AARHUS
UNIVERSITET

Optimization and Data Analytics

Henrik Karstoft
Carl Peter Leslie Schultz
Alexandros Iosifidis
@
Aarhus University, Department of Engineering

Optimization and Data Analytics

Course outline, formal stuff

› Prerequisite

› Lectures

› Homework

› Textbook

› Blackboard

What is this and Goals?

Some definitions

- **Data Analytics** is the discovery, interpretation and communication of meaningful patterns in data
- **Optimization** is the process for *selecting the best element* (with regard to some criterion) from some set of available alternatives

Some course goals

- Increased mathematical skills
- A broad knowledge of the classical and modern mathematical optimization techniques
- Understand the basic Machine Learning models
- Get experience with Data Analysis through project work
- Understand data visualization concepts
- Get introduced to Scientific Writing

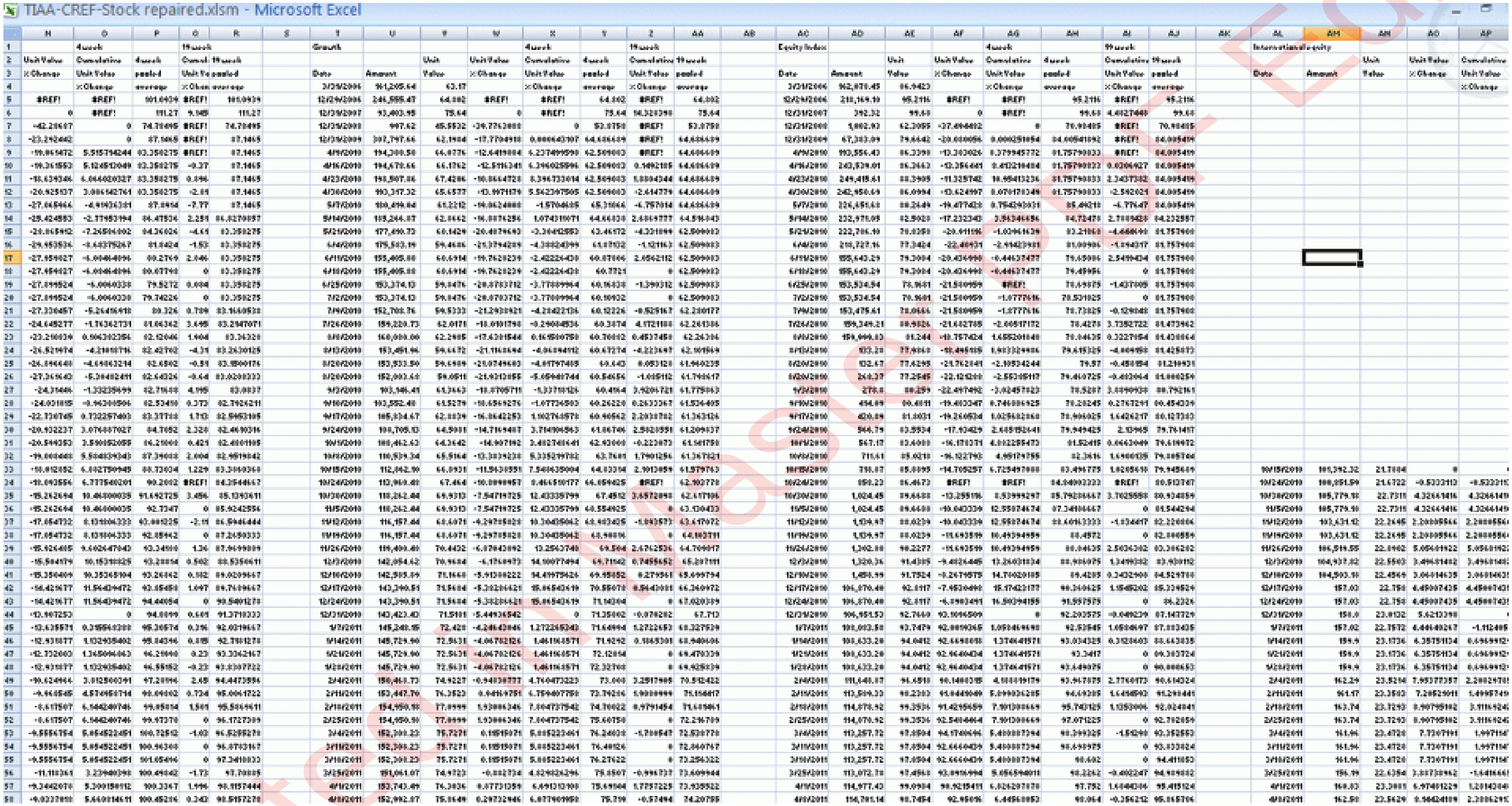
machine learning

optimization

communication,
visualization

today: data overload

data **useless** without interpretation



anything interesting happening here?

"interpretation" = analysis

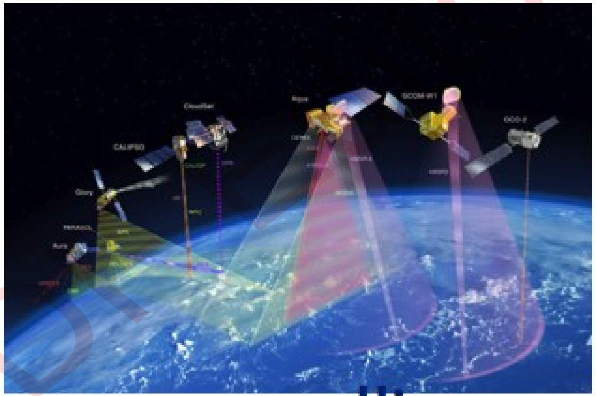
= finding patterns

what are examples of **data**?

what are examples of **patterns**?

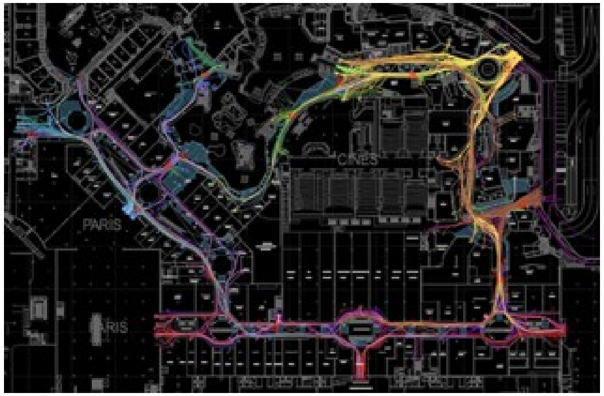
video


audio


satellite


images


stocks


movement

computers can help!
with lots of examples computers can build “pattern finder”

machine learning

optimization

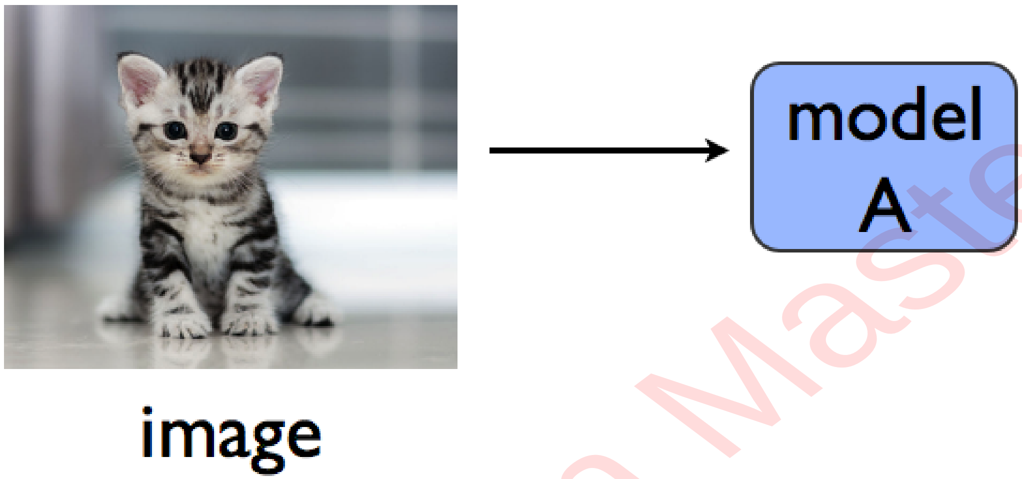
communication, visualization

models can "find patterns"

machine learning: building models from lots of examples

model
A

models can “find patterns”



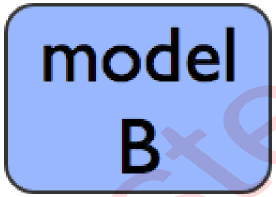
models can "find patterns"

image

model A

donkey!

models can “find patterns”

