

Neural Networks and Learning Systems  
TBM126 / 732A55  
2023

**Lecture 1**  
**Introduction**

***Magnus Borga***

Department of Biomedical Engineering (IMT)

*magnus.borga@liu.se*

# Course information

- All information will be available on Lisam
- You must register for classes on Lisam (Signup)
  - Chose group A, B , or C for classes
  - Do not change group!
- Laboratory exercises will be done at home with scheduled supervision on Teams
  - Dual occasions for each lab. You may chose occasion (Wed or Thu)

# Staff

- **Examiner:** Magnus Borga, IMT (magnus.borga@liu.se)
- **Course admin:** Martin Hultman, IMT ([martin.o.hultman@liu.se](mailto:martin.o.hultman@liu.se))
- Lectures:
  - Magnus Borga
  - Anders Eklund
- Classes:
  - Martin Hultman
  - Iulian Tampu
- Laboratory exercises:
  - Martin Hultman
  - Iulian Tampu
  - Elisabeth Klint
  - Gustav Magnusson

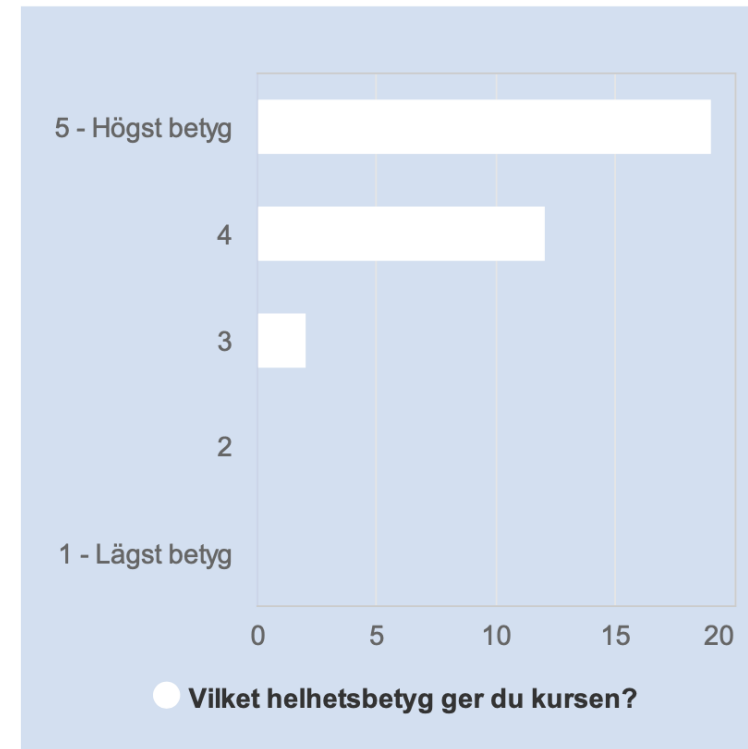
# Course evaluation and development

## 9. Vilket helhetsbetyg ger du kursen?

Vilket helhetsbetyg ger du kursen?	Antal svar
5 - Högst betyg	19 (58%)
4	12 (36%)
3	2 (6%)
2	0 (0%)
1 - Lägst betyg	0 (0%)
Summa	33 (100%)

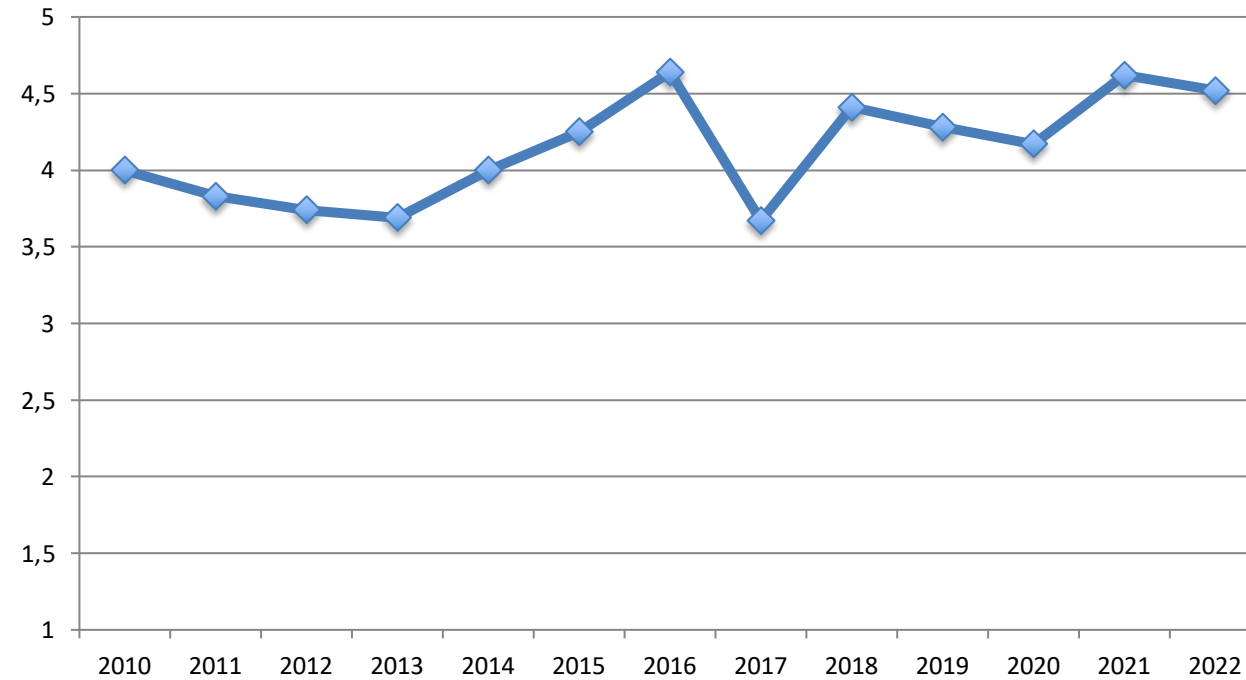
Average score 2022: 4.52

## Course evaluation 2022



	Medelvärde	Standardavvikelse	Variationskoefficient	Min	Undre kvartil	Median	Övre kvartil	Max
Vilket helhetsbetyg ger du kursen?	4.52	0.62	13.70 %	3.00	4.00	5.00	5.00	5.00

# Overall credit over time



# Course evaluation and development

## **Changes 2023**

- Further improvements on the laboratory exercises

# The Course - Lectures

PPT lectures, handouts on course page

1. Introduction
2. Supervised learning - Linear classification
3. Neural networks
4. Deep learning
5. Generative Adversarial Networks
6. Ensemble learning & Boosting methods
7. Reinforcement learning
8. Unsupervised learning – Dimensionality reduction, Clustering
9. Kernel methods

# The Course - Lessons

- Pen & paper exercises
- Complementary presentations
- Preparations and help with lab assignments
- Choose group (A,B, or C) on Lisam and follow that group



# The Course – Lab assignments

- 4 laboratory exercises/assignments:
  1. Pattern recognition using linear classifiers and neural networks
  2. Deep learning
  3. Face recognition in images using Boosting techniques
  4. Reinforcement learning
- Programming in Python
- Assignments are done in pairs. (Not more than 2 students together!)
- Supervision in Teams
  - time scheduled (“LA” in schedule)
  - Important to turn on notifications in Teams
- Possible oral examination during scheduled supervision if time allows! (The code still needs to be submitted in Lisam.)
- Deadlines for written reports (unless passed on oral examination) one week after scheduled lab!
- Late reports may not be corrected until next re-exam (June/August).
- Labs not passed at the August exam will be registered as failed in Ladok and need to be redone next year

# Course literature

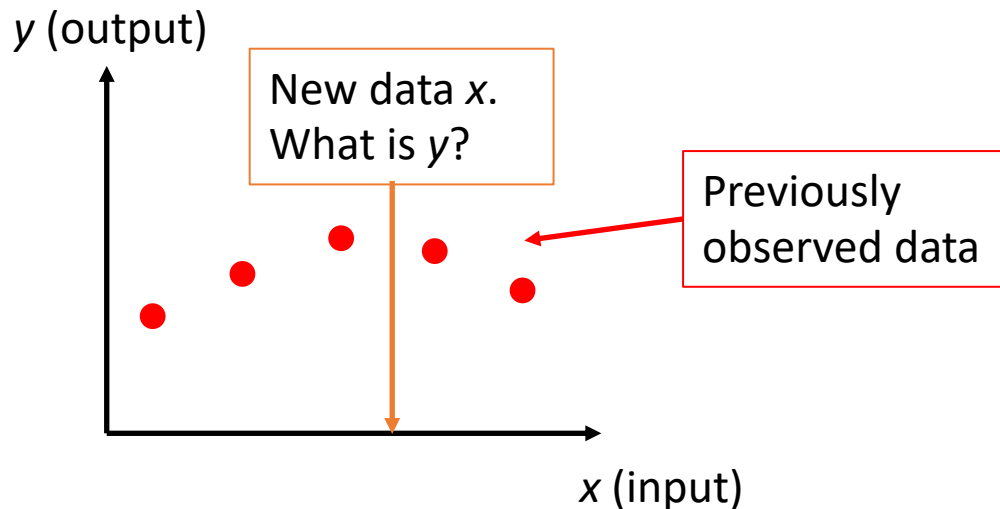
- Lecture notes
- Exercise collection
- Assignments
- Additional links in lecture notes (not required reading)

# Machine Learning

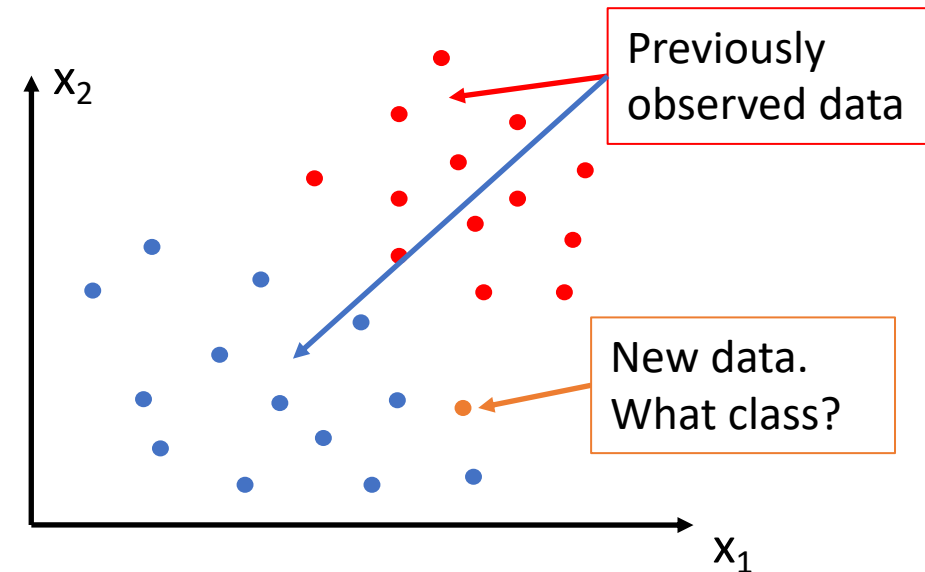
# What is machine learning?

