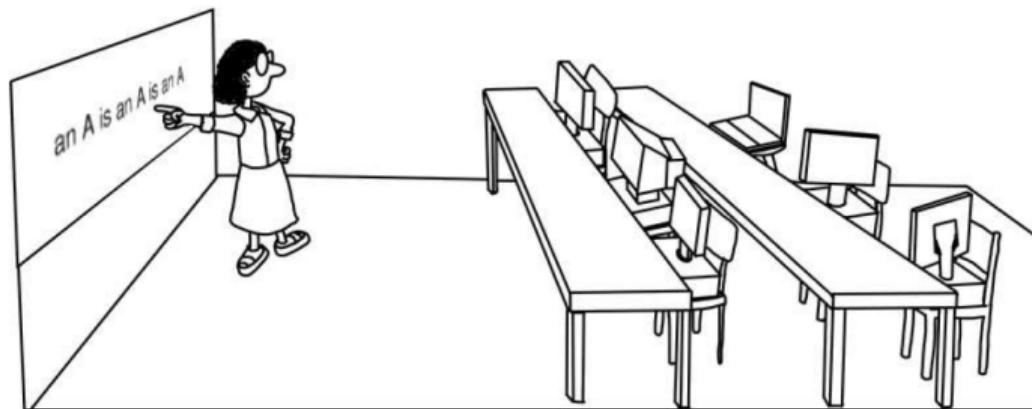


# Machine Learning

Krystian Mikolajczyk & Deniz Gunduz

Department of Electrical and Electronic Engineering  
Imperial College London



# Machine Learning - Part 1 Summary

Department of Electrical and Electronic Engineering

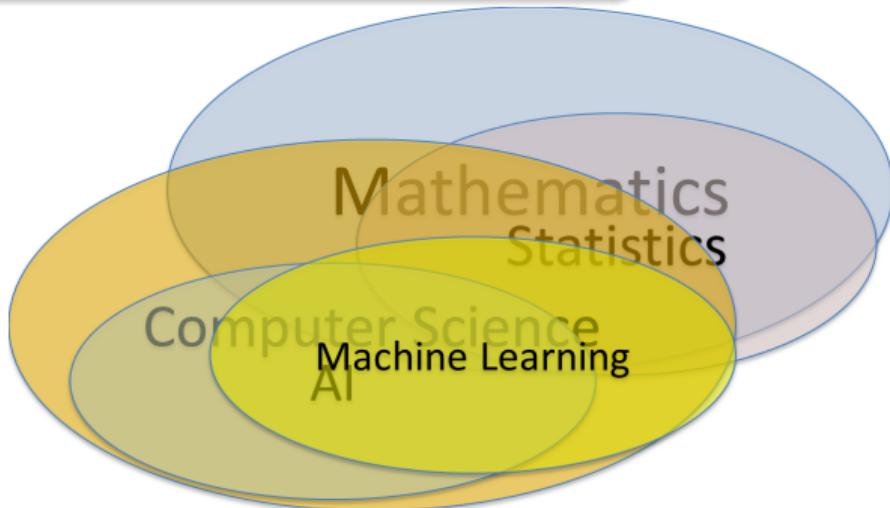
Imperial College London

- Components of learning
- Tasks, types of learning
- Types of data
- ML problem formulation
- Example applications

## What is Machine Learning?

- A pattern exists in a problem
- We cannot solve it analytically  
 $f(x) \neq ax + b$
- We have data on it
- Mathematics
  - Study of quantity, structure, space and change
- Statistics
  - Data analysis: from hypothesis to validation by data
- Computer Science
  - Theory, experimentation and engineering for design and use of computers

Extract a description (approximate)  
of the pattern!



