AARHUS
UNIVERSITET

Optimization and Data Analytics

Alexandros Iosifidis
@
Aarhus University, Department of Engineering

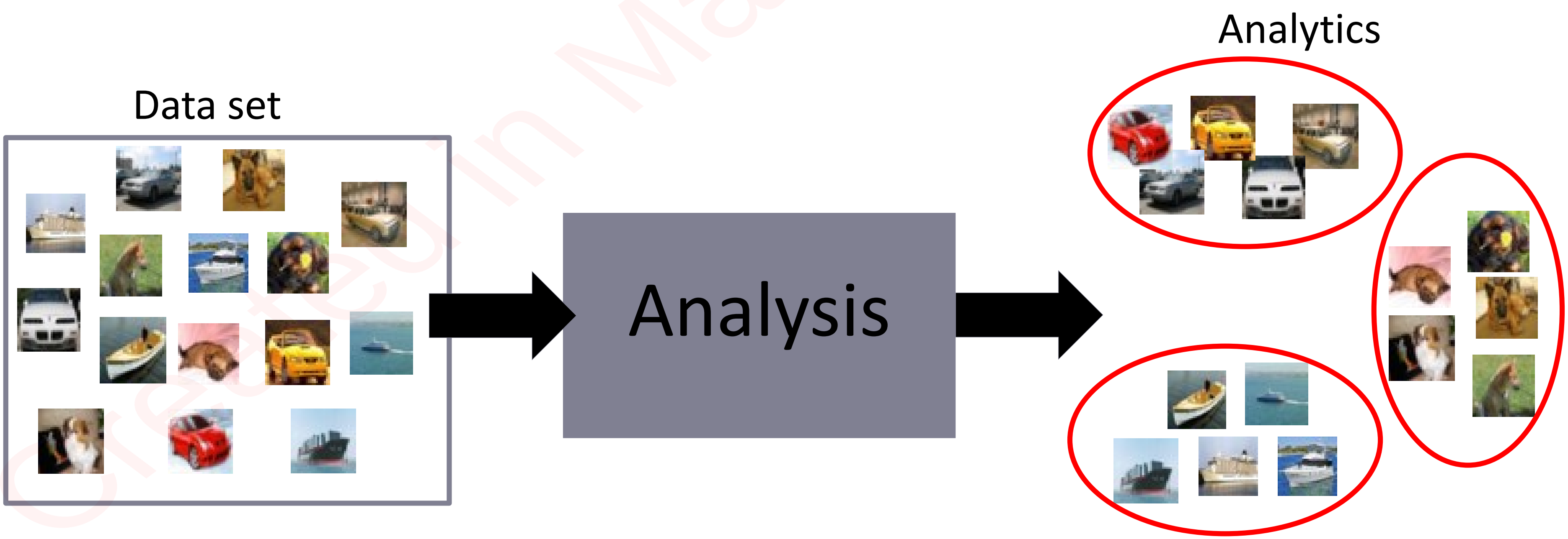
What is Data Analytics?

Meaningful depends on the problem at hand.

Discovery: because we don't know that there are patterns
Communication: there are different ways of communicating the results/discovery.
This depends on the problem that we have what we want to say.

A definition

› **Data Analytics** is the discovery, interpretation and communication of meaningful patterns in data