(My) **definition:** The ability of a system to extract relevant information from previous data and generalize to new data

Example: Regression



Example: Classification



# How can a machine learn?

Learning: "Any relatively permanent change in behaviour resulting from past experience."

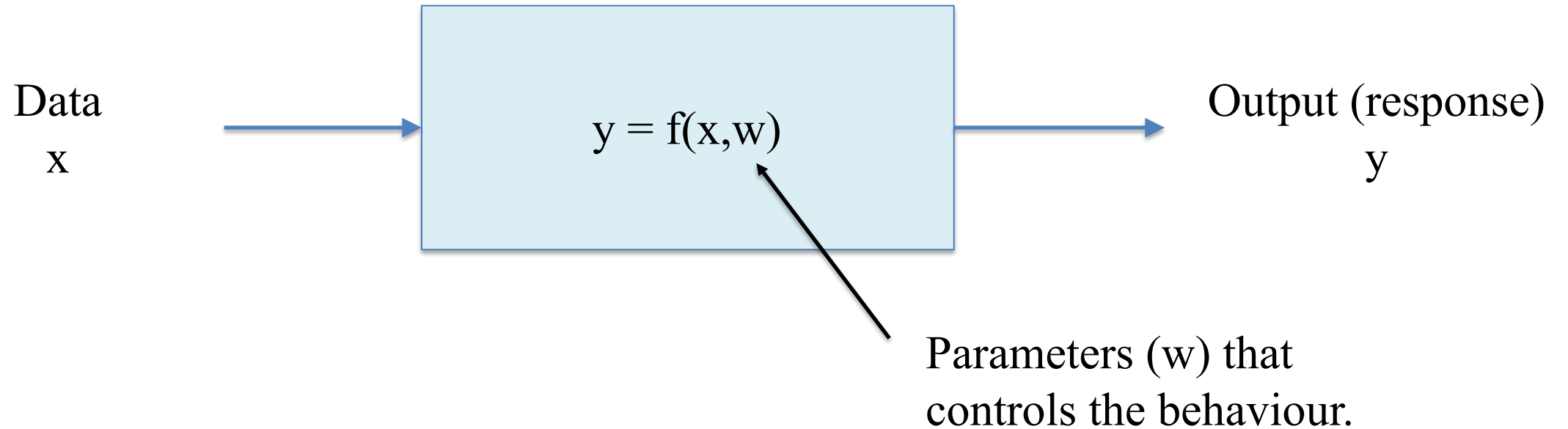
(Encyclopaedia Britannica 1964)

- The "behaviour" of the machine is determined by its parameters.
- "past experience" is previously observed data.



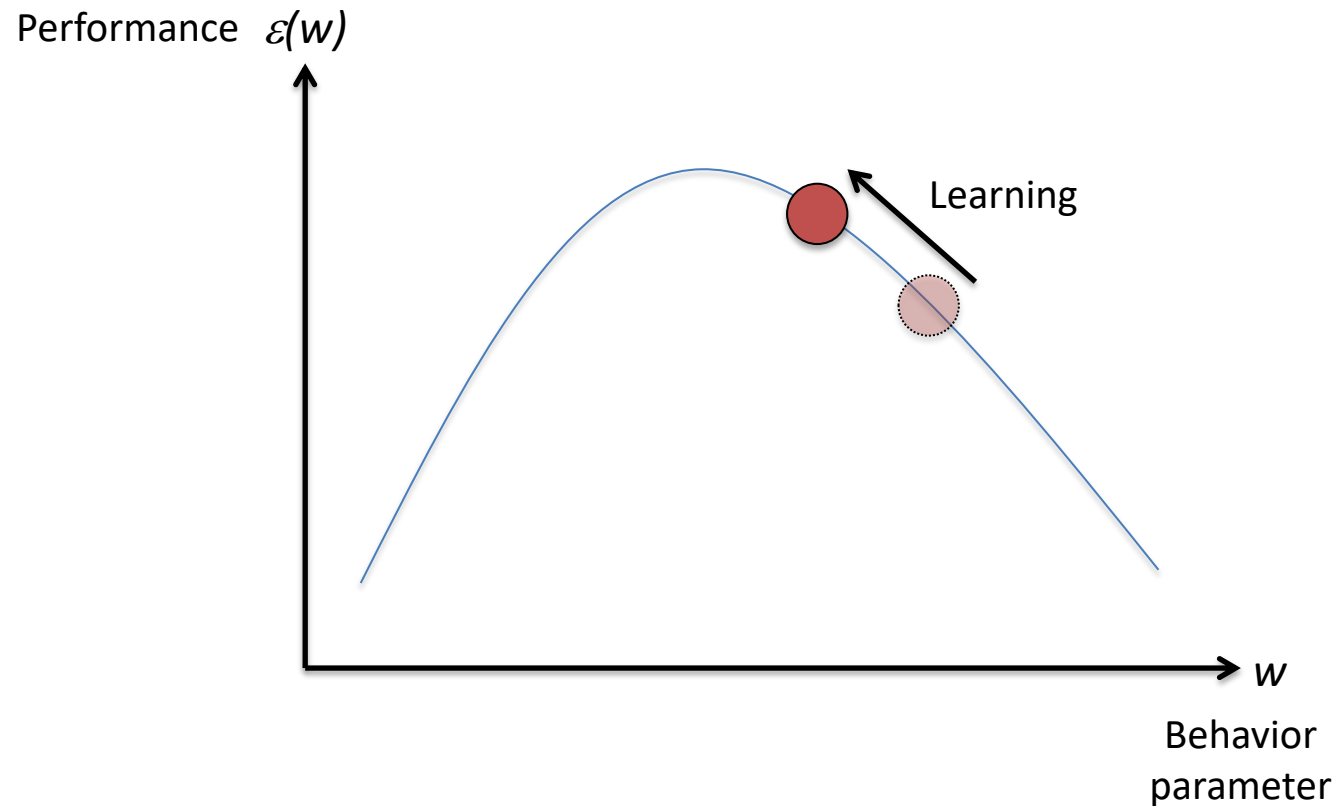
- Machine learning = changing parameter values (behaviour) as a result of observed data (experience).

# Machine Learning



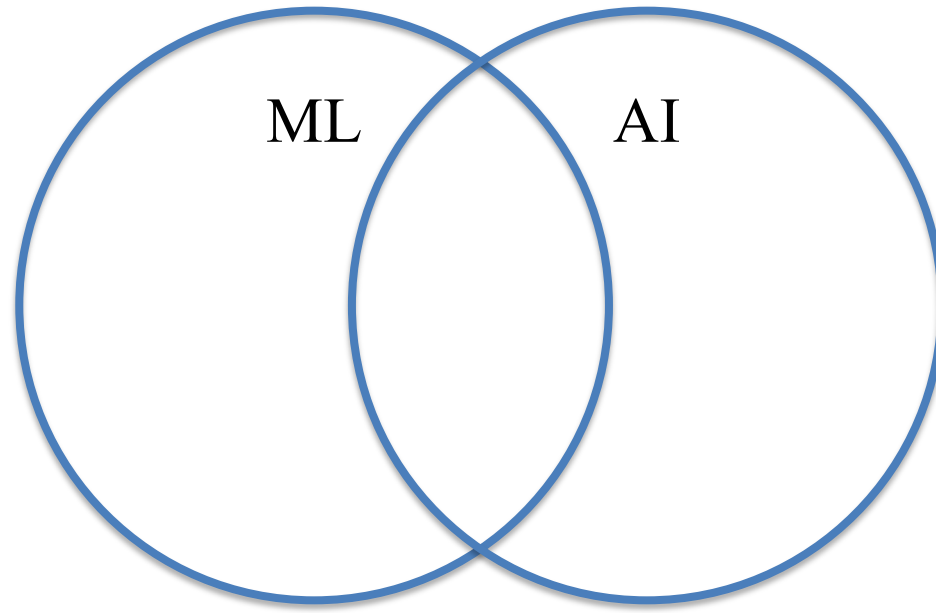
Machine learning: The system changes its parameters ( $w$ ) in order to improve its performance

# Learning = optimization of performance



What is being learnt depends on how we define performance!