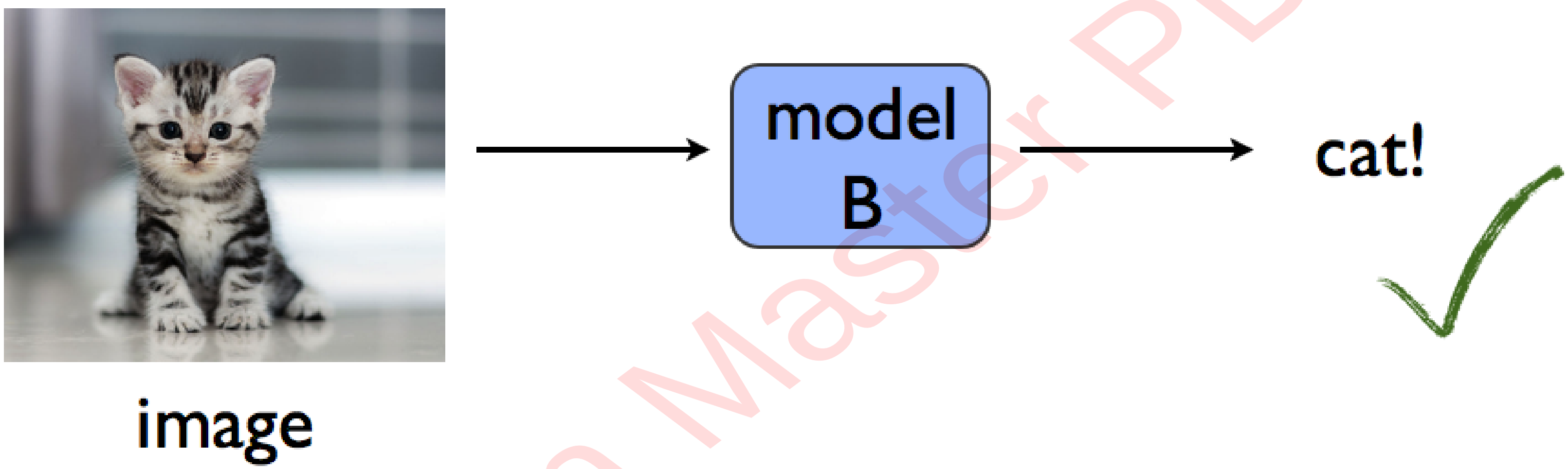


models can “find patterns”

image

model
B

models can “find patterns”



machine learning

optimization

communication,
visualization

imagine set of **all possible** models

model
A

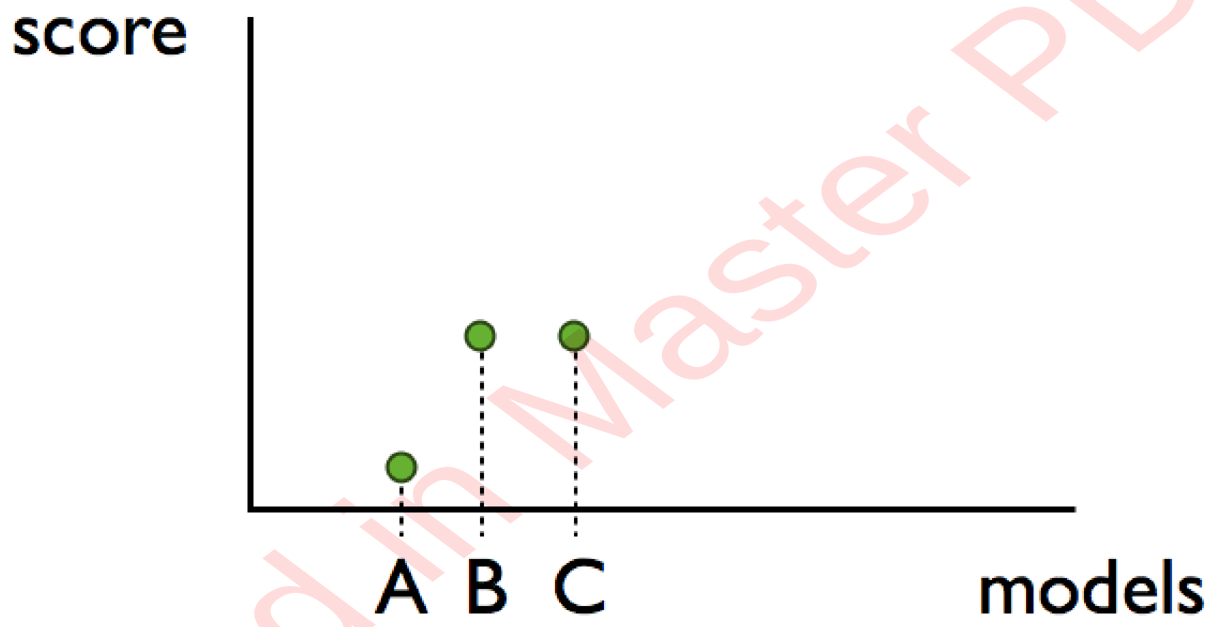
model
B

model
C

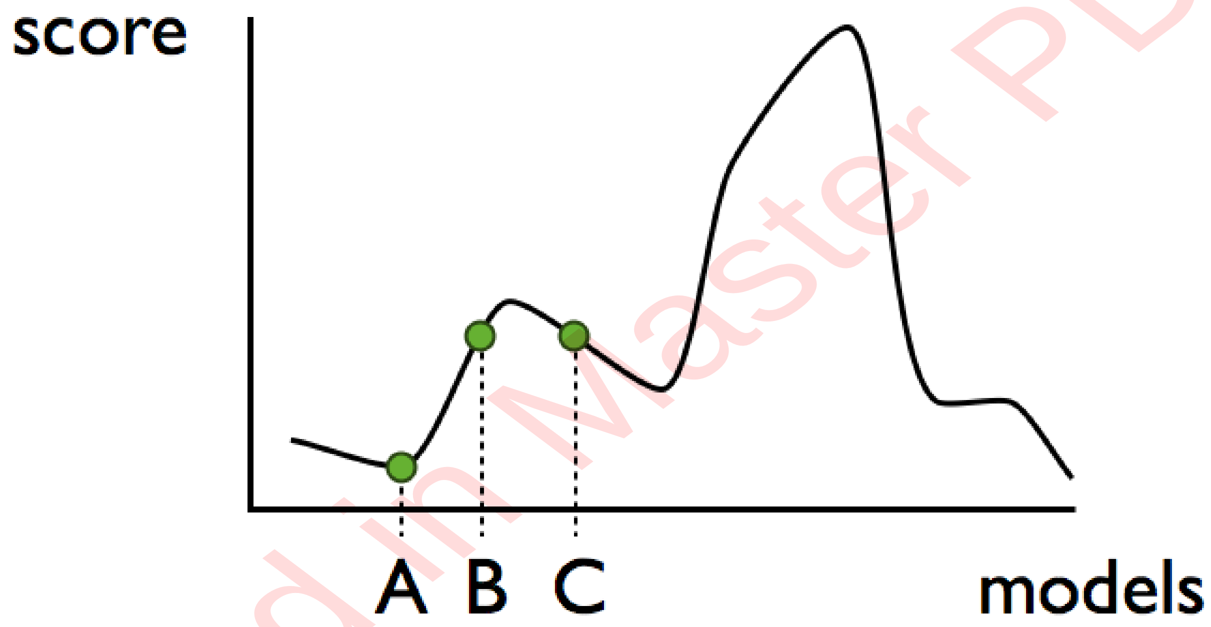
...