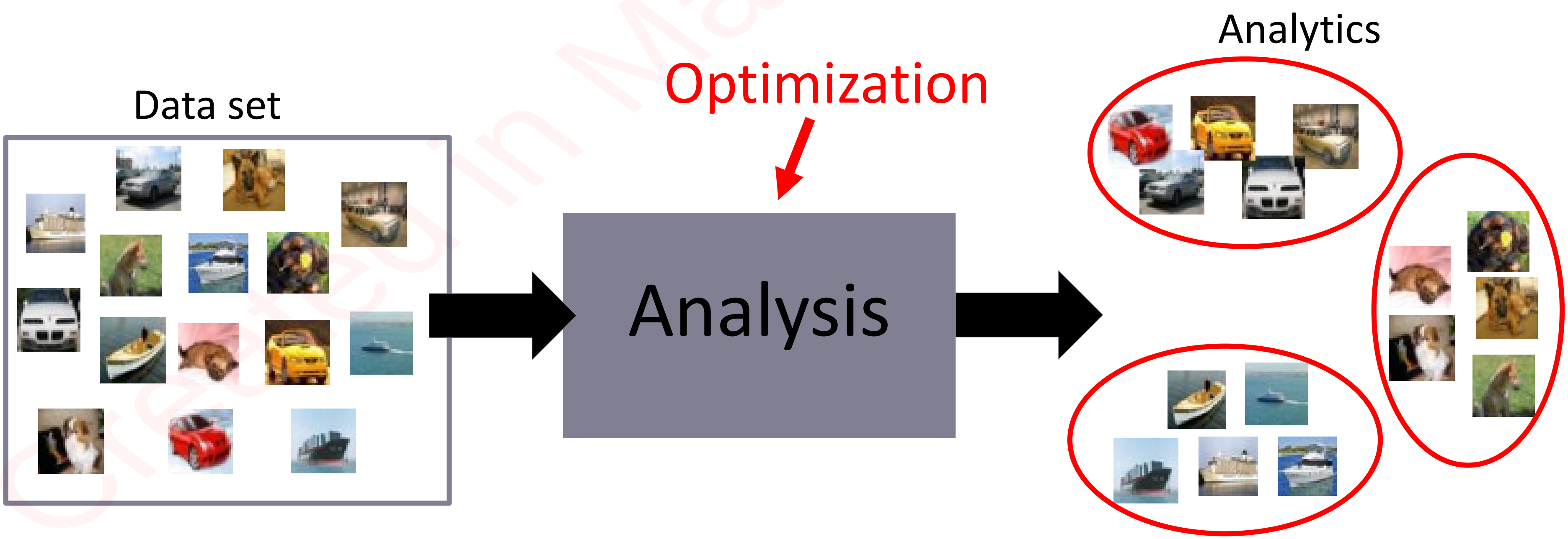


Connection to Optimization?

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



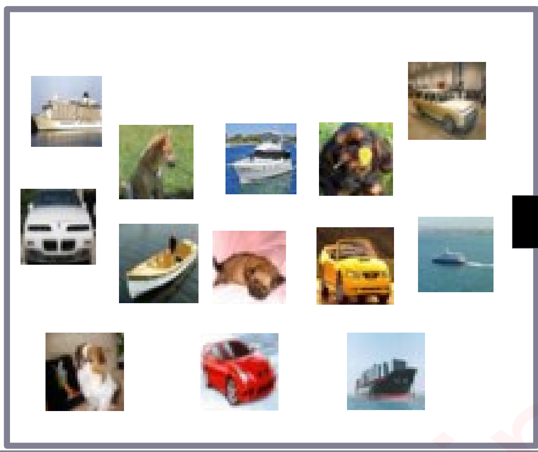
We assume that the training set and test sets are from the same distribution, so when we have trained our model, then on the test data the model should perform well

Optimization for parameter estimation

Empirical error = error in training data

Training phase

Optimization



Parameters of the Model

Trained Model

Test phase/Evaluation/Online process

Expected Error: the error in test data

Trained Model

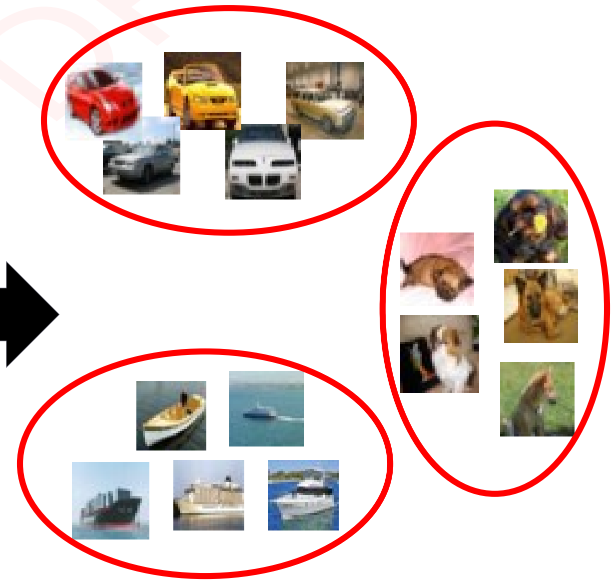
Car

Bottom-up view of the analysis

Data set

Optimization

Analysis



Data pre-processing → Data representation → Repr. pre-processing → Model selection/ training

Data pre-processing

For example:

- Image segmentation
- Temp. video segm.