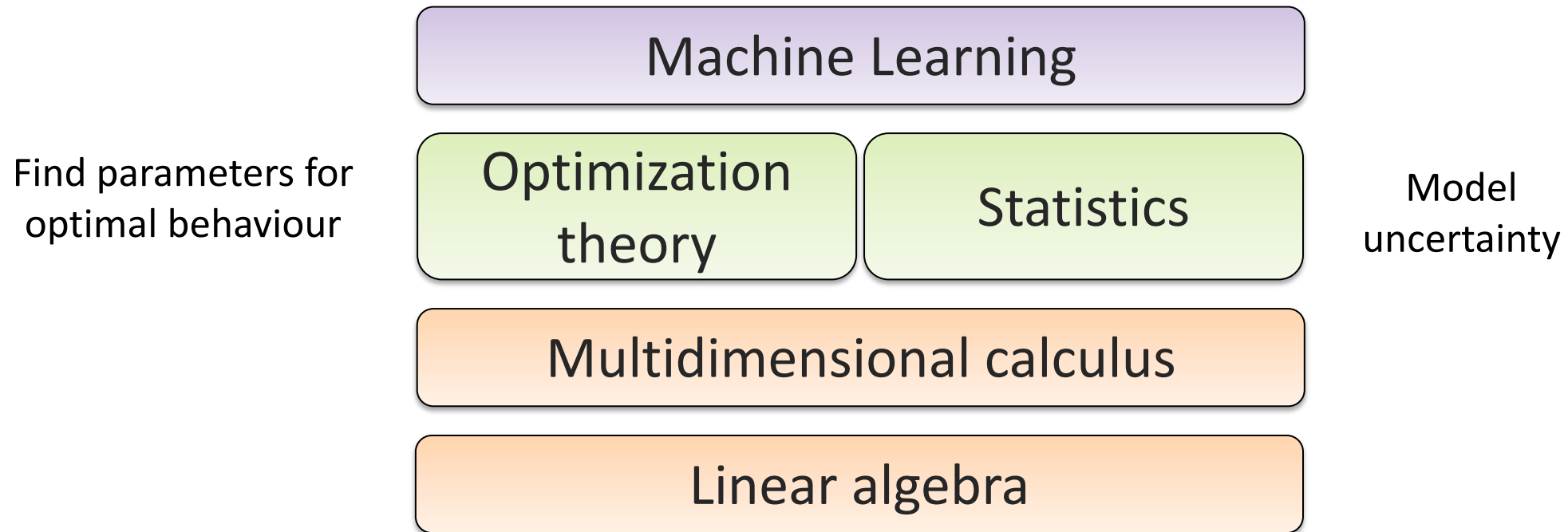
# ML vs AI



- AI aims at simulating “intelligent” behaviour, but not necessarily by learning
- ML does not always aim at doing something “intelligent”



# Mathematical foundations of machine learning



# Prerequisites

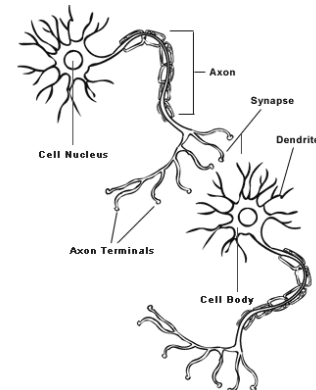
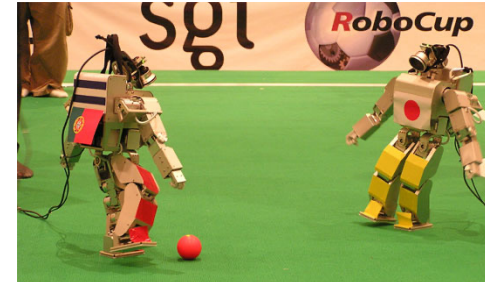
- Linear algebra
  - Vectors, scalar products, eigenvectors and eigenvalues
- Multidimensional calculus
  - Gradients, partial derivatives
- Mathematical statistics
  - Mean, variance, covariance, correlation, Gaussian distribution,
- Programming
  - Some programming experience
  - Python experience helps a lot

# Why machine learning?

- Algorithm too complex for a human to design, but we can easily provide examples of what the algorithm should do.
- Relationships in high-dimensional data too complex for a human to see, but a computer can find these.
- The computer should learn and adapt continuously to new situations.

# Applications of machine learning

- Pattern and speech recognition
- Robots & autonomous systems
- Big data
- Expert systems & decision support
- Games
- Models of the brain



# Three main categories of machine learning methods

- **Supervised learning (predictive)**  
Learn to generalize and classify new data based on labelled training data.
  - Pattern recognition
  - Classification
  - Regression
- **Unsupervised learning (descriptive)**  
Discover structure and relationships in complex high-dimensional data.
- **Reinforcement learning (active)**  
Generate policies/strategies that lead to a (possibly delayed) reward. Learning by doing.

# Classification vs. regression vs. ranking

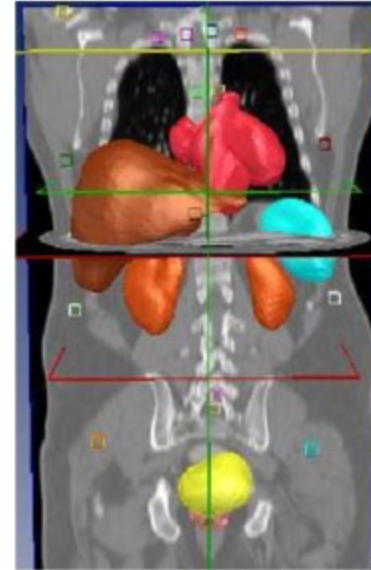
- **Classification:** Select one of a discrete set of classes (the set  $\Omega$  is discrete).
  - Which horse is going to win this race?
  - Which letter does this image depict?
  - Is this email spam (yes/no)?
- **Regression:** Learn to predict a continuous value ( $\Omega = \mathbb{R}$ ).
  - Learn to predict the temperature tomorrow.
  - What is the *probability* that this image depicts the letter 'a'?
- **Ranking:** Learn to rank a set of items ( $\Omega = \mathbb{R}$ ).
  - Rank webpages, movies, etc.

# Pattern recognition examples

Face recognition



Organ segmentation



# Pattern recognition examples

## Optical Character Recognition (OCR)

0 0 0 0..  
1 1 1 1..  
2 2 2 2..  
9 9 9 9..

Source: <http://blog.damiles.com>

How do I optimize the text conversion?  
You must write from left to right.  
The lines of text must be horizontal.  
Try and maintain a steady writing direction.  
Keep the size of the letters relatively constant.  
An upper-case letter is twice the size of a lower-case.  
Leave enough space between words.  
You cannot edit a sentence, once it has been captured.  
Add unrecognized words to your dictionary.

The following sentences are random and historical facts;  
they allow us to collect the remainder of the writing samples.  
TINTIN was first published in 1929.  
The Cape Verde Islands are in the Atlantic Ocean.  
The 1000 Lakes Rally takes place in Finland.  
Goulash is a Hungarian beef stew!  
Dunlop invented the bicycle wheel in 1888.  
Rio de Janeiro is overlooked by Sugarloaf Mountain.  
Concordia's first flight was on 2 March 1969.  
An alexandrine is a verse of twelve syllables.  
The top of Mount EVEREST is 8,848m high (Himalayas).  
On 21 July 1969, Neil Armstrong walked on the Moon.  
Oliver Stone made the film Platoon in 1987.  
Honshu is the largest island in the Japanese archipelago.  
A sheet of A4 paper measures 21x29.7 cm.  
The island of Cuba is 180 km south of Florida.  
The Richter scale measures the magnitude of earthquakes.

How do I optimize the text conversion?  
You must write from left to right.  
The lines of text must be horizontal.  
Try and maintain a steady writing direction.  
Keep the size of the letters relatively constant.  
An upper-case letter is twice the size of a lower-case.  
Leave enough space between words.  
You cannot edit a sentence, once it has been captured.  
Add unrecognized words to your dictionary.

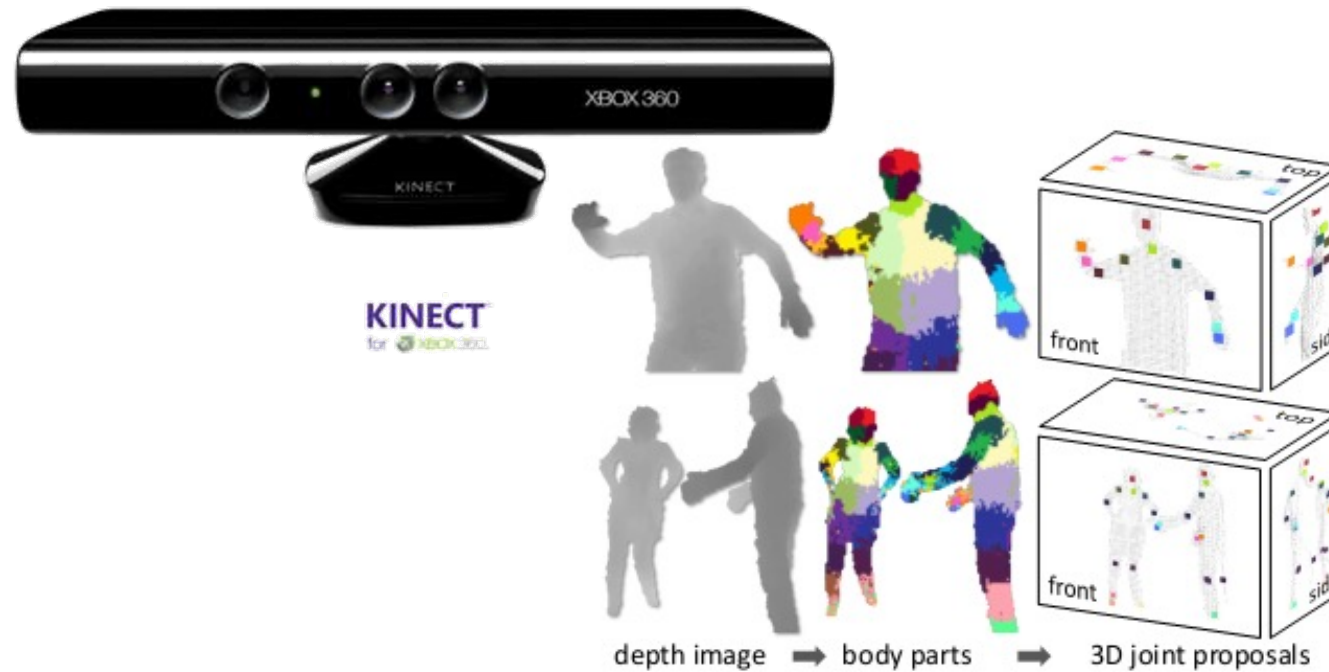
The following sentences are random and historical facts;  
they allow us to collect the remainder of the writing samples.  
TINTIN was first published in 1929.  
The Cape Verde Islands are in the Atlantic Ocean.  
The 1000 Lakes Rally takes place in Finland.  
Goulash is a Hungarian beef stew!  
Dunlop invented the bicycle wheel in 1888.  
Rio de Janeiro is overlooked by Sugarloaf Mountain.  
Concordia's first flight was on 2 March 1969.  
An alexandrine is a verse of twelve syllables.  
The top of Mount EVEREST is 8,848m high (Himalayas).  
On 21 July 1969, Neil Armstrong walked on the Moon.  
Oliver Stone made the film Platoon in 1987.  
Honshu is the largest island in the Japanese archipelago.  
A sheet of A4 paper measures 21x29.7 cm.  
The island of Cuba is 180 km south of Florida.  
The Richter scale measures the magnitude of earthquakes.