- AI, Machine Intelligence
  - Intelligent agents with perception and actions to achieve goals
- Machine Learning
  - Ability to learn: from data to hypothesis

# Components of learning

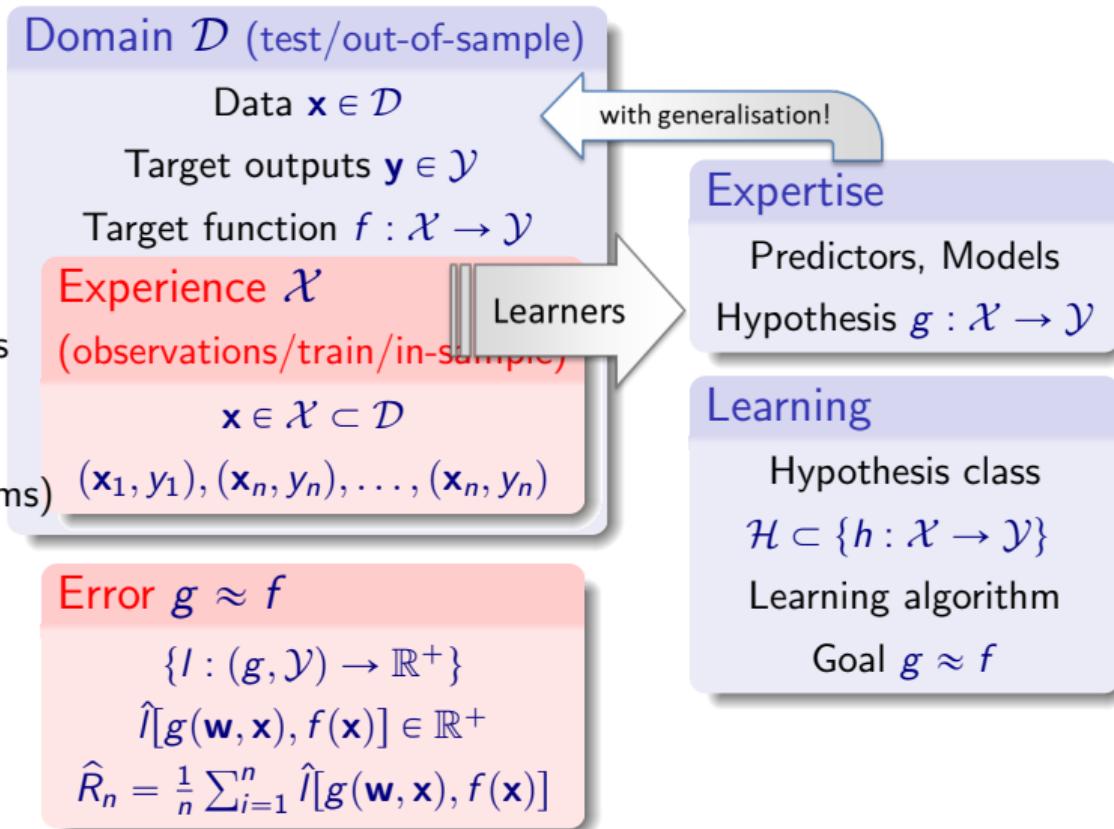
- Theory
  - ML problem formulation
  - ML types and tasks
  - Data representations
  - Metrics, error, loss and bounds
  - Empirical Risk Minimization

- Predictors and Learners (algorithms)

- Linear and non linear
- Neural Networks
- SVM

- Learning frameworks

- Supervised
- Unsupervised

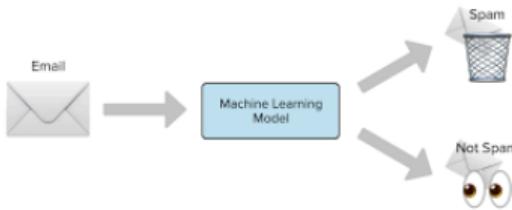


# Learning tasks

- Classification

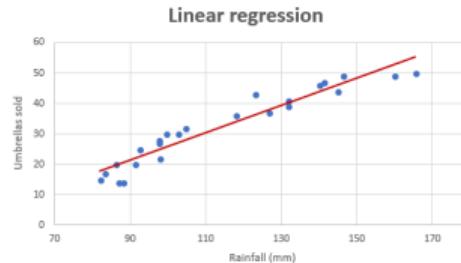
- ▶ Binary
- ▶ Multiclass
- ▶ Multilabel