Data representation

For example:

- Vector per sample
- Vectors per sample

Repr. pre-processing

For example:

- Centering
- Standardization

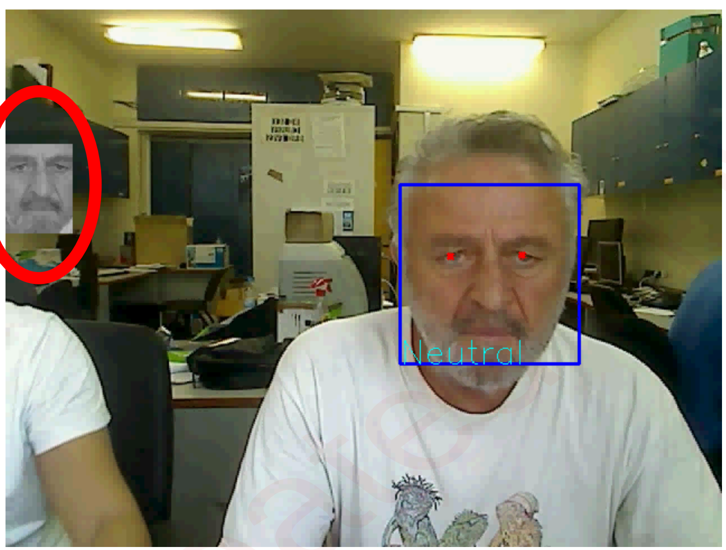
Model selection/ training

For example:

- Classification
- Clustering

Bottom-up view of the analysis

Data pre-processing → Data representation → Repr. pre-processing → Model selection/ training



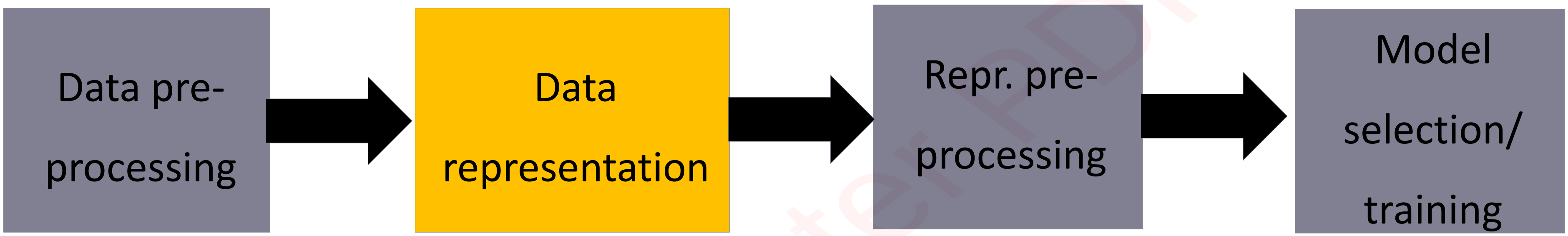
Data pre-processing is a process which is guided by the domain of the problem to be solved.

For example in facial image analysis:

- Face detection and segmentation
- (Possibly) image frontalization
- Illumination normalization
- Facial image resizing

What about human action recognition from videos?

Bottom-up view of the analysis



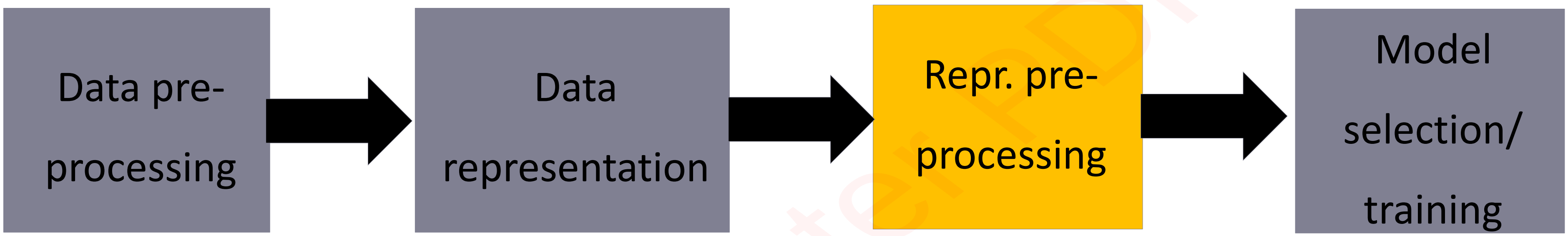
Data representations can also be guided by the domain of the problem to be solved.

For example we can represent a facial image based on:

- Image intensities
- Edges → texture
- Locations and sizes of each facial feature

What about a human action?

Bottom-up view of the analysis



After obtaining data representations, we (usually) obtain a D-dimensional vector **x** for each sample.