let's give each model a **score**

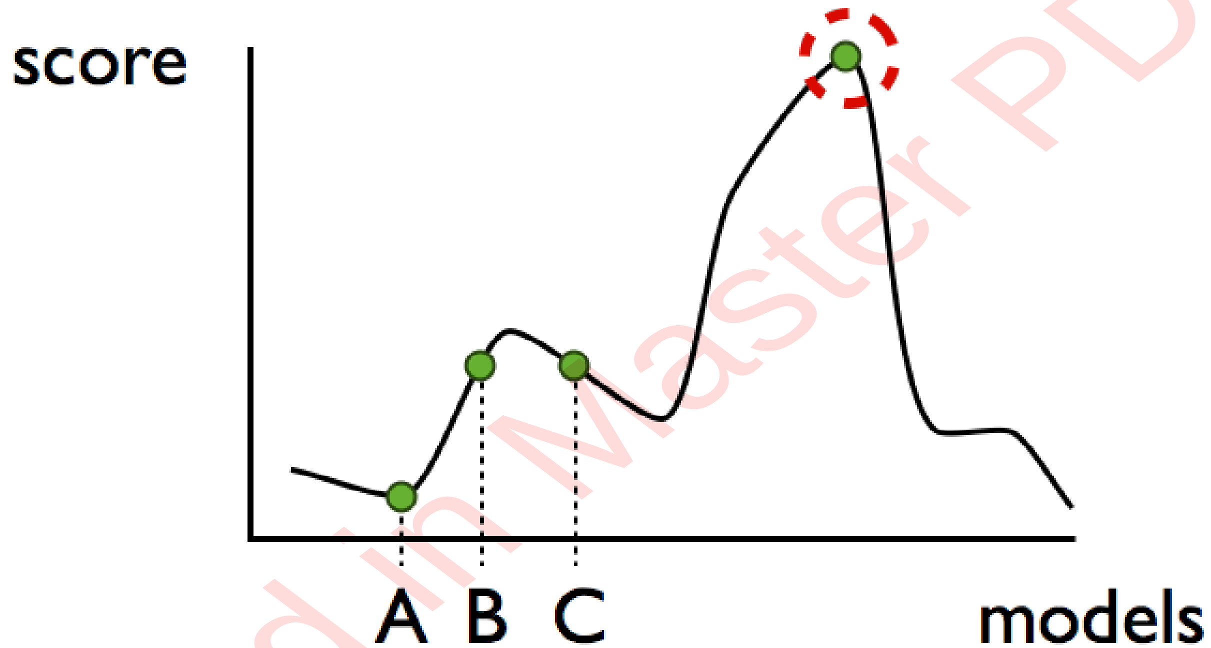
we want **best model** we can find



we want **best model** we can find

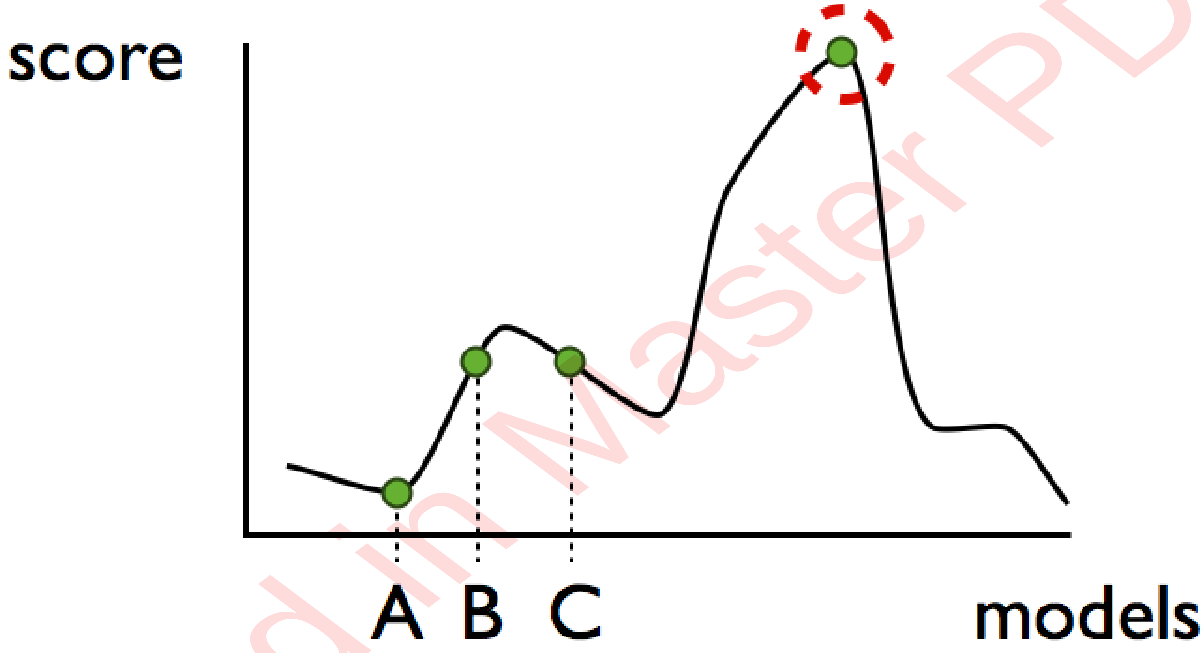


we want **best model** we can find



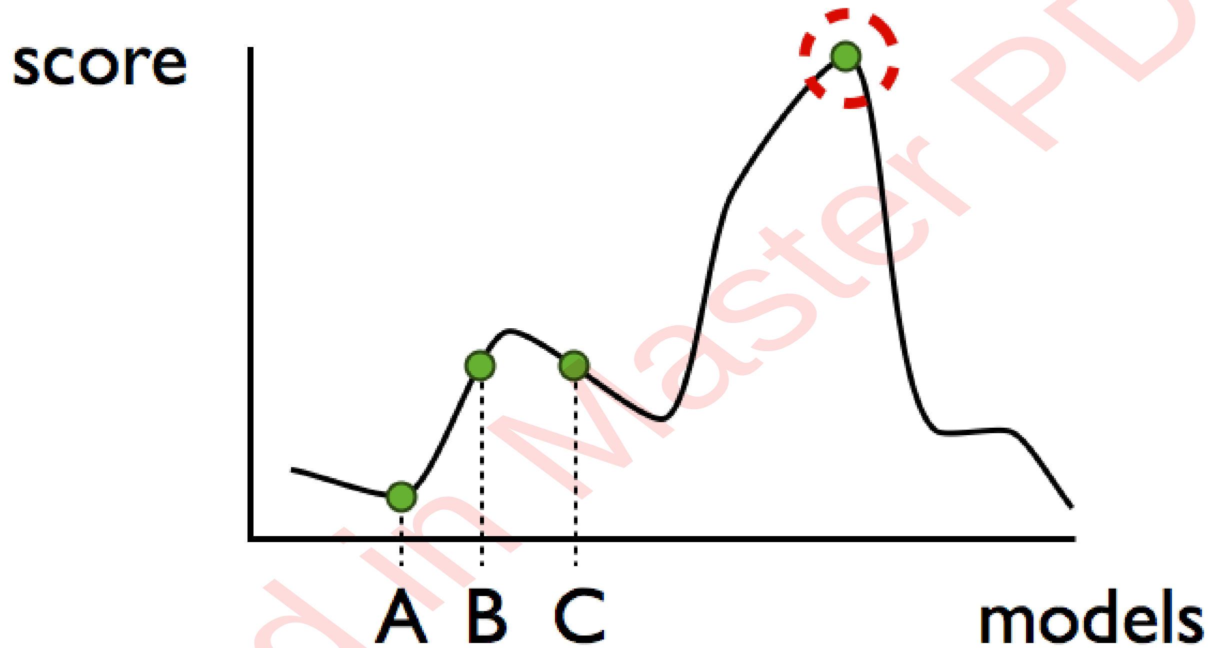
optimization: find model that maximises score

how did I explain this concept to you?



how did I explain this concept to you?