C = 3	Multi-Class			Multi-Label		
	Samples	Samples	Samples	Labels (t)	Labels (t)	Labels (t)
				[0 0 1]	[1 0 0]	[0 1 0]
				[1 0 1]	[0 1 0]	[1 1 1]



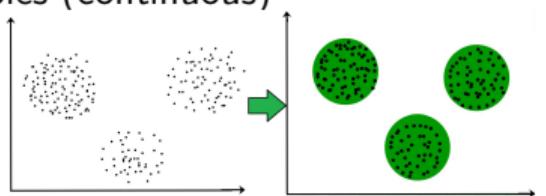
- Regression

- ▶ Univariate (continuous)
- ▶ Multivariate, group of variables (continuous)

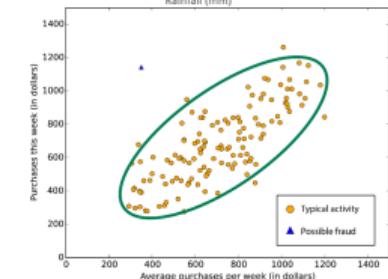


- Clustering

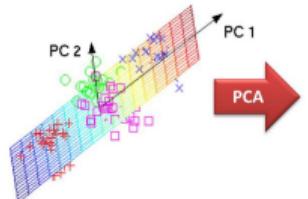
- ▶ Hierarchical, Flat



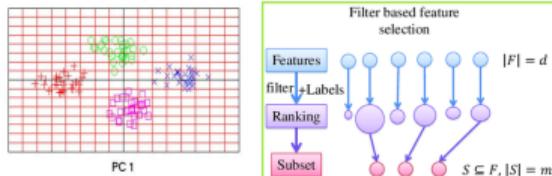
- Anomaly detection



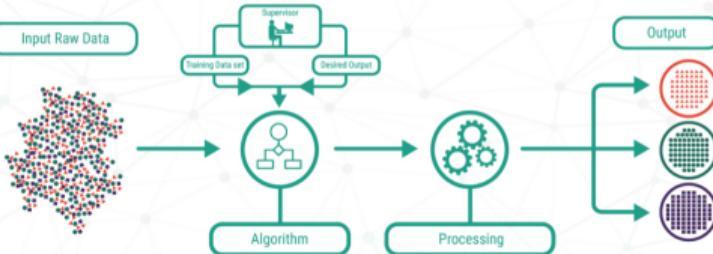
- Dimensionality reduction



- Feature selection



# Types of learning



## ● Supervised learning

- ▶ Learns explicitly, data with clearly defined input–output pairs
- ▶ Predicts outcome, future
- ▶ Direct feedback is given, prediction error
- ▶ Classification, regression, ranking

## ● Unsupervised learning

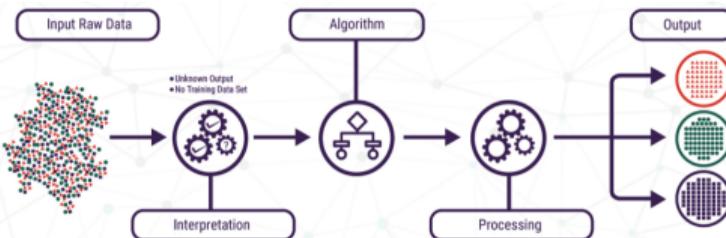
- ▶ Attempts to understand the data, find patterns/structures
- ▶ Only input is given, no specific predictions made
- ▶ Evaluation is qualitative, indirect
- ▶ Clustering, association mining, dimensionality reduction