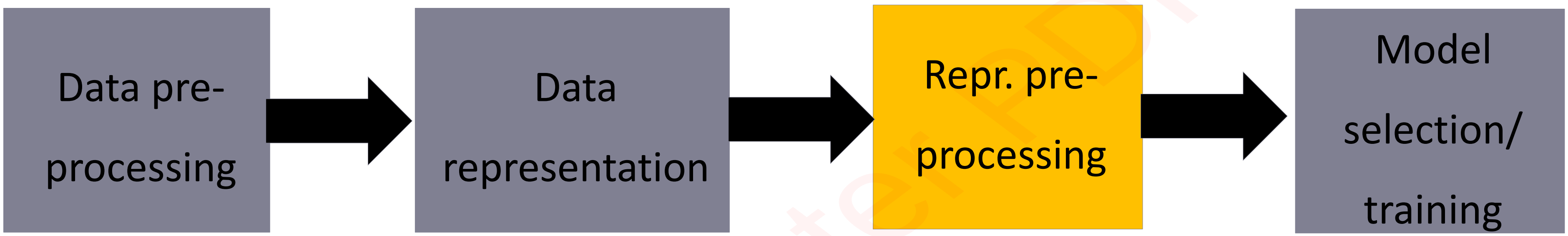
Bottom-up view of the analysis

Data pre-processing → Data representation → Repr. pre-processing → Model selection/ training

After obtaining data representations, we (usually) obtain a D-dimensional vector **x** for each sample.

Let us assume that we have a set of N samples. We use a subscript i=1,…,N to denote a specific sample. That is, we write $\mathbf{x}_i$ to denote the i-th sample in the data set. Some times it is convenient to write $\mathbf{X} = [\mathbf{x}_1,\ldots,\mathbf{x}_N]$.

Bottom-up view of the analysis



After obtaining data representations, we (usually) obtain a D-dimensional vector $\mathbf{x}$ for each sample.

Let us assume that we have a set of N samples. We use a subscript i=1,…,N to denote a specific sample. That is, we write $\mathbf{x}_i$ to denote the i-th sample in the data set. Some times it is convenient to write $\mathbf{X} = [\mathbf{x}_1,\ldots,\mathbf{x}_N]$.

If D>2, sometimes it is convenient to 'map' the D-dimensional data to the 2-dimensional space. We say that we map (or project) the data from $\mathbb{R}^D$ to $\mathbb{R}^d$ (d<D)

Bottom-up view of the analysis

Data pre-processing → Data representation → Repr. pre-processing → Model selection/ training

If D>2, sometimes it is convenient to 'map' the D-dimensional data to the 2-dimensional space. We say that we map (or project) the data from $\mathbb{R}^D$ to $\mathbb{R}^d$ (d<D)

For a linear mapping, we have: $\mathbf{z}_i = \mathbf{W}^T \mathbf{x}_i$

We call **W** as the 'projection matrix'.

Bottom-up view of the analysis

Data pre-processing

Data representation

Repr. pre-processing

Model selection/ training

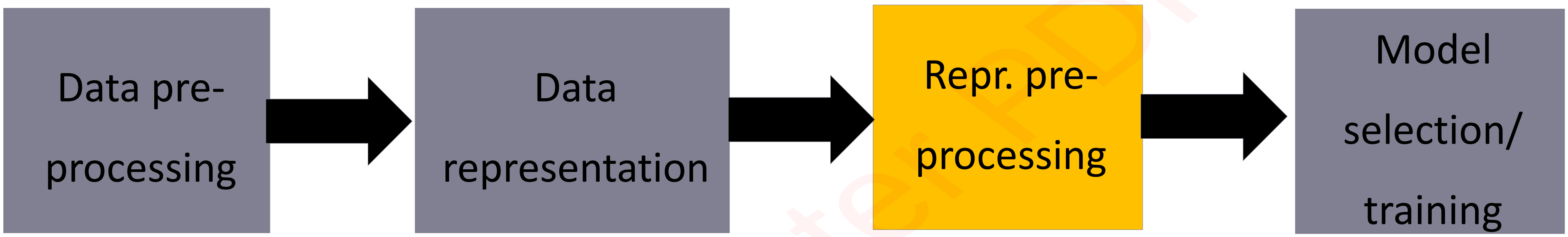
If D>2, sometimes it is convenient to 'map' the D-dimensional data to the 2-dimensional space. We say that we map (or project) the data from $\mathbb{R}^D$ to $\mathbb{R}^d$ (d<D)

For a linear mapping, we have: $\mathbf{z}_i = \mathbf{W}^T \mathbf{x}_i$

We call **W** as the 'projection matrix'.

W is a matrix with D rows and d columns.