I **visualized** set of models as a **chart**

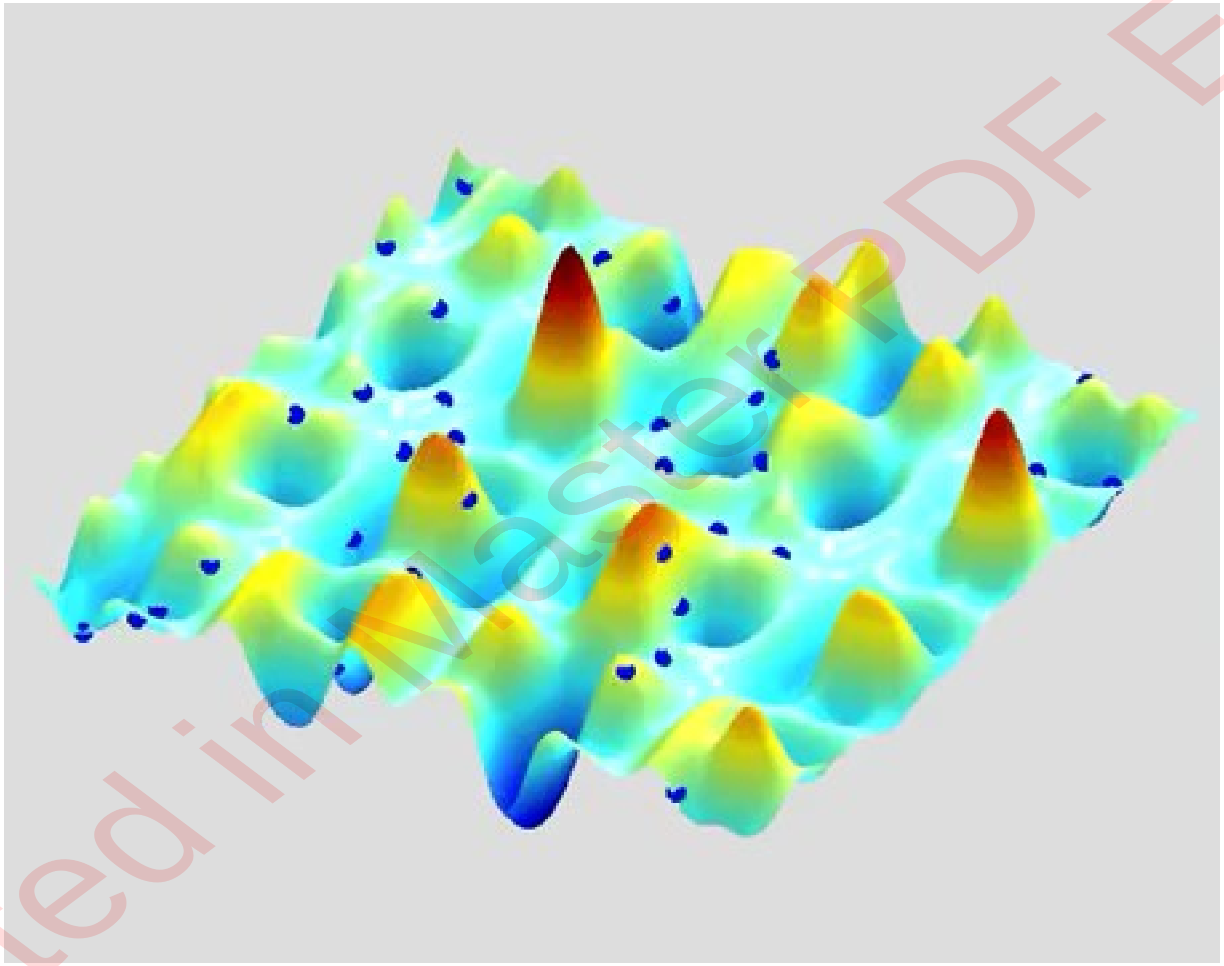


machine learning

optimization

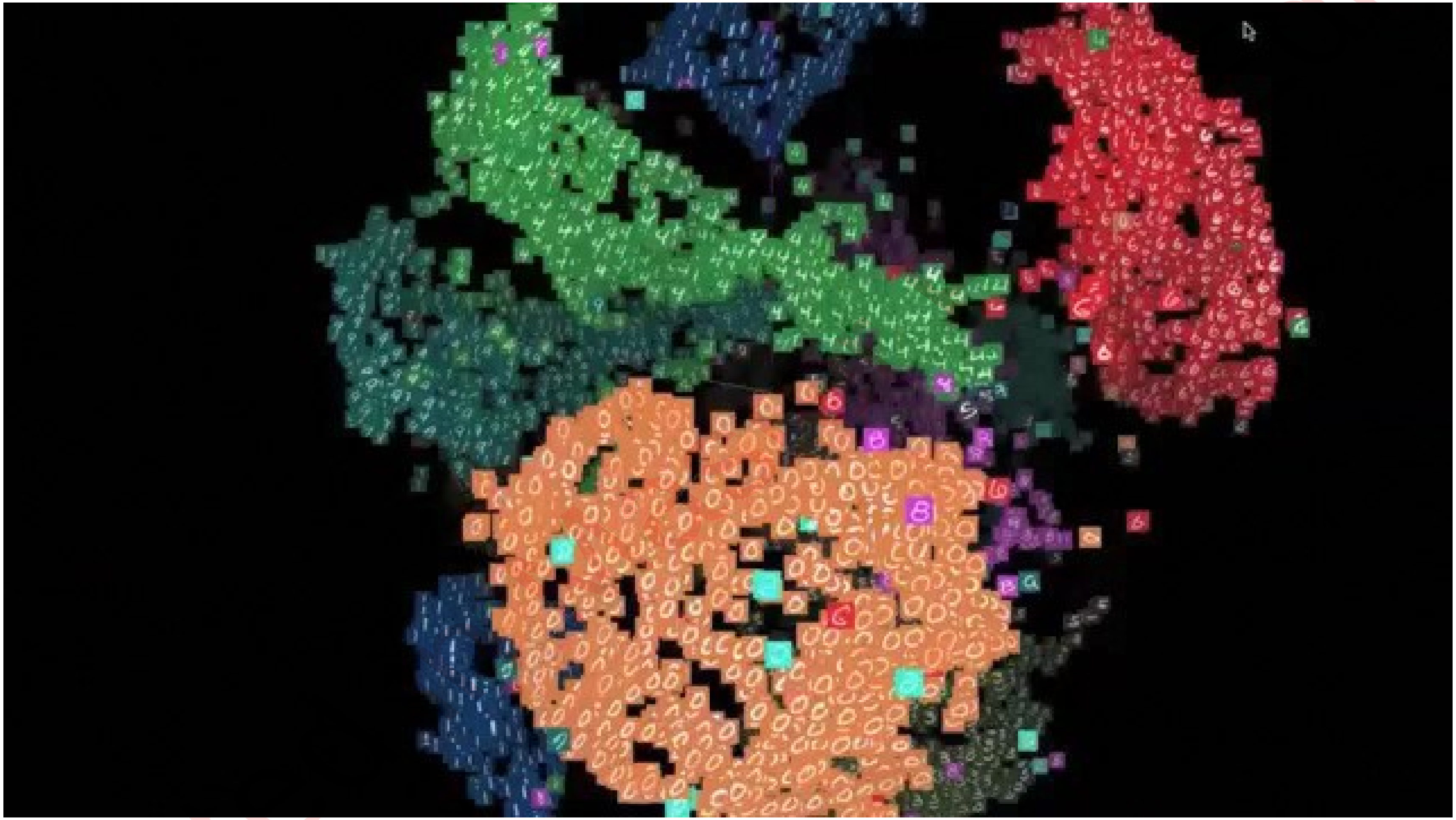
communication,
visualization

tools for thinking, understanding, communicating



learning process https://www.youtube.com/watch?v=VAASmSSsFaY

tools for thinking, understanding, communicating



learnt classification (from Google)

https://www.youtube.com/watch?v=wvsE8jm1GzE

part 3
machine learning

part 1
optimization

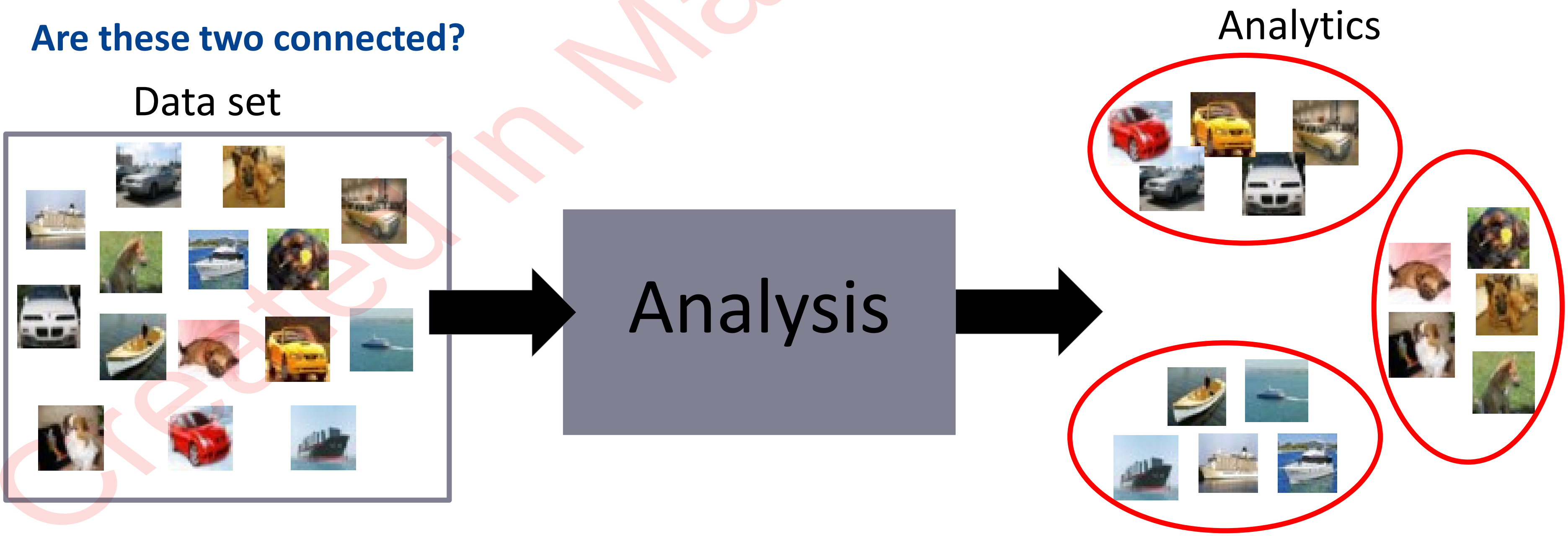
part 2
communication,
visualization

What is this?

Some definitions

- **Data Analytics** is the discovery, interpretation and communication of meaningful patterns in data
- **Optimization** is the process for *selecting the best element* (with regard to some criterion) from some set of available alternatives

Are these two connected?