MJ



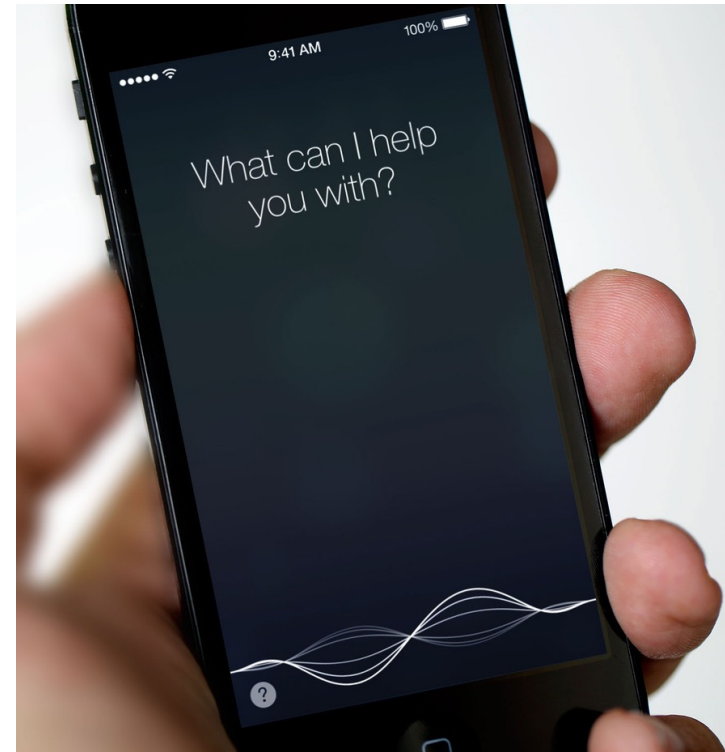
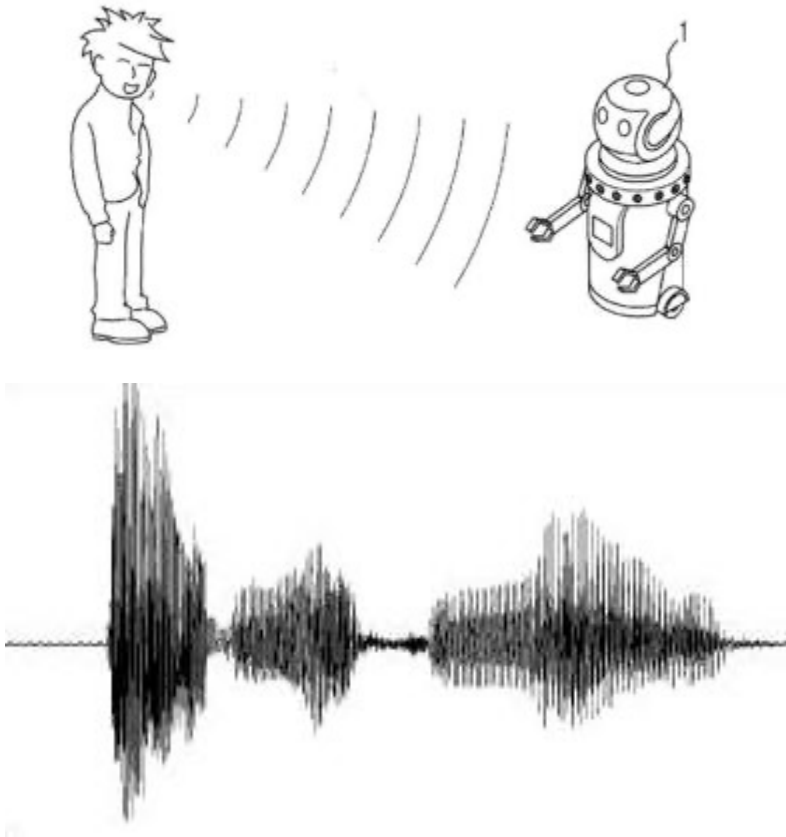
# Pattern recognition examples

## Xbox Kinect – Pose estimation



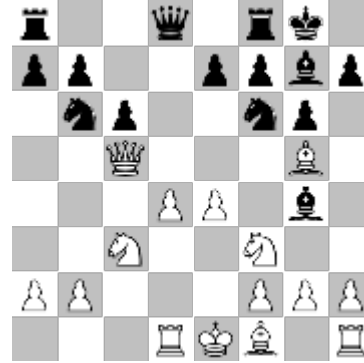
# Pattern recognition examples

## Speech recognition



# Pattern recognition examples

## Game positions



## Spam filters



#1  
100% satisfied  
4U  
Accept credit cards  
Act Now!  
Additional Income  
Affordable  
All natural

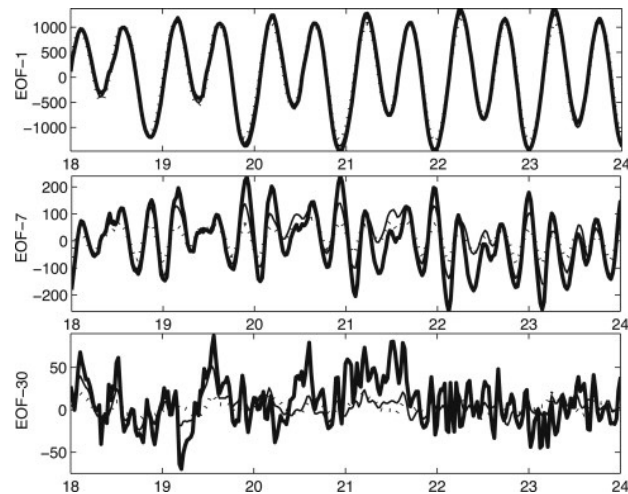
## Movie & music recommendation



# Regression examples

## Prediction and forecasting

### Weather and natural phenomena



### Financial markets



# Features

- A feature is a measurement or scalar number that describes some aspect of a phenomenon or object
  - Size, length, shape, velocity
  - Intensity and color (RGB)
  - Position (x,y)
  - Signal frequency
  - Sensor measurements (e.g., temperature)
  - Game piece present at certain location (yes/no)
  - Word present in an email (yes/no)
- Feature extraction is the process of measuring features from data.

# Features – Iris dataset



Iris setosa



Iris versicolor



Iris virginica

Fisher's Iris Data

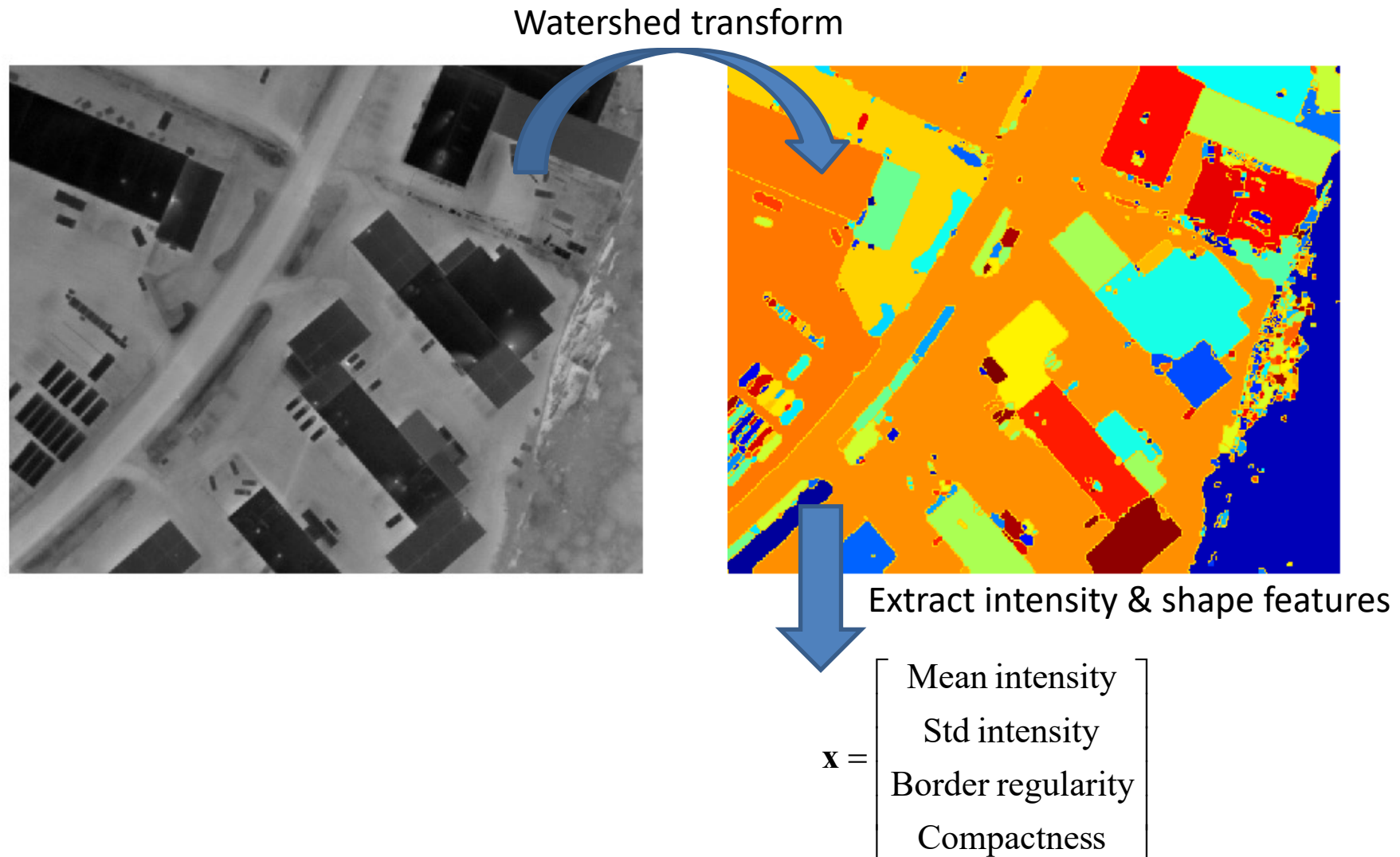
Sepal Length ♦	Sepal Width ♦	Petal Length ♦	Petal Width ♦	Species ♦
5.7	4.4	1.5	0.4	<i>I. setosa</i>
5.8	2.6	4.0	1.2	<i>I. versicolor</i>
5.8	2.7	5.1	1.9	<i>I. virginica</i>