Bottom-up view of the analysis



Other data representations pre-processing types:

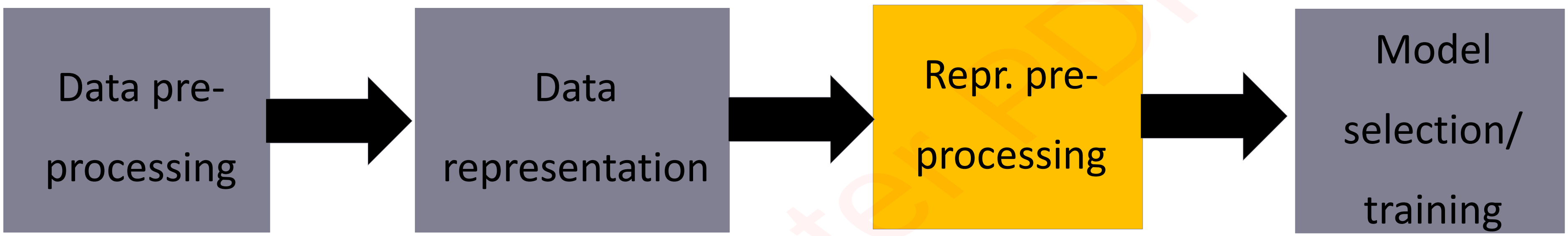
Data centering

$$\bar{\mathbf{x}}_i = \mathbf{x}_i - \boldsymbol{\mu}, \;\; i = 1, \ldots, N$$

$$\boldsymbol{\mu} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i$$

We move the data points to origin

Bottom-up view of the analysis



Other data representations pre-processing types:

Data standardization

$$\hat{x}_{id} = \frac{x_{id} - \mu_d}{s_d}$$

$$s_d = \frac{1}{N-1}\sqrt{\sum_{i=1}^{N}(x_{id} - \mu_d)^2}$$

Standardisation:
We center the data and each dimension is scaled so they have standard diviation is 1.

Why do we do this?
We want small changes in the weights to have small impacts.

Bottom-up view of the analysis

Data pre-processing → Data representation → Repr. pre-processing → Model selection/ training

Other data representations pre-processing types:

Data normalization (l2)

$$\tilde{\mathbf{x}}_i = \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|_2} = \frac{\mathbf{x}_i}{\sqrt{\mathbf{x}_i^T \mathbf{x}_i}}$$

All the data points are mapped to the unit circle

Data normalization (l1)

(usually for positive-valued vectors)

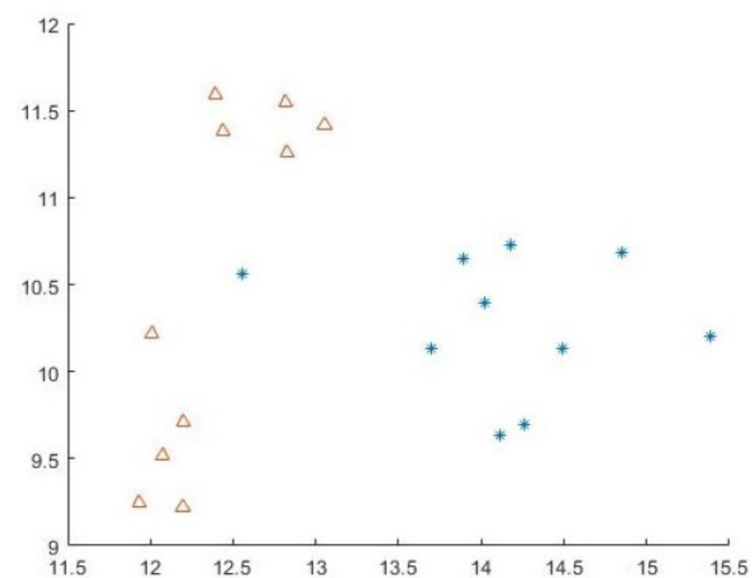
$$\dot{\mathbf{x}}_i = \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|_1} = \frac{\mathbf{x}_i}{\sum_{d=1}^{D} x_{id}}$$