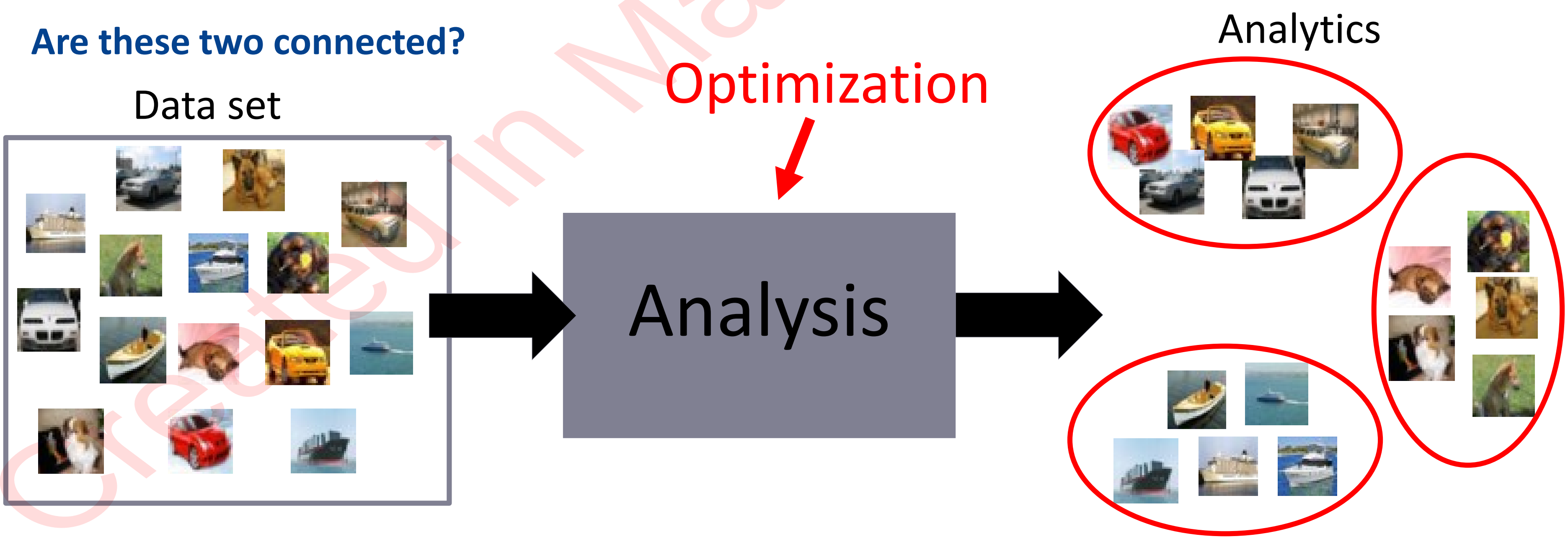
AARHUS
UNIVERSITET

What is this?

Some definitions

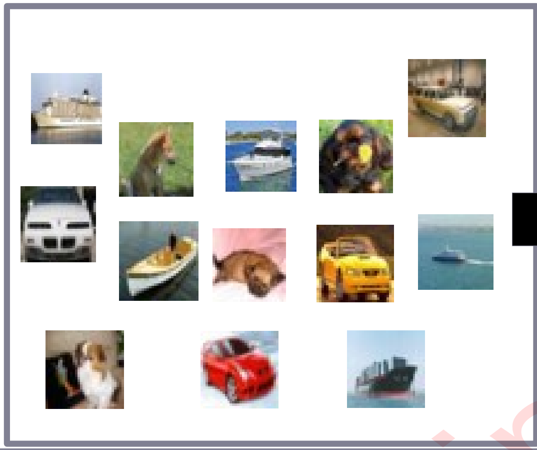
- **Data Analytics** is the discovery, interpretation and communication of meaningful patterns in data
- **Optimization** is the process for *selecting the best element* (with regard to some criterion) from some set of available alternatives

Are these two connected?