## ● Reinforcement learning

- ▶ Learns how to act in a given environment to maximize rewards
- ▶ Input, some output, and some valuation is given
- ▶ Decision process, reward or recommendation systems

## ● Weakly-, semi-supervised learning

- ▶ Output given for part of data, or partially/auxiliary/intermediate outputs are available



## Types of learning

### Passive vs. active learners

- Data is given vs. learner actively decides which data to use during training

### Oblivious vs. adversarial teacher

- Randomly sampled data vs. selected hard samples

### Batch vs. online learning

- All training data available vs. training and testing at the same time

# Applications

## Supervised

- Classification
  - Spam detection
  - Diagnostics
  - Fraud detection
  - Image/audio/text classification
- Regression
  - Risk assessment
  - Score prediction
  - Energy consumption
- Ranking
  - Search engines
  - Information retrieval
  - Machine translation

## Unsupervised

- Dimensionality reduction
  - Big data analytics
- Clustering
  - Biology
  - City planning
  - Marketing

## Reinforcement

- Gaming
- Finance
- Manufacturing
- Inventory management
- Robot navigation

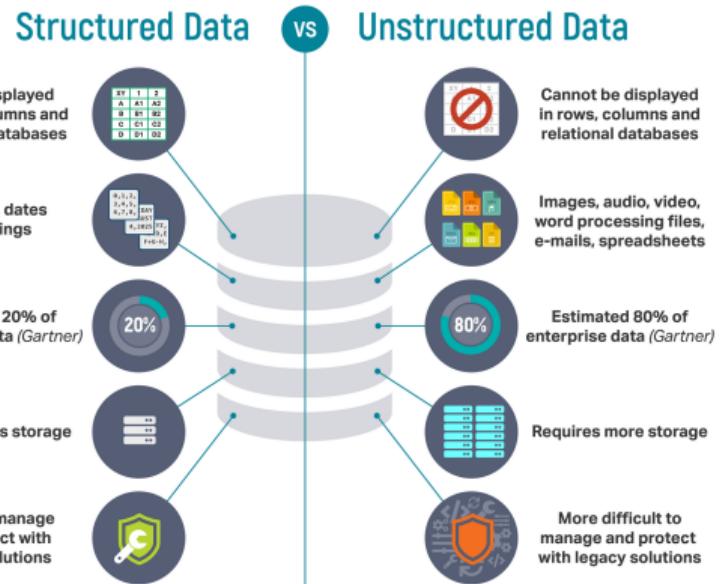
# Types of data

- Structured

- ▶ Organized into specific types
- ▶ Can be used as direct inputs into ML algorithms

- Unstructured

- ▶ Raw data
- ▶ Different types and modalities of data are mixed
- ▶ Needs preprocessing for ML algorithms



# ML problem formulation

- Input:  $\mathbf{x} \in \mathcal{X} \subset \mathcal{D}$ 
  - text, image, audio clip, sensor output
- Output  $\mathbf{y} \in \mathcal{Y}$ 
  - label, output value
- Target function  $f : \mathcal{X} \rightarrow \mathcal{Y}$ 
  - Actual true function - never known in practice
- Data:  $(\mathbf{x}_1, y_1), (\mathbf{x}_n, y_n), \dots, (\mathbf{x}_n, y_n)$ 
  - available for training
- Error function  $\hat{R}_n = \frac{1}{n} \sum_n \hat{l}(g(\mathbf{w}, \mathbf{x}), y)$ 
  - Defining  $\hat{l}$  is crucial in supervised learning



Hypothesis:  $g : \mathcal{X} \rightarrow \mathcal{Y}$ .

learnt model, one of many possible

## Learning

- Hypothesis class:  $\mathcal{H} \subset \{h : \mathcal{X} \rightarrow \mathcal{Y}\}$
- Learning algorithm
- Goal:  $g \approx f$

# Applications

- Spam Filtering

- Goal: Identify spam emails.