Bottom-up view of the analysis

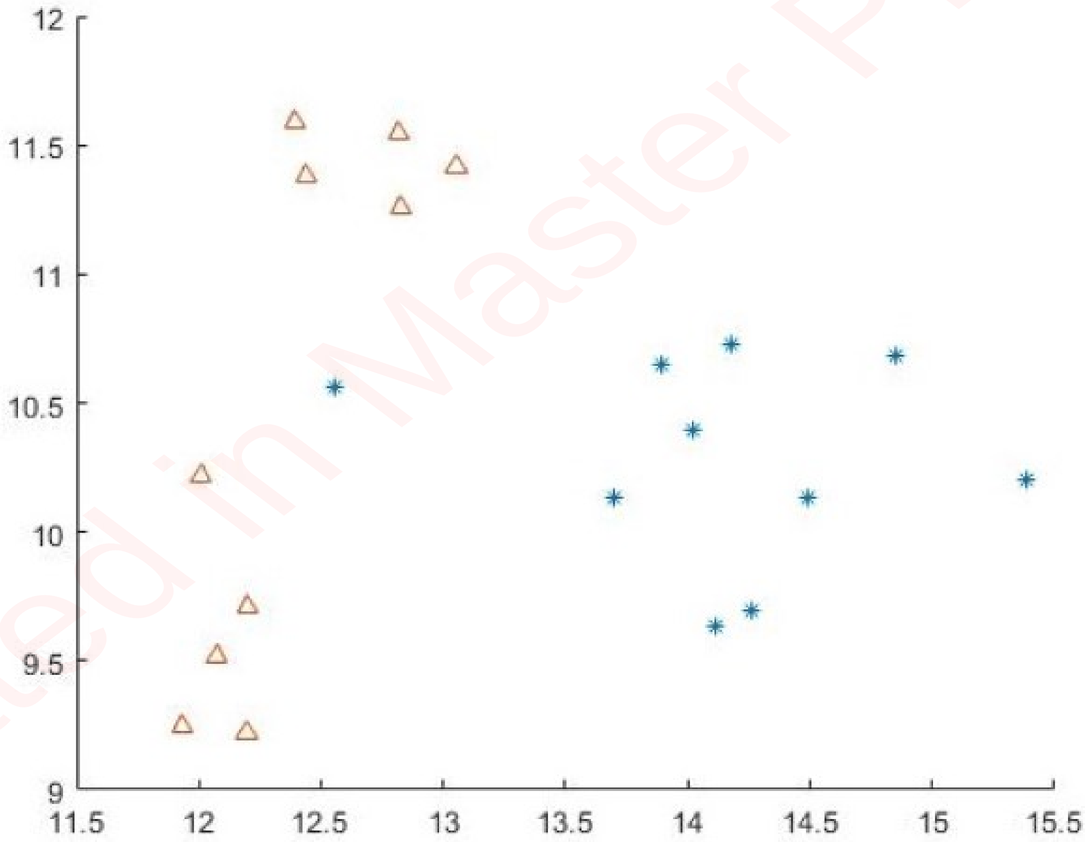
Data pre-processing → Data representation → Repr. pre-processing → Model selection/ training

Let us assume that we have the following set of data

Can you draw a ‘good decision function’ classifying the stars and the triangles correctly?