From Wikipedia

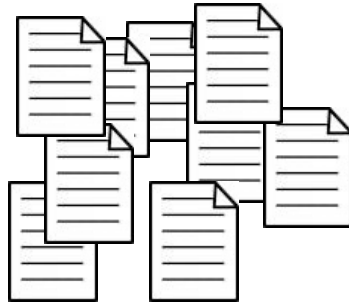
Feature vectors:

$$\mathbf{x}_1 = \begin{bmatrix} 5.7 \\ 4.4 \\ 1.5 \\ 0.4 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 5.8 \\ 2.6 \\ 4.0 \\ 1.2 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} 5.8 \\ 2.7 \\ 5.1 \\ 1.9 \end{bmatrix}$$

# Features – Image classification



# Features – Document analysis

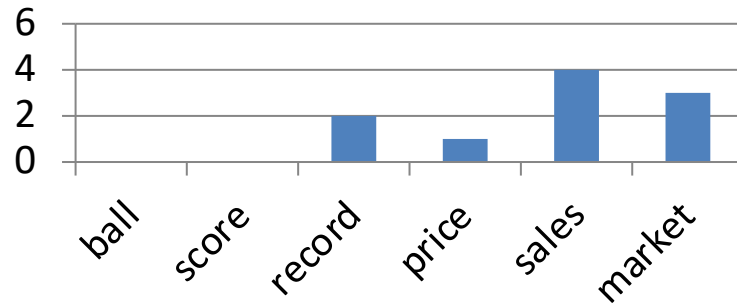


Articles, mail, web pages, ...

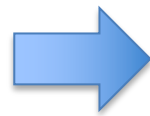
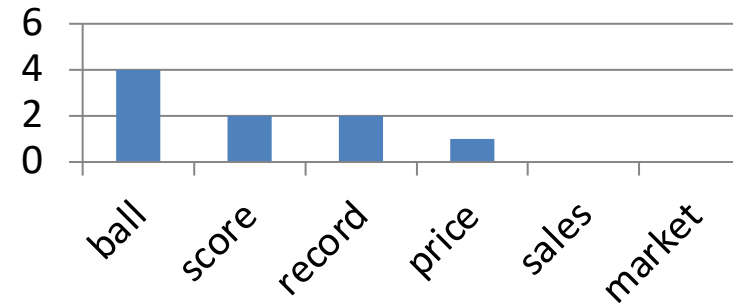
## Dictionary:

{'ball', 'score', 'record', 'price', 'sales', 'market'}

Financial document



Sports document

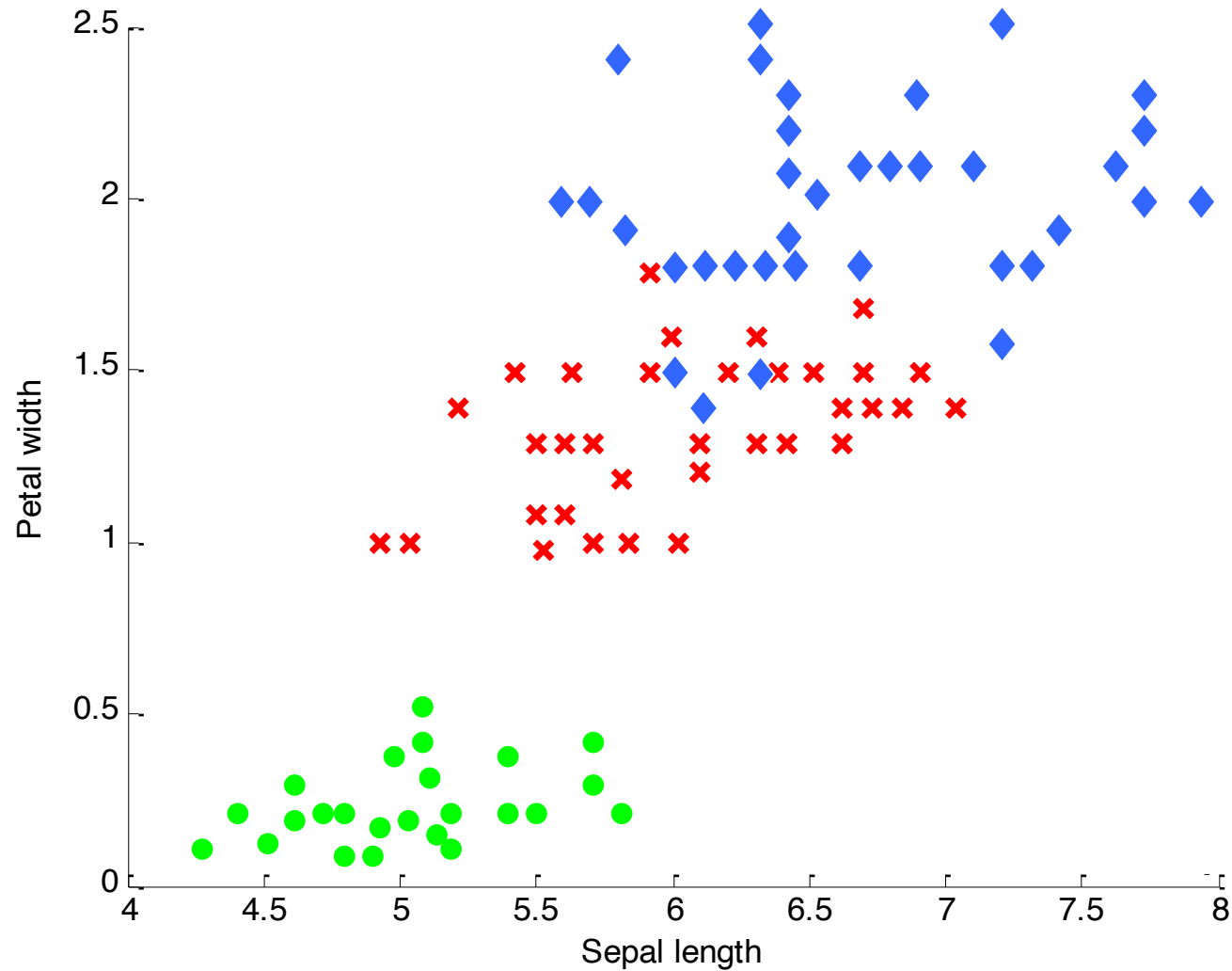


$$\mathbf{x} = \begin{bmatrix} \# 'ball' \\ \# 'score' \\ \# 'record' \\ \# 'price' \\ \# 'sales' \\ \# 'market' \end{bmatrix}$$

Bag-of-words model



# Feature space



Iris setosa



Iris versicolor



Iris virginica



# Supervised Learning

# Supervised learning

- **Task:** Learn to predict/classify new data from labeled examples.
- **Input:** Training data examples  $\{\mathbf{x}_i, y_i\}$ ,  $i=1\dots K$ , where  $\mathbf{x}_i$  is a feature vector and  $y_i$  is a class label in the set  $\Omega$ .
- **Output:** A function  $f(\mathbf{x}; w_1, \dots, w_N) \rightarrow \Omega$

Find a function  $f$  and adjust the parameters  $w_1, \dots, w_N$  so that new feature vectors are classified correctly. **Generalization!!**

# k-Nearest Neighbours (k-NN)

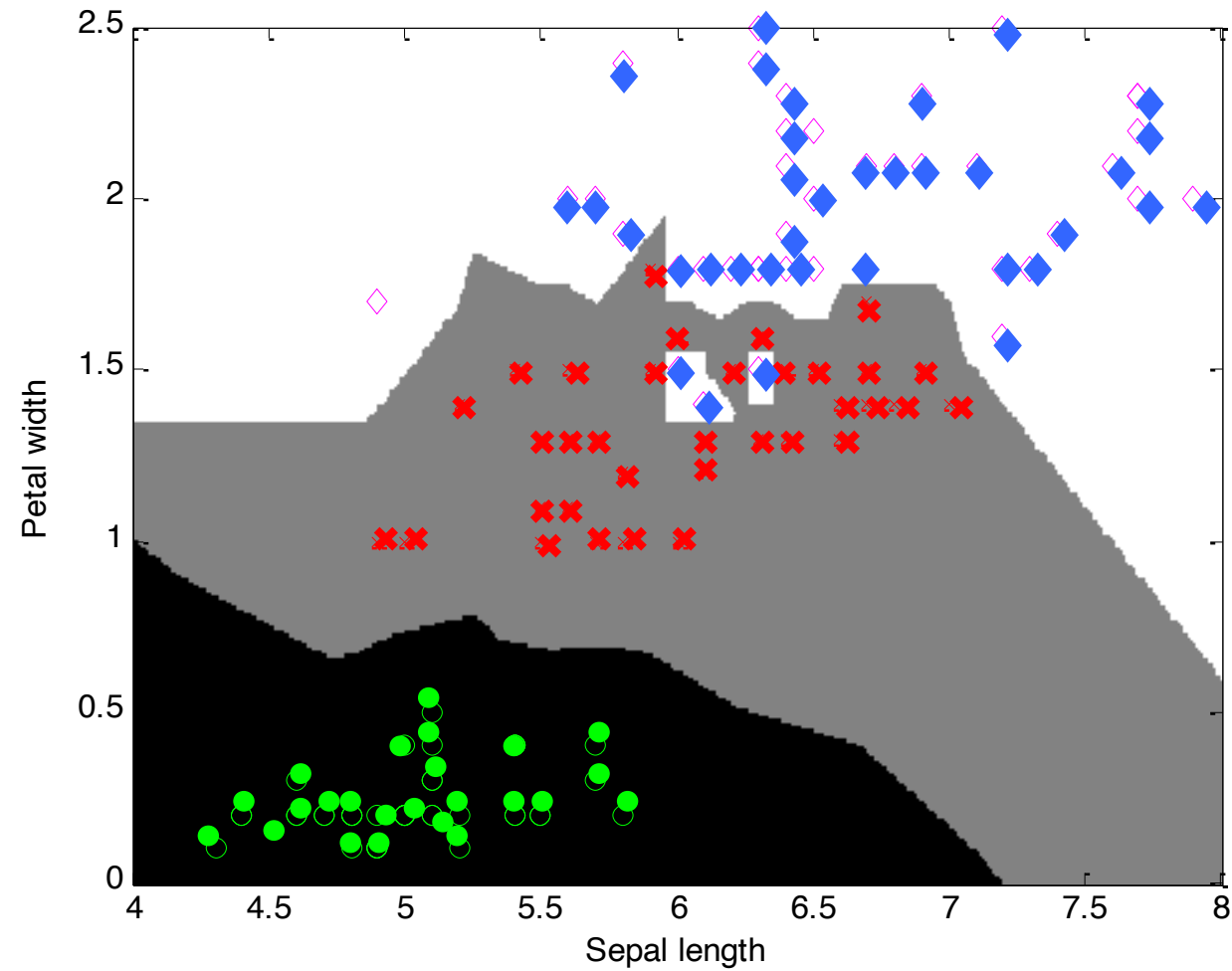
- Save all training data.
- For a new case, find similar examples among the training data.
- Requires a similarity measure (metric), for example the Euclidian distance

$$\|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_i (x_i - y_i)^2}$$

- A majority vote among the  $k$  nearest neighbours decides the class, where  $k$  can be 1,2,3,4...