The screenshot shows an email client interface with the following details:

- To:** ee\_ee3-23
- Cc:** information
- Subject:** Dear All,
- Content:** The content of the lectures will be available on Blackboard.  
Andras
- Signature:** András György  
Senior Lecturer
- Contact Information:** Department of Electrical and Electronic Engineering  
Imperial College London  
South Kensington Campus, London SW7 2AZ, UK  
Phone: +44 (0)20 7594 6173  
Web: <http://www.imperial.ac.uk/people/a.gyorgy>  
Email: a.gyorgy@imperial.ac.uk

An email is represented by the vector of word counts (bag of words model):

information	dear	all	content	blackboard	Friday	business	proposal	late	husband	...
1	1	1	1	1	0	0	0	0	0	...

Determine if an email is spam based on its word-count vector!

# Applications

- Character recognition
- Face recognition
- Object recognition

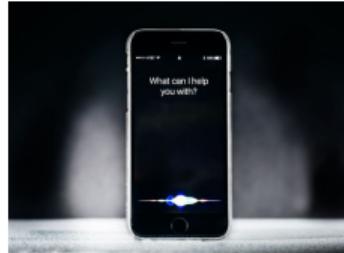
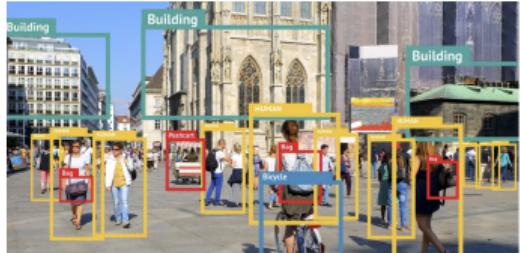
0 0 0 0 0 0 0 0 0 0 0 0 0  
1 1 1 1 1 1 1 1 1 1 1 1 1  
2 2 2 2 2 2 2 2 2 2 2 2 2  
3 3 3 3 3 3 3 3 3 3 3 3 3  
4 4 4 4 4 4 4 4 4 4 4 4 4  
5 5 5 5 5 5 5 5 5 5 5 5 5  
6 6 6 6 6 6 6 6 6 6 6 6 6  
7 7 7 7 7 7 7 7 7 7 7 7 7  
8 8 8 8 8 8 8 8 8 8 8 8 8  
9 9 9 9 9 9 9 9 9 9 9 9 9

MNIST



Baba Dash ([mathworks.com](http://mathworks.com))

- Intelligent assistants
  - Alexa, Google Now, Siri, Cortana



# Applications

- Ad placement

- ▶ Google, Bing, Facebook,  
Yahoo

- Recommendations

- ▶ Amazon, Ebay, Netflix

- Assembly of web pages

- ▶ Yahoo, CNN

Google barbados

All Maps Images News Videos More Settings Tools

About 93,200,000 results (0.57 seconds)

**Barbados Holidays - TUI - Discover Your Smile - TUI.co.uk**  
<http://www.tui.co.uk/barbados/Holidays> • Choose From A Wide Range Of Holidays In Barbados. ABTA/ATOL Prot...

**Barbados - Wikipedia**  
<https://en.wikipedia.org/wiki/Barbados> • Barbados is an island country in the Lesser Antilles, in the Caribbean region of North America. It is 34 kilometres (21 miles) in length and up to 23 km (14 mi) in width, covering an area of 432 km<sup>2</sup> (167 sq mi). It is situated in the western area of the North Atlantic and 100 km (62 mi) east of the Windward Islands and the ... History of Barbados · Indians In Barbados · Flag of Barbados · Economy of Barbados

**Top stories**

Rihanna wipes away tears at murdered cousin's funeral in Barbados as she l... Canada issues travel advisory against Barbados due to sewage mess Joe McElory vigorously works out in shorts in Barbados

Metro 13 hours ago St. Lucia Times Daily Mail

1 day ago 9 hours ago

→ More for barbados

**Barbados (@Barbados) - Twitter**  
<https://twitter.com/Barbados>

Combine your rum with the sun, sea and sand and we promise you couldn't find anything more relaxing! #BirthplaceOfRum #LoveBarbados : wonderleighs on IG pic.twitter.com/lT7pfIZV N...