Optimization for parameter estimation

Training phase



Optimization

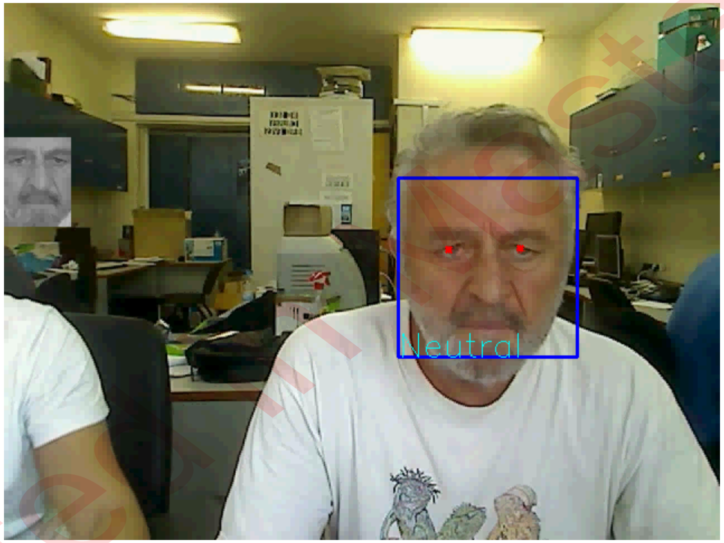
Parameters of the Model → Trained Model

Test phase/Evaluation/Online process

→ Trained Model → Car

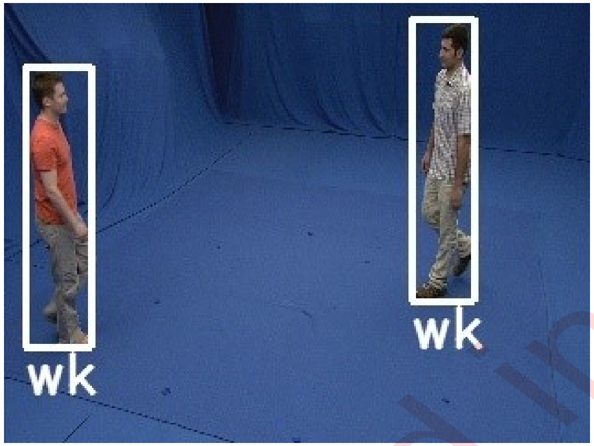
Examples

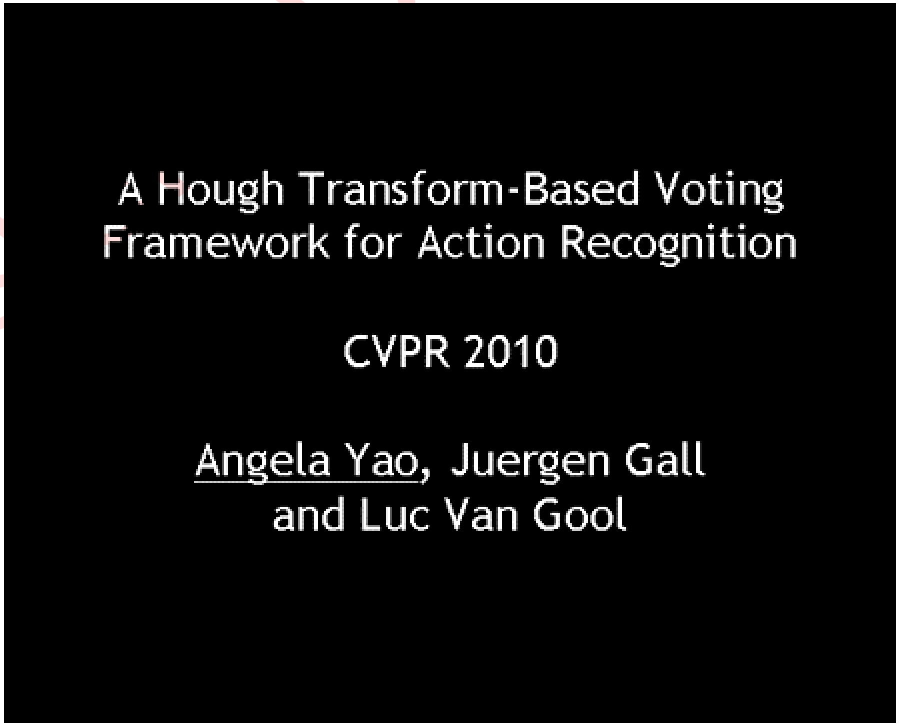
Face Analysis



Examples

Action recognition/analysis

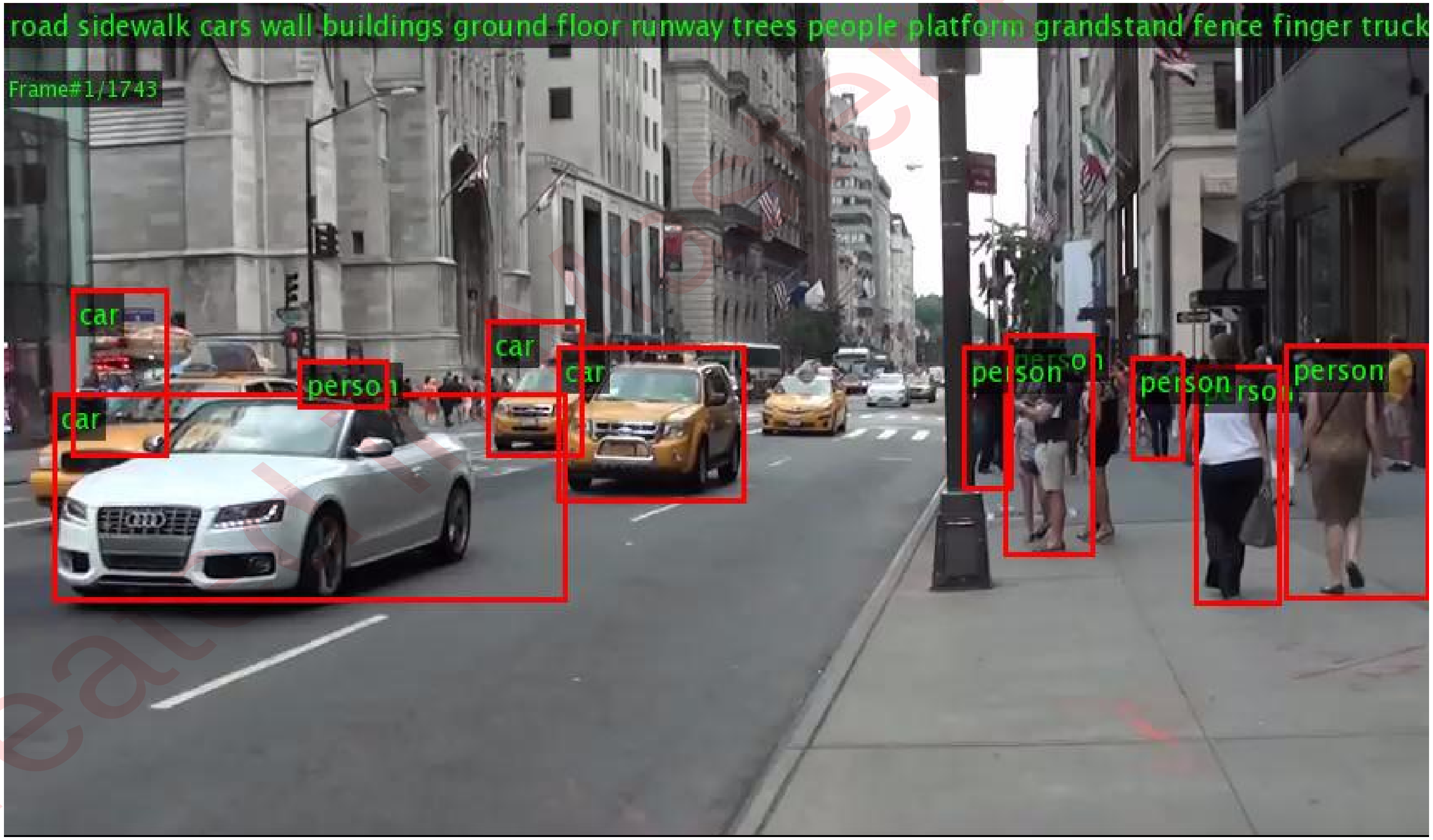




Source: https://www.youtube.com/watch?v=sgcVOOZI8bw

Examples

Object localization/recognition, scene understanding



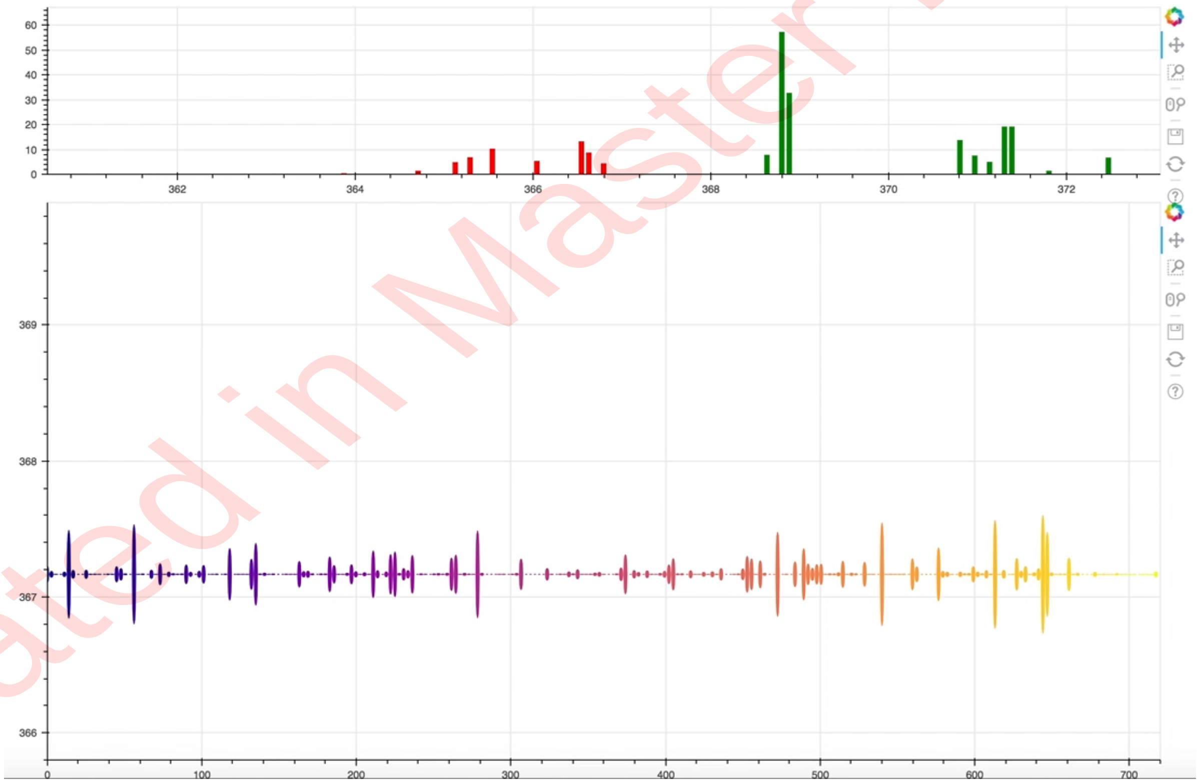
Examples

Image/video captioning

Source: https://www.youtube.com/watch?v=FmSsek5luHk

Examples

Time-series forecasting



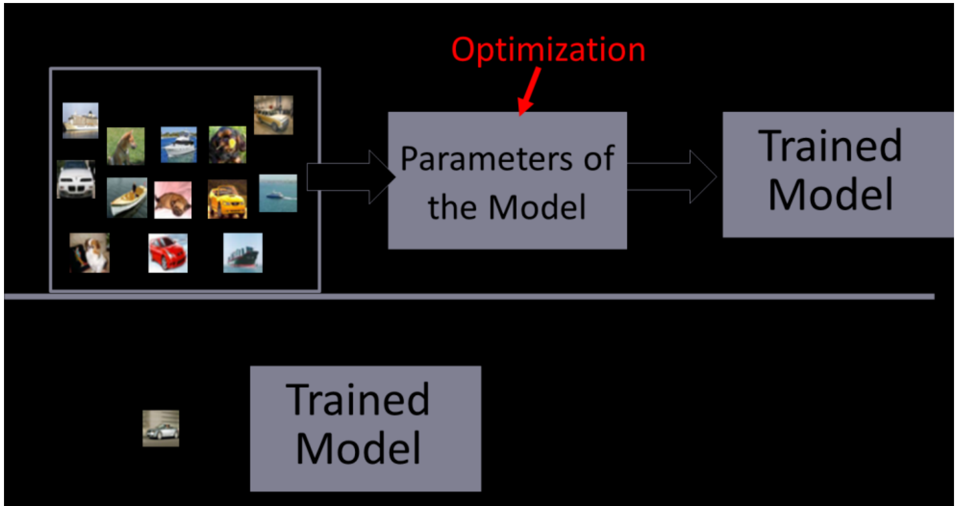
Examples

Everyday examples

- Natural language processing
- Affective computing
- Medical diagnosis
- Robotics
- Speech and handwritten recognition
- Translation
- …

Why should I learn about Optimization?

Machine learning models are in fact formed by a set of parameters which need to be 'optimized', i.e. to find good values for these parameters in order to achieve our goal.



Why should I learn about Optimization?

Machine learning models are in fact formed by a set of parameters which need to be 'optimized', i.e. to find good values for these parameters in order to achieve our goal.