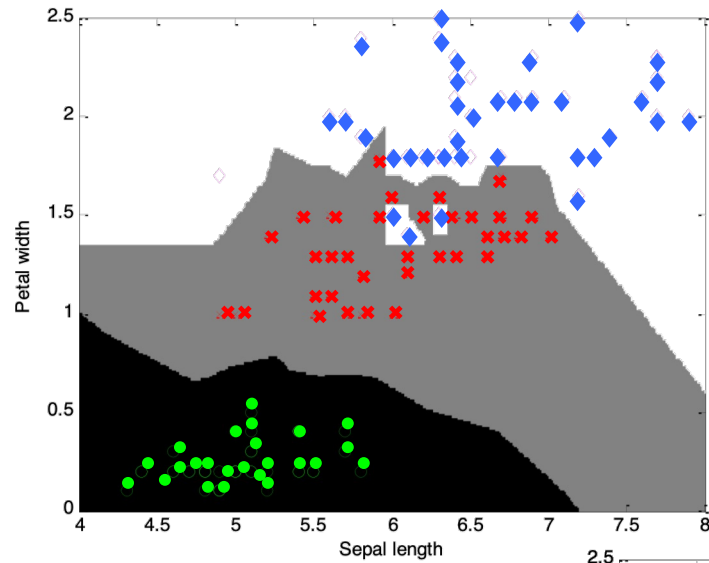
# Classification boundaries

$k = 1$

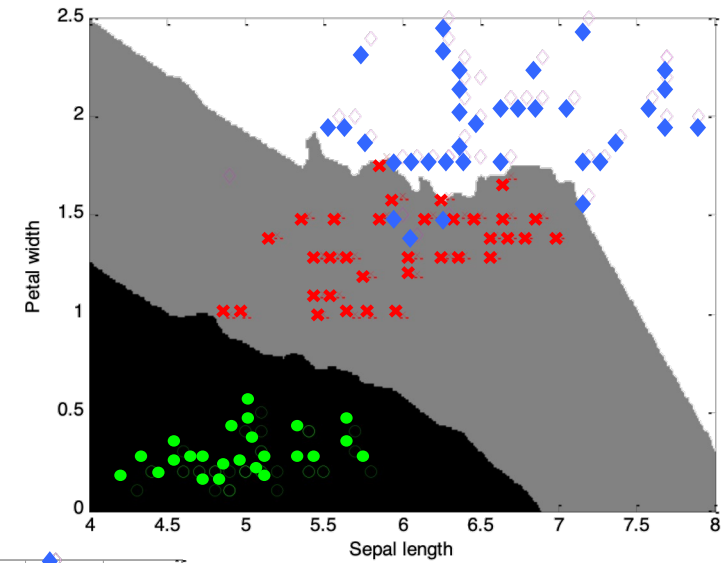


# Classification boundaries

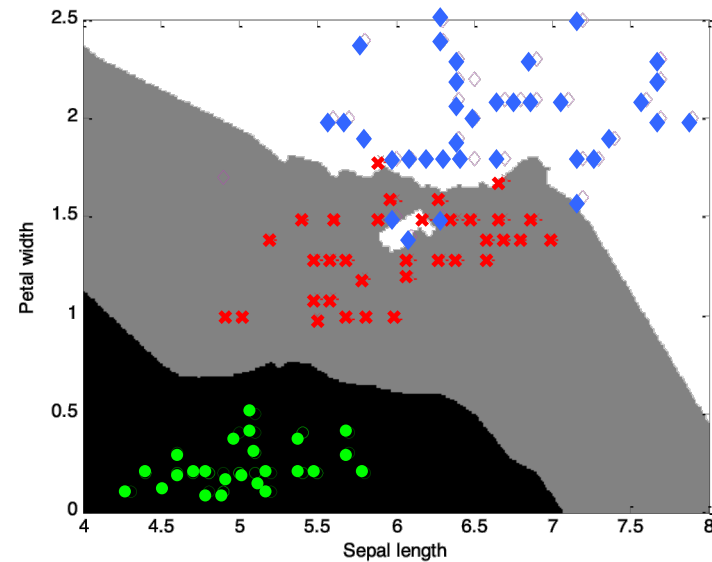
$k = 1$



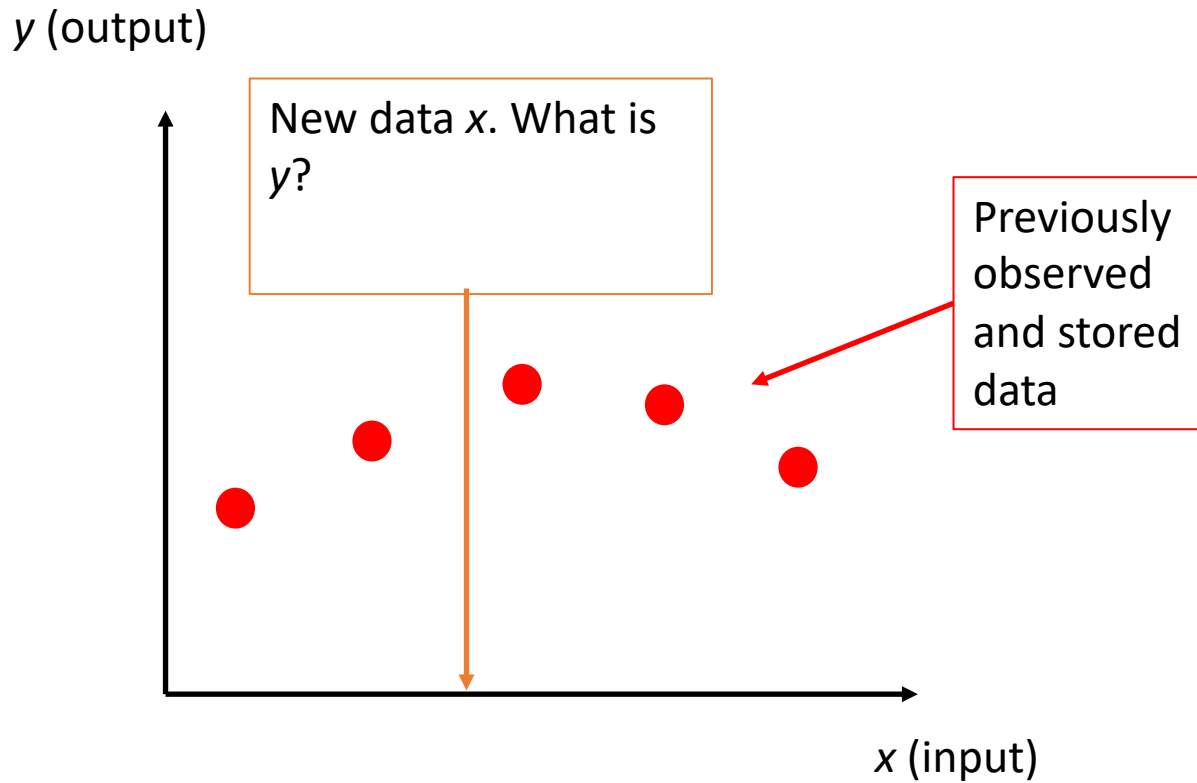
$k = 7$



$k = 3$



# k-NN for regression



Local regression model (instead of a global model)

## Examples:

- $k = 1$ 
  - Nearest neighbour interpolation
- $k = 2$ 
  - Use linear interpolation between the two closest points