Highlighted by its London brick buildings, discover the rich heritage of the Garrison Historic Area. A rare English military memorabilia including extremely unique canons, maps, bottles and even blueprints! #UNESCOWorldHerit age #LoveBarbados : theresaathomson on IG pic.twitter.com/DqkukZIE...

Freights Bay, tucked away in Christ Church, is a great spot for surfing, picnics and sunsets! #LoveBarbados : #VisitBarbados : theresathomson on IG pic.twitter.com/Qe8J P...

**Barbados**  
Country in the Caribbean

Barbados is an eastern Caribbean island and an independent British Commonwealth nation. Bridgetown, the capital, is a cruise-ship port with colonial buildings and Nidhe Israel, a synagogue founded in 1654. Around the island are beaches, botanical gardens, the Harrison's Cave formation, and 17th-century plantation houses like St. Nicholas Abbey. Local traditions include afternoon tea and cricket, the national sport.

**Capital:** Bridgetown  
**Currency:** Barbadian dollar  
**Recognised regional languages:** Bajan Creole  
**Capital and largest city:** Bridgetown; 13°06'N 59°37'W / 13.100°N 59.617°W

**Plan a trip**

Barbados travel guide 3-star hotel averaging £132, 5-star averaging £98 8 h 45 min flight, from £558 Destinations: Bridgetown, Oistins, Holetown, Bathsheba, Barbados, MORE

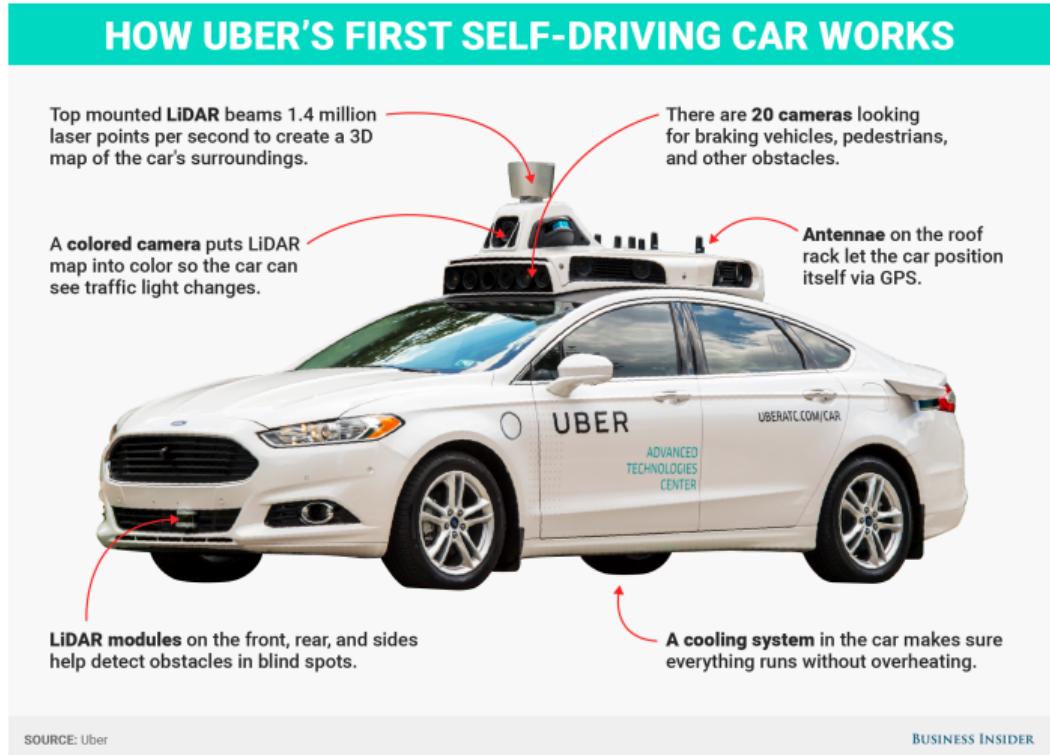
People also search for View 15+ more

Bah... Carib... Saint Lucia Belize Antigua and Barb...