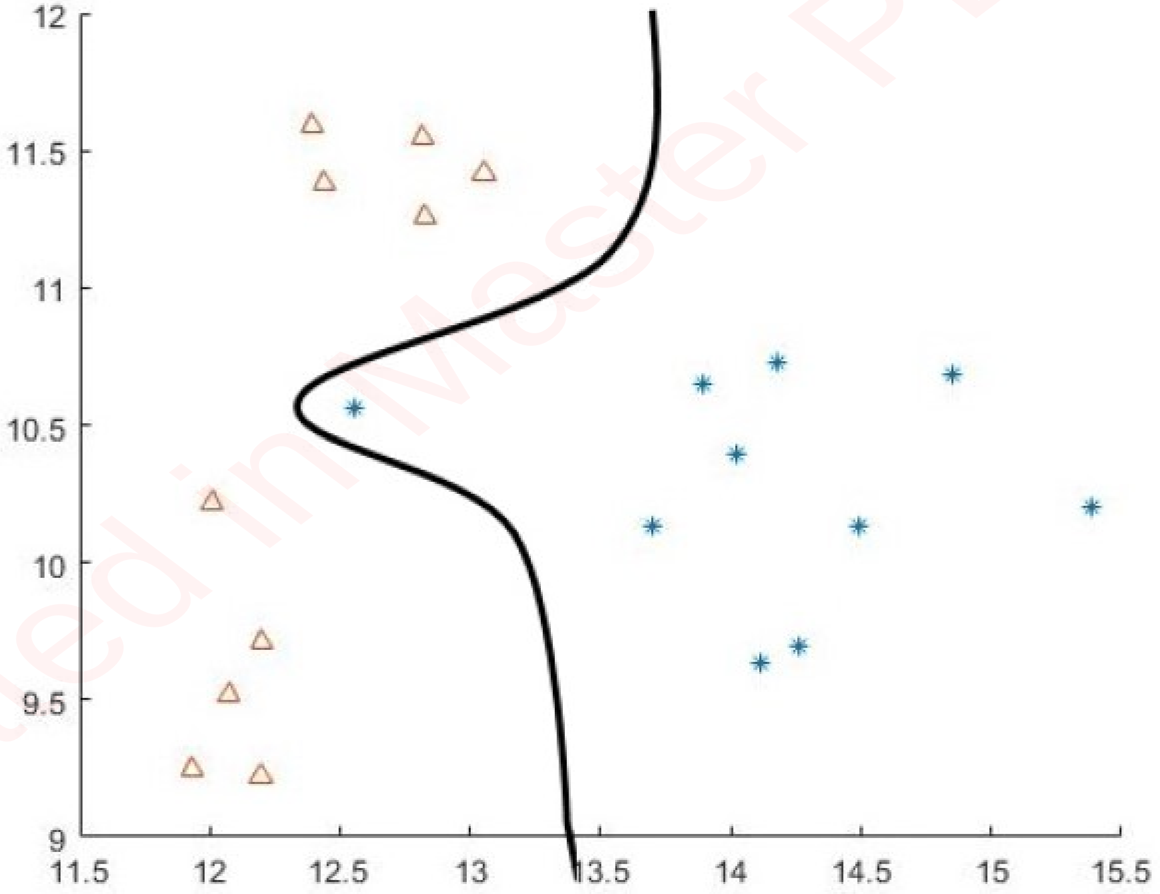


Model selection



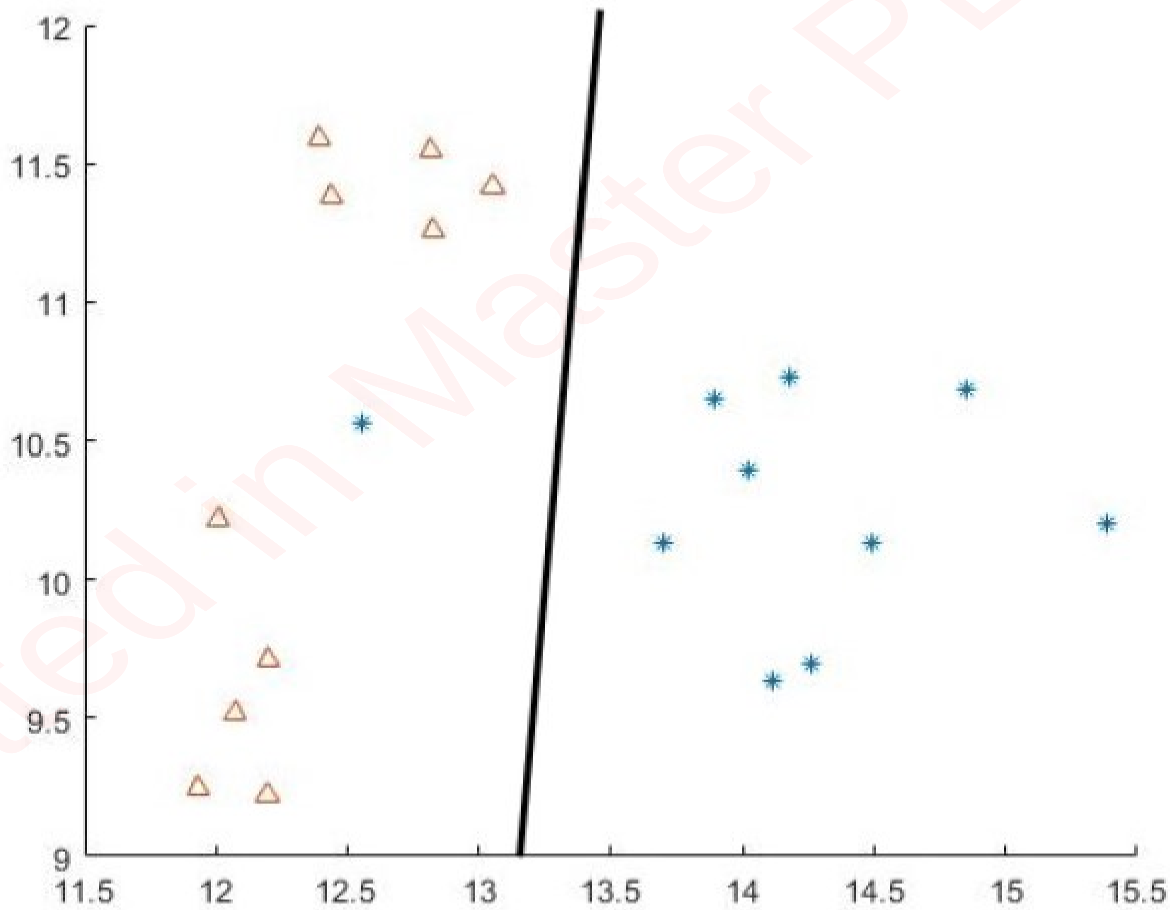
Model selection

This model has zero emperical error



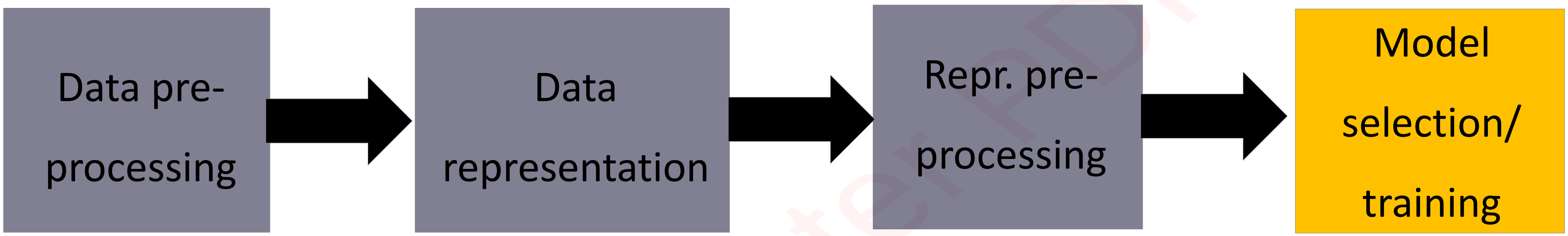
Model selection

This model has relatively low emperical error.
But it is a simple model because it is linear



Karstoft-Schultz-Iosifidis
Department of Engineering

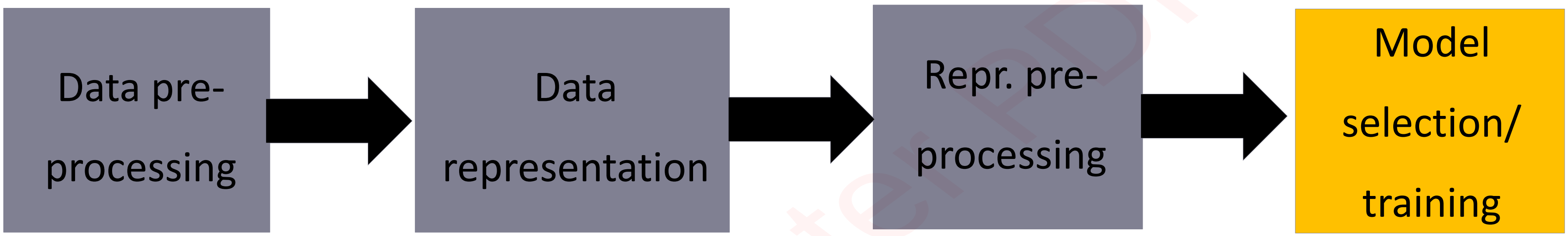
Bottom-up view of the analysis



Model selection is the process of:

- Defining what technique will be used (e.g. linear classifier, nonlinear classifier, etc.)
- Defining the 'optimal parameters' of the model. This includes:
    - The best 'architecture of the model' (e.g. in neural networks)
    - The **optimal values** of the parameters of the model

Bottom-up view of the analysis



Model selection is the process of:

- Defining what technique will be used (e.g. linear classifier, nonlinear classifier, etc.)
- Defining the 'optimal parameters' of the model. This includes:
    - The best 'architecture of the model' (e.g. in neural networks)
    - The **optimal values** of the parameters of the model

In order to define optimality, we formulate an optimization criterion involving the parameters of the model. Optimizing it (i.e. finding the parameter values corresponding to the minimal/maximal criterion value) gives us the solution.

Categorization of ML models