# Pros and cons of k-NN

- Very simple – only one parameter ( $k$ )
- Simple to add new data
- Very fast to “train”
- Must store all training data – problem for large data sets:

$$f(\mathbf{x}; \underbrace{w_1, \dots, w_k}) \rightarrow \Omega$$

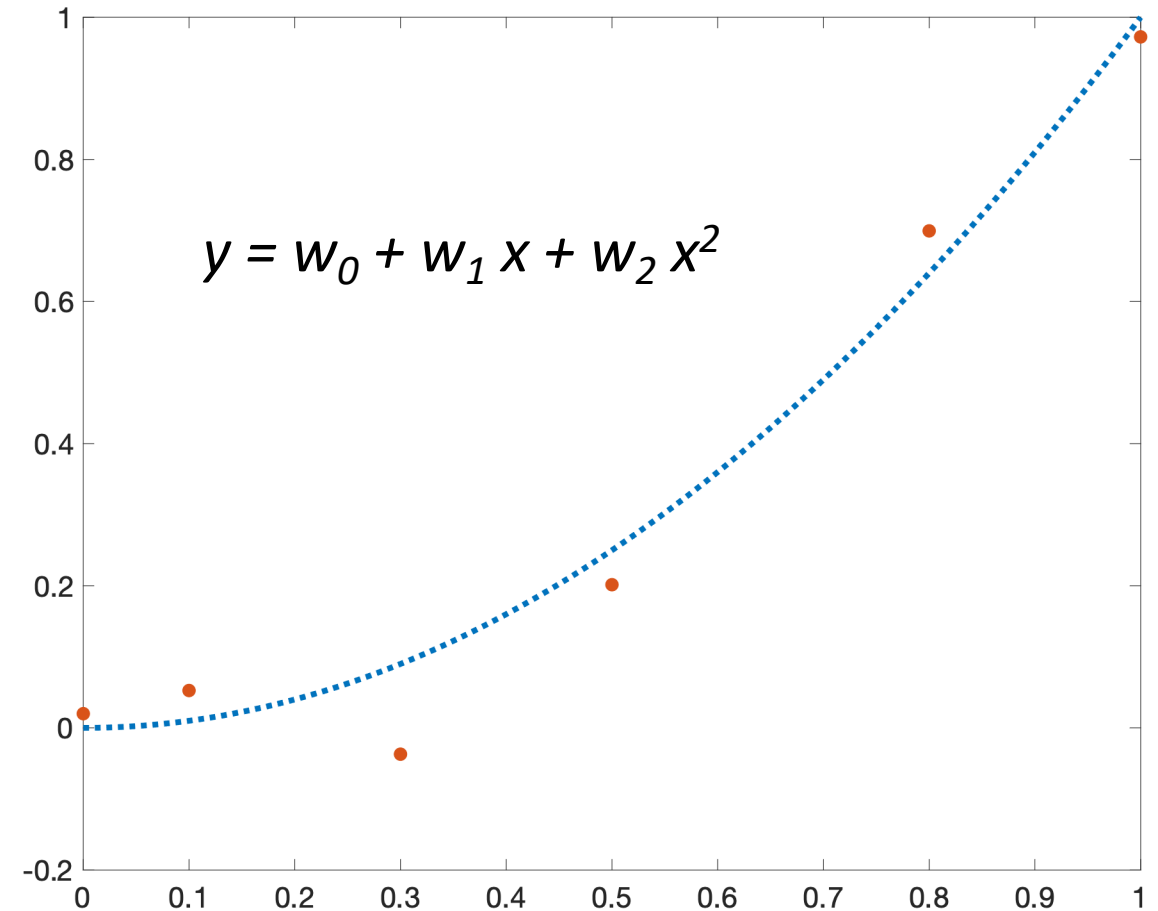
Parameters equal to training data  $\mathbf{x}_i$

- Computationally heavy for large data sets – must compare new samples with all stored samples.



# Parametric models

- Instead of storing all training data, a parametric model can be fitted
- Only need to store the parameters
- Very fast to compute the output
- May take long time to train the model



# Discriminant functions

- Classification can be seen as a function that takes one value e.g. +1, for one class and another value, e.g. -1, for the other class
- Such function is called a **discriminant function**
- This means that training a classifier becomes a function approximation or regression problem!

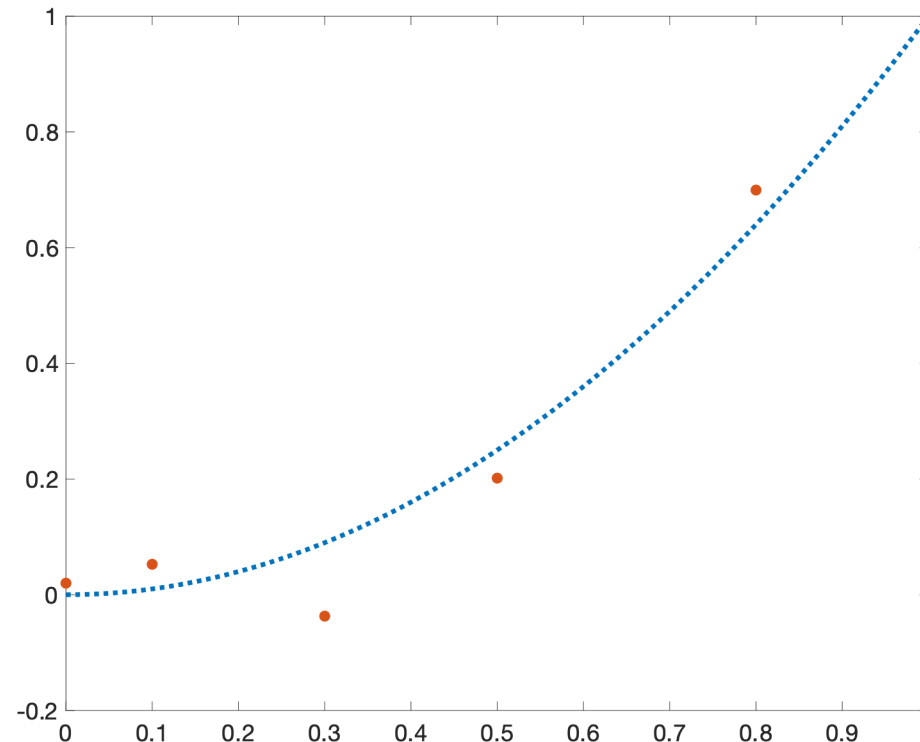
# Generalization is the key

- The important question is not how well a model can fit to existing data, but how well it can predict future data!
- Any (finite) data set can be perfectly fitted by a model complex enough.
- Do we need a perfect fit?
- Do we even *want* a perfect fit?!

# Overfitting

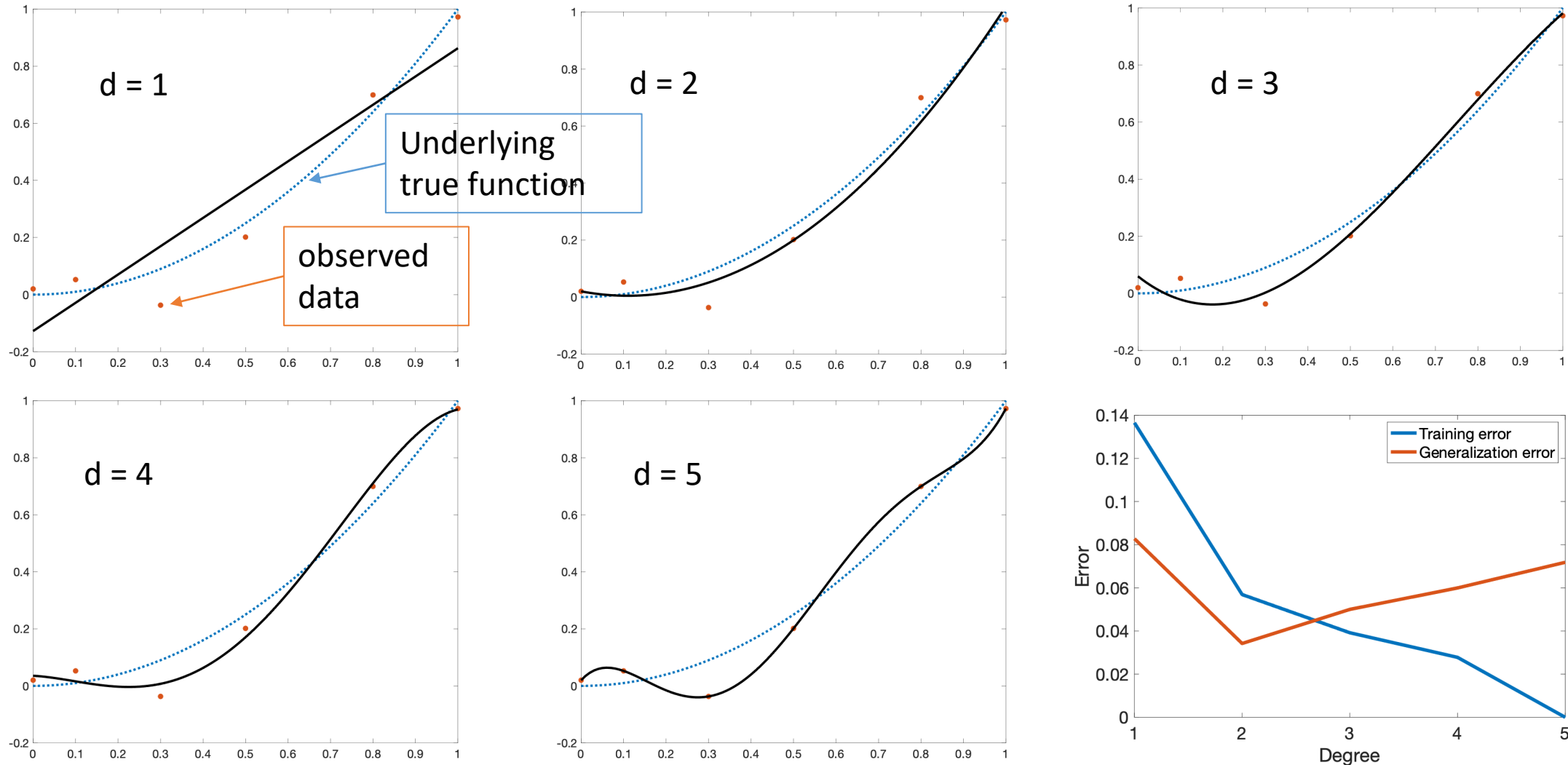
Example:

Assume we want to learn a relationship between  $x$  and  $y$  that in reality is quadratic. But we don't know that, we only have these sample points, which (as always) contain some measurement errors or “noise”:



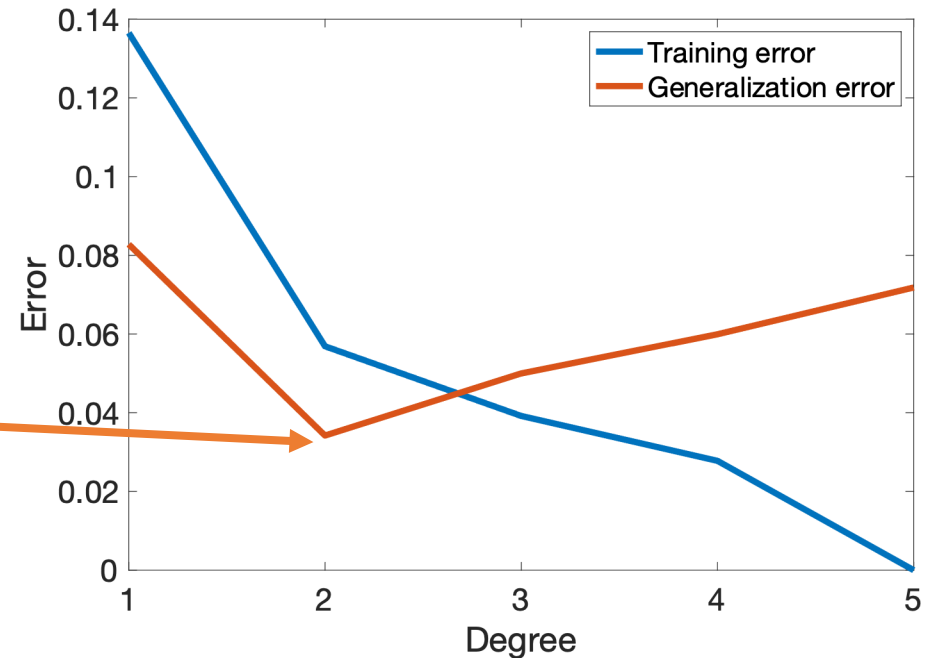
# Overfitting

This is how polynomials of different degrees would fit the data:



# Generalization error

- The best fit to the data is clearly not the model that is closest to the underlying true relationship!
- We want to find the model that **minimizes the generalization error**, not the error on the training data!
- Hence, we need to train our model on one part of the data and validate the model on another part that has not been used for training.