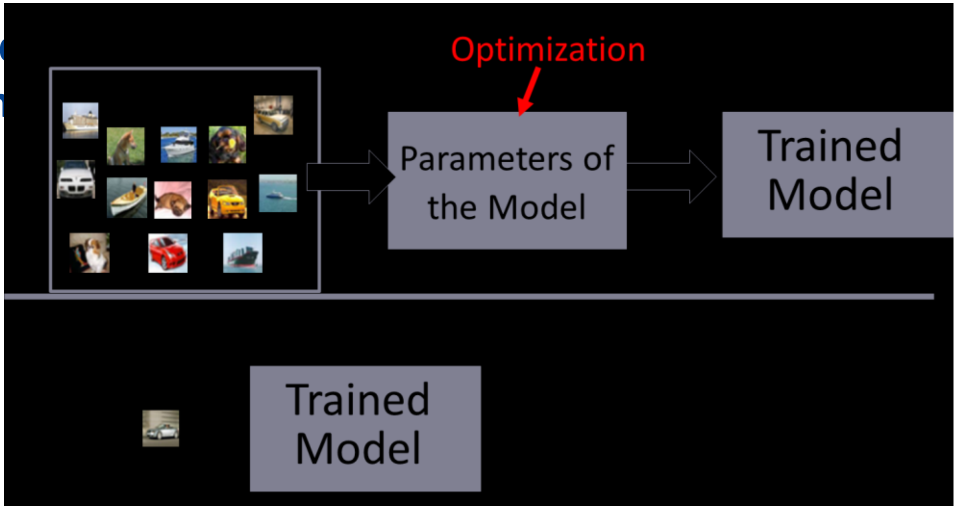
These parameters are usually stored in vectors and matrices. For example a linear model has the form:

$$\mathbf{y}_i = \mathbf{W}^T \mathbf{x}_i$$

W can be random values if x is high dimensions and y is low dimensions. For classifier, W cannot be random.

where the parameters of the linear model are stored in the matrix W.

In this case, the goal of optimization is to find an 'optimal W', which means a W that, given the input x_i can generate a good (desired) y_i.



Optimization topics

In the course, we will cover the following topics:

- Linear programming
- Unconstrained optimization
- Nonlinear constrained optimization
- Solving linear equations
- Unconstrained optimization: 1D search methods
- Unconstrained optimization: Gradient methods
- Global search methods

Mathematical optimization

A large number of optimization problems can be written in the following form:

$$\text{minimize } f_0(x)$$
$$\text{subject to } f_i(x) \leq b_i, \text{ where } i = 1, \dots, m$$

Contraints are used to define a solution that is practical.

- $x = (x_1, \dots, x_n)$, Optimization variables, or decision variables
- f_0: $R^n \to R$, objective function
- $f_i(x)$ $R^n \to R$, where $i = 1, \dots, m$: constraint functions
- x^* **optimal solution**, smallest value of f_0 among $x = (x_1, \dots, x_n)$, satisfying the constraint

Mathematical optimization

A large number of optimization problems can be written in the following form:

$$\text{minimize } f_0(x)$$
$$\text{subject to } f_i(x) \le b_i, \text{ where } i = 1, \dots, m$$

- $x = (x_1, \dots, x_n)$, Optimization variables, or decision variables
- f_0: $R^n \to R$, objective function
- $f_i(x)$ $R^n \to R$, where $i = 1, \dots, m$: constraint functions
- x^* **optimal solution**, smallest value of f_0 among $x = (x_1, \dots, x_n)$, satisfying the constraint

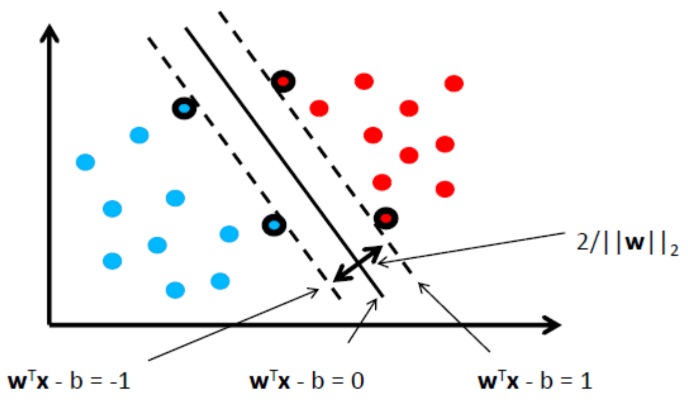
Mathematical optimization

A large number of optimization problems can be written in the following form:

$$\text{minimize } f_0(x)$$
$$\text{subject to } f_i(x) \leq b_i, \text{ where } i = 1, \ldots, m$$

- $x = (x_1, \ldots, x_n)$, Optimization variables, or decision variables
- f_0: $R^n \rightarrow R$, objective function
- $f_i(x)$ $R^n \rightarrow R$, where $i = 1, \ldots, m$: constraint functions
- x^* **optimal solution**, smallest value of f_0 among $x = (x_1, \ldots, x_n)$, satisfying the constraint

Example of ML model: SVM



Linear Programming

Linear programming problems have the following form

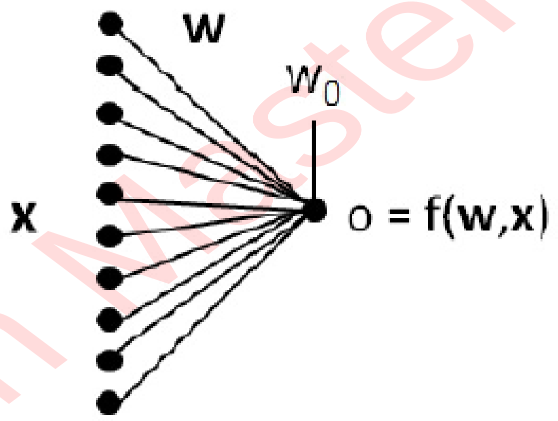
if we had c^T c then we would not be able to use linear programming

$$\text{minimize } c^T x$$
$$\text{subject to } {a_i}^T x \leq b_i \;\; i = 1, \dots, , m$$

- *no* analytic solution
- reliable and efficient algorithms and software
- computational time proportional to $n^2 m$, where $A \in R^{m \times n}$, in some case less than this
- mature technology

Linear Programming

Example of ML model: Perceptron



Unconstrained optimization

Unconstrained optimization is used in cases where there are no constraints on the model's parameters

For example a linear unconstrained problem has the form

$$\text{minimize } c^T x$$

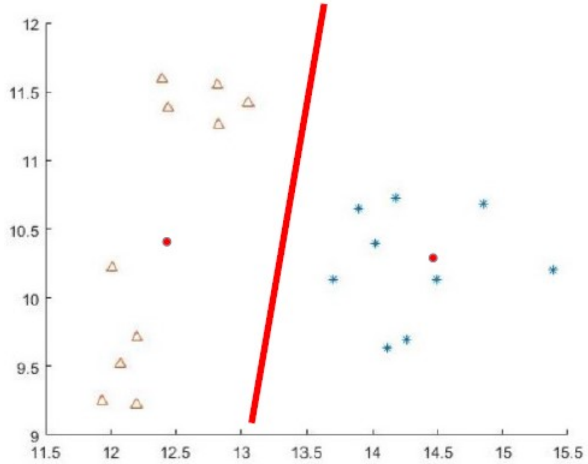
Unconstrained optimization

Unconstrained optimization is used in cases where there are no constraints on the model's parameters

For example a linear unconstrained problem has the form

$$\text{minimize } c^T x$$

Example of ML model: Linear classifiers



Solving linear equations

A system of linear equations has the form:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$$

Example of ML model: Linear/nonlinear Regression

$$\mathcal{J}_s = \sum_{i=1}^{N} (\mathbf{w}^T \mathbf{x}_i - b_i)^2 = \|\mathbf{X}^T \mathbf{w} - \mathbf{b}\|_2^2$$

1D search methods

Very often, optimization cannot be obtained using a closed form solution. Then we need to follow an iterative process which at every step gives a set of parameters which are better than before