In *Supervised Learning*, the information available for model selection is formed by a set of training samples and their labels. This means that the training set has been labeled by a human expert and these labels can be used by the model selection process in order to define a decision function which is in line with the decisions made by the expert. While in this way the model selection process can exploit valuable information related to the problem at hand, labeling is a tedious and expensive process and this is why labeled sets are rare and small in size.

Can you describe some supervised learning problems?

Categorization of ML models

In *Unsupervised Learning*, there is no expert who can provide domain knowledge for the problem at hand. Instead, there is the assumption that samples naturally form groups based on a similarity criterion and the model selection process will be able to reveal this similarity function and identify the underlying groups of patterns.

Can you describe some unsupervised learning problems?

Categorization of ML models

In *Reinforcement Learning*, model selection is guided by an expert who can provide only binary feedback. This means that the training process follows an iterative process where the model is presented with a sample and makes a guess. The expert observes this guess and indicates if it is correct or not. Given this feedback, the training process can either update the model (in case of an error) or continue by using another training sample.

It is extremely slow to train this model compared to the supervised model where we tell the agent where it should go!

But it has many advantages.

Can you describe some reinforcement learning problems?

https://cs.stanford.edu/people/karpathy/convnetjs/demo/rldemo.html
https://www.youtube.com/watch?v=PSQt5KGv7Vk