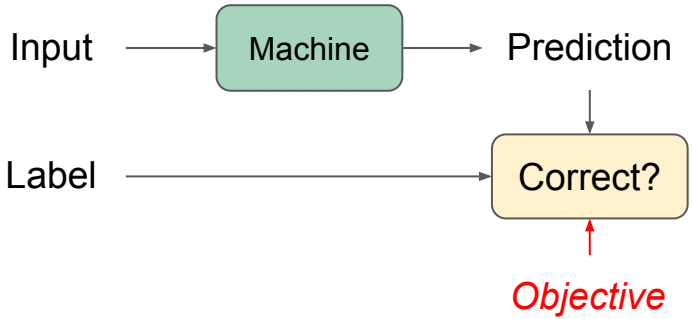
- Corresponding value / label for each datapoint
- e.g. temperature or “car”

Objective

- Feedback signal
- Measures whether algorithm does a good job



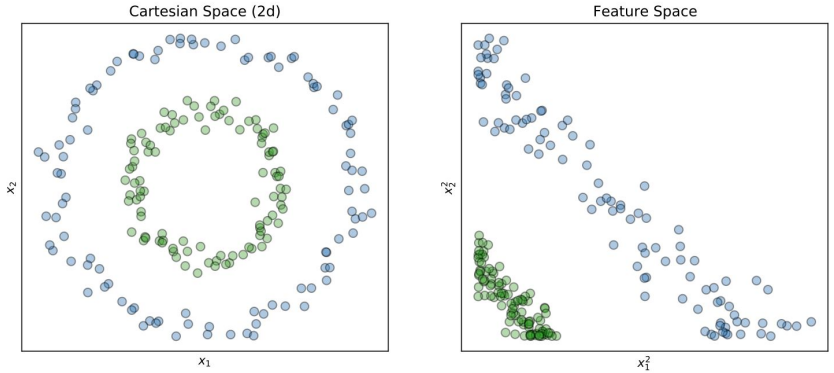
https://www.analyticsvidhya.com/blog/2020/02/learn-image-classification-cnn-convolutional-neural-networks-3-datasets/



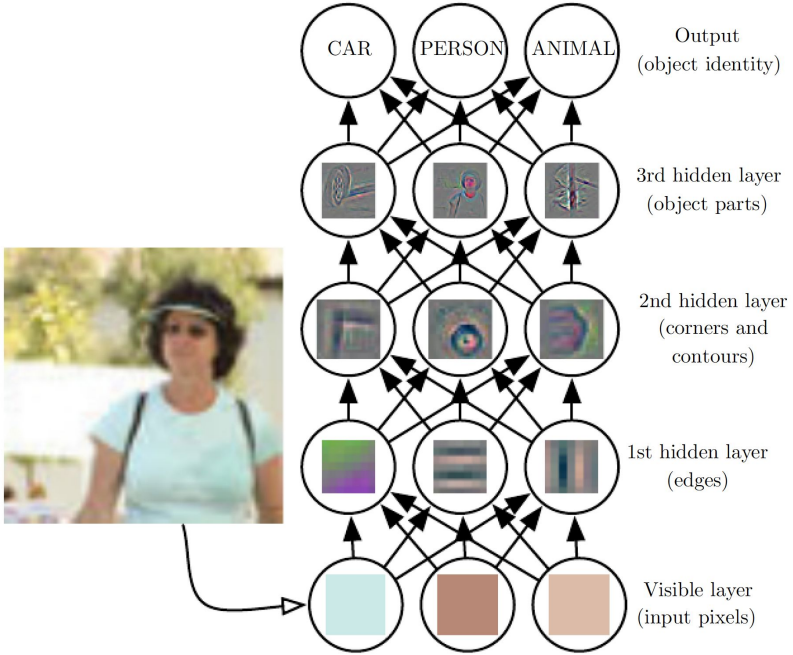
Machine Learning Terms

Representation / Feature Learning

⇒ Learn features or different representation from data



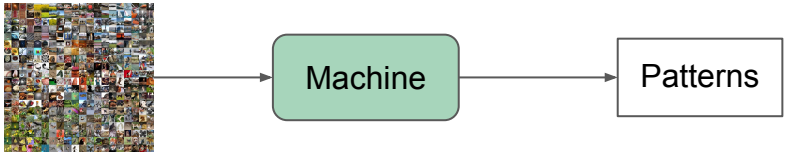
https://sthalles.github.io/a-few-words-on-representation-learning/



Goodfellow, Bengio, & Courville (2016)

Machine Learning Terms

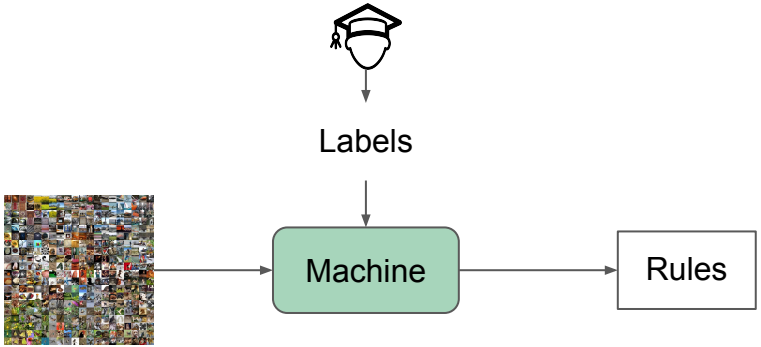
Unsupervised ⇒ Only input and no labels are available



https://paperswithcode.com/dataset/imagenet

Next

Supervised ⇒ Pairs of input and labels are available



https://paperswithcode.com/dataset/imagenet

Later

For the coming weeks…

Bring your own laptop with you!

We will also start with homework next week