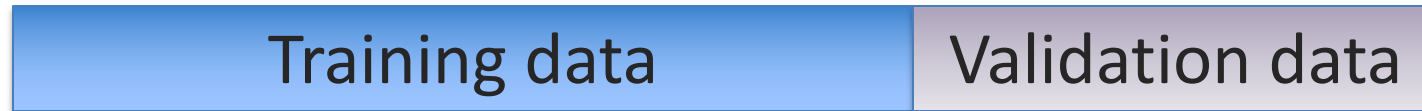


# Validation data

- Three ways to perform the validation:
  - Hold out
  - Cross validation
  - Leave one out

# Hold out

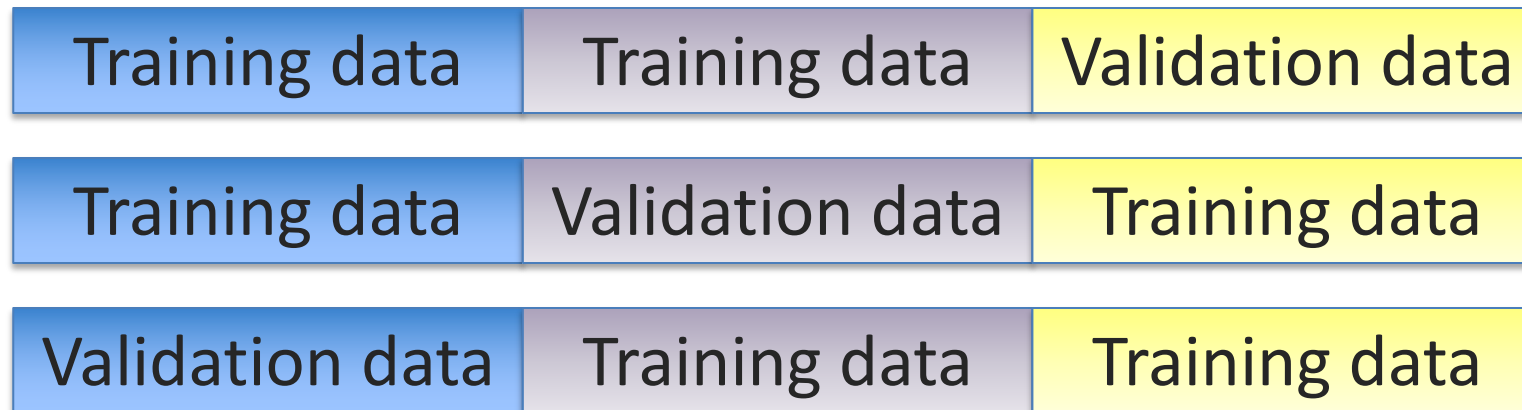
- Simplest approach, hold out one part of the entire data set as validation data.





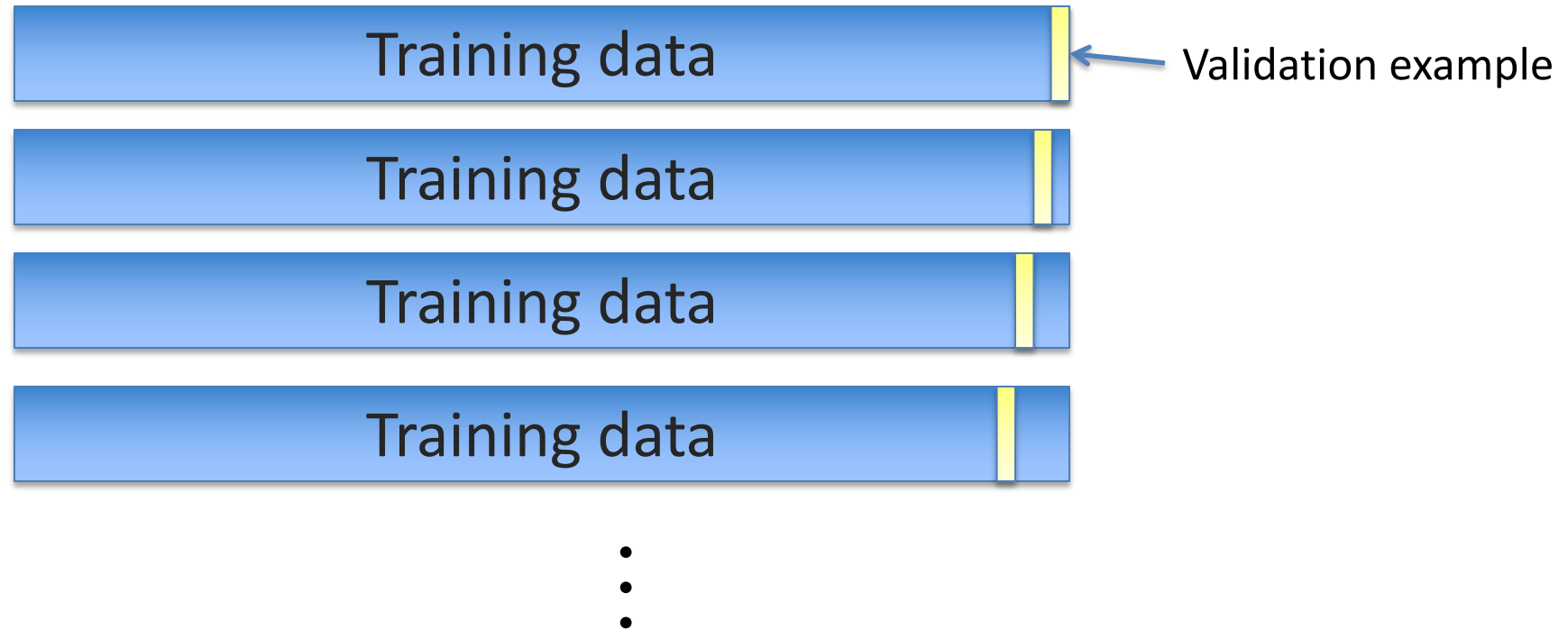
# n-fold Cross-Validation

- Divide data set into  $n$  segments. Train using  $n-1$  segments and validate using the  $n$ :th.
- Example of 3-fold Cross-Validation:



# Leave-one-out

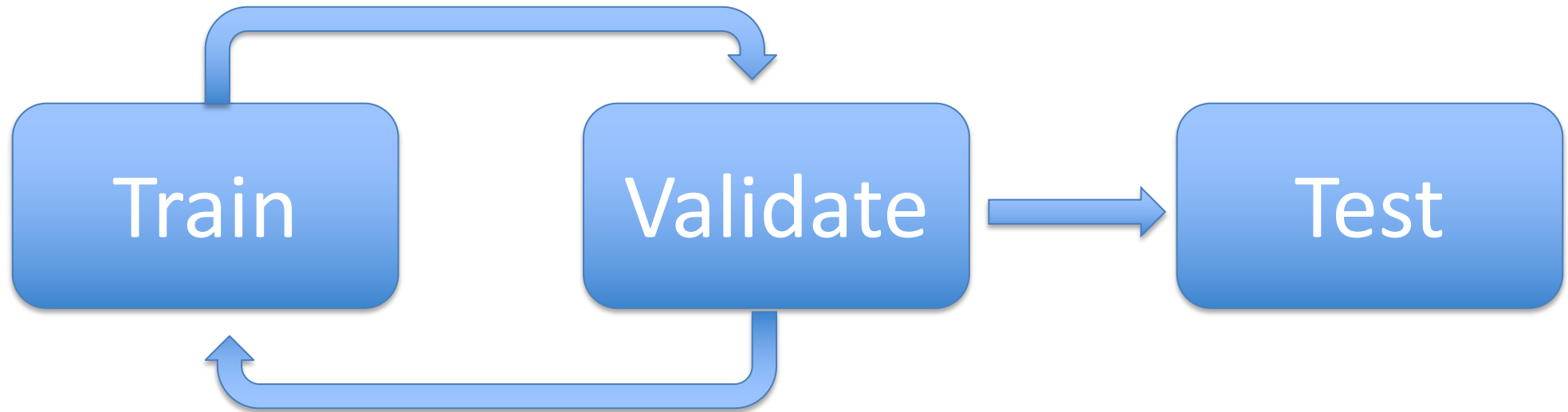
- Extreme case of Cross-Validation: Use all data but one example for training and use the last one to evaluate



# How can we find the minimum generalization error?

- What happens if the generalization error is not low enough?
- Modify the classifier (change the model) and train again...
- But – then the validation data is used to select the model!
- How do we know how well the new model generalizes?
- Need new validation data to test the final model – this dataset is called test data.
- Test data must never be used more than once!

# Training – Validation –Testing



# Evaluating classifiers – The Confusion matrix



Iris setosa



Iris versicolor



Iris virginica

		Predicted class		
		Setosa	Versicol.	Virginica
Actual class	Setosa	50	0	0
	Versicol.	0	45	5
	Virginica	0	7	43

Accuracy: 
$$\frac{50 + 45 + 43}{150} = 92\%$$