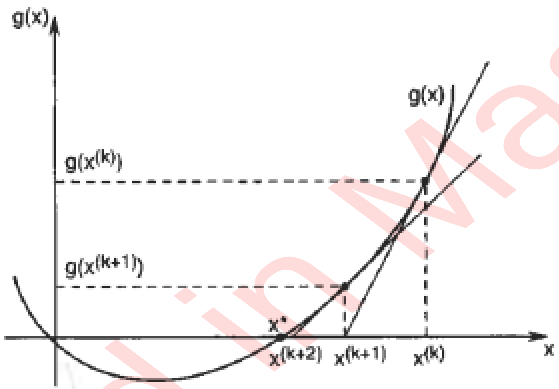
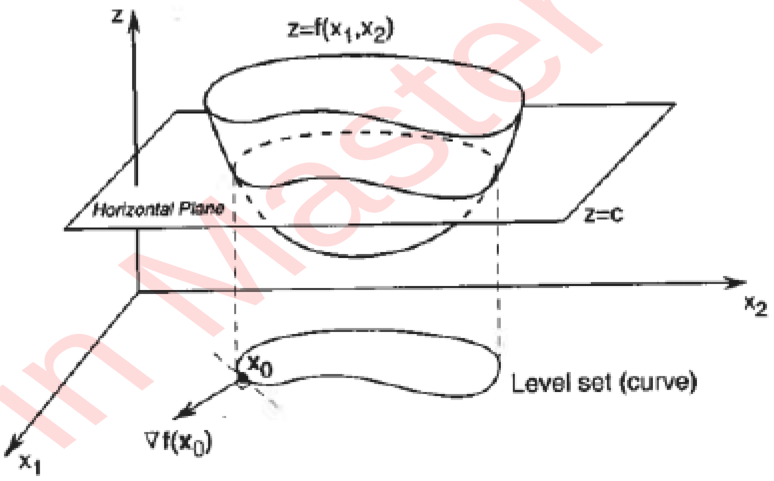


Figure 7.8 Newton's method of tangents.

$$x^{(k+1)} = x^{(k)} - \frac{g(x^{(k)})}{g'(x^{(k)})}$$

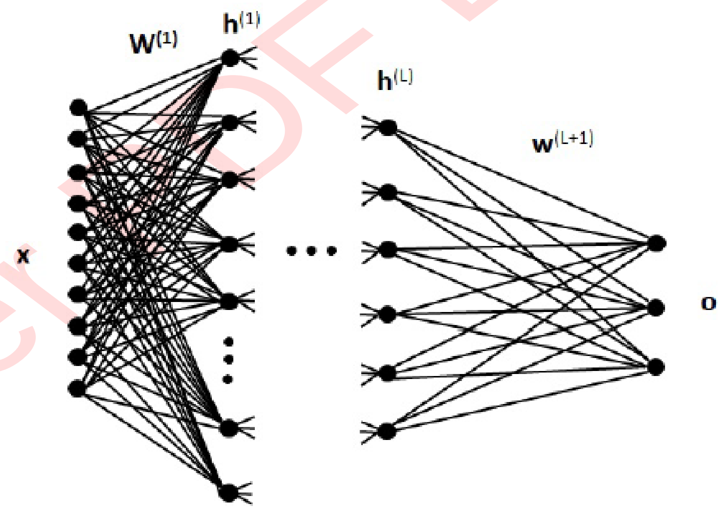
Gradient methods

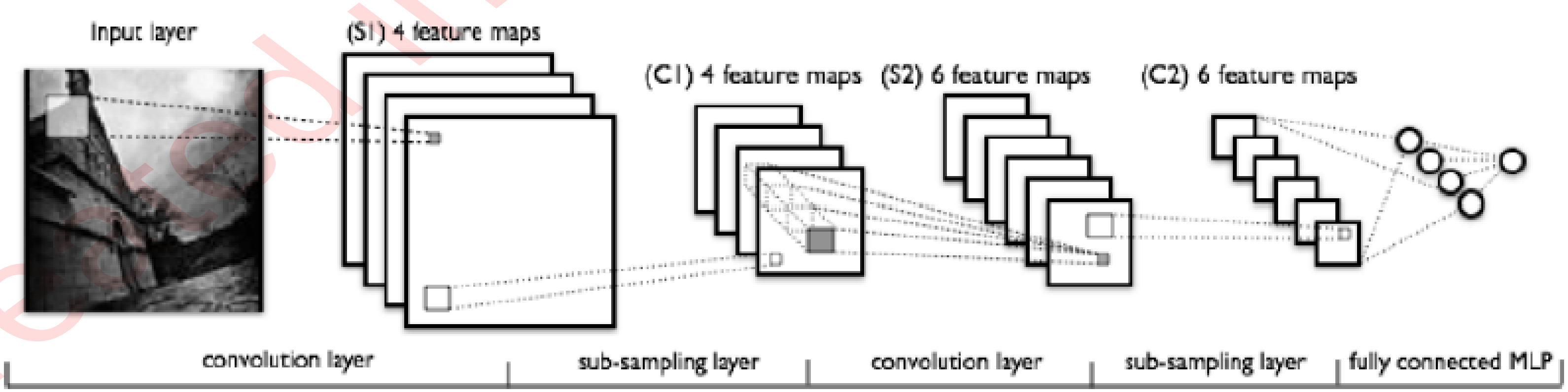
When the number of parameters is higher than one, the iterative optimization process uses the gradient of the function w.r.t. the parameters for updating:



Gradient methods

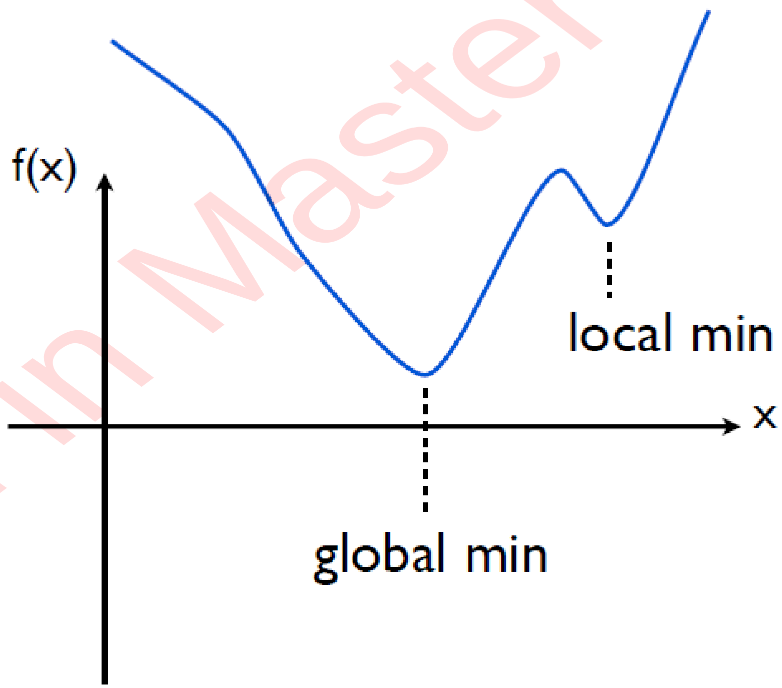
Example of ML model: Neural networks





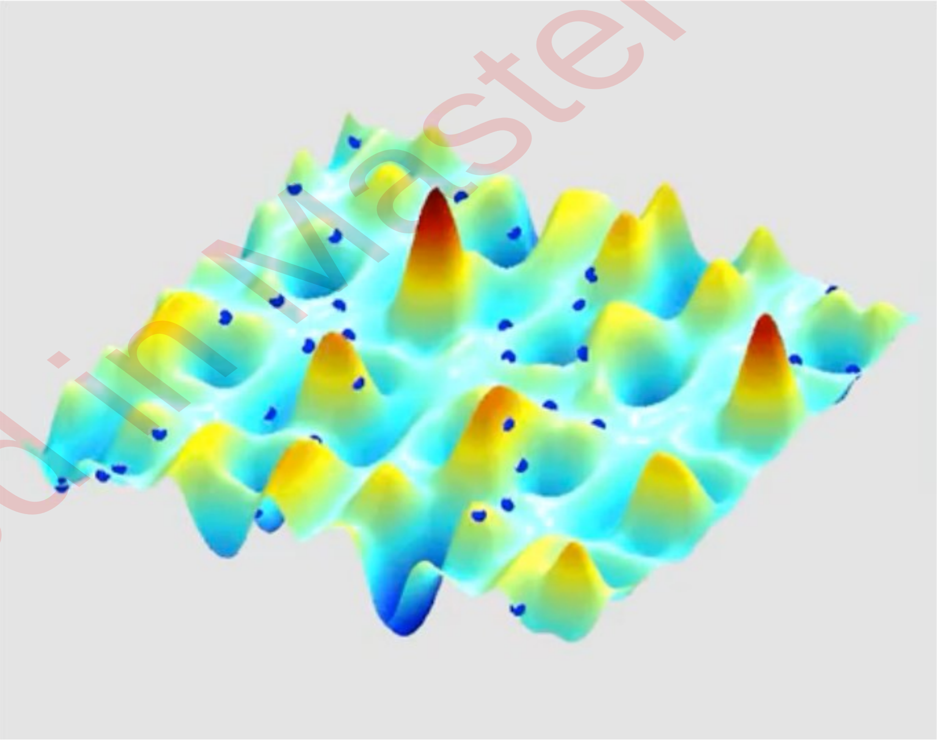
Global search methods

One of the problems of iterative optimization methods is that their performance depends on the starting point:



Global search methods

One of the problems of iterative optimization methods is that their performance depends on the starting point:



https://www.youtube.com/watch?v=VAASmSSsFaY