Feedback

# Applications

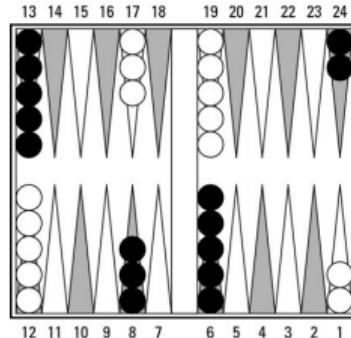
- Self-driving car
  - ▶ Google, Uber, several car manufacturers



## Applications

Computer Games:

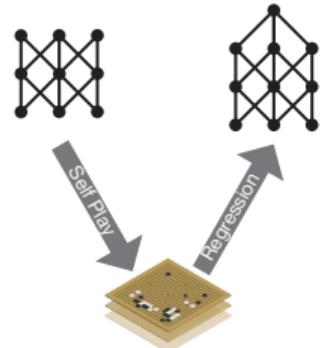
- Backgammon (1992)



- Poker (Texas hold'em) (2015, 2017)



- Go (2016, 2017)



Silver et al., Nature 2016

## ML datasets – example

- UCI Machine Learning Repository
  - <https://archive.ics.uci.edu/ml>
  - Wine Quality Data Set
  - Human Activity Recognition Using Smartphones

## Wine Quality Dataset – example

- Two datasets were created, using red and white wine samples
- The inputs include objective tests (e.g. PH values)
- The output is based on sensory data
  - Median of at least 3 evaluations made by wine experts
  - Each expert graded the wine quality between 0 (very bad) and 10 (very excellent)

Input variables (based on physicochemical tests):

1 - fixed acidity  
2 - volatile acidity  
3 - citric acid  
4 - residual sugar  
5 - chlorides  
6 - free sulfur dioxide  
7 - total sulfur dioxide  
8 - density  
9 - pH  
10 - sulphates  
11 - alcohol

Output variable (based on sensory data):

12 - quality (score between 0 and 10)

# Wine Quality Dataset – example

Data Set Characteristics:	Multivariate	Number of Instances:	4898
Attribute Characteristics:	Real	Number of Attributes:	12
Associated Tasks:	Classification, Regression	Missing Values?	N/A

- Domain
    - red and white variants of wine
    - $\mathcal{D} \subset \{\mathbf{x} : \mathbf{x} \in \mathbb{R}^{11}\}$
  - Data samples
    - physicochemical (inputs) and sensory (the output)
    - $\mathcal{X} \subset \mathcal{D}, \mathbf{x}_n \in \mathbb{R}^{11}, n = 1, \dots, N = 4898$
  - Target function
    - Assign quality score, unbalanced
    - Regression
    - $f : \mathcal{X} \rightarrow \mathcal{Y}, f(\mathbf{x}) = y, \mathbf{x} \in \mathcal{X}, y \in \mathbb{R}, y \in [1, 10]$
  - Hypothesis class  $\mathcal{H} \subset \{h : \mathcal{X} \rightarrow \mathcal{Y}\}$
  - Error function  $\hat{R} = \frac{1}{n} \sum_i \dots$
- Input variables (based on physicochemical tests):  
 1 - fixed acidity  
 2 - volatile acidity  
 3 - citric acid  
 4 - residual sugar  
 5 - chlorides  
 6 - free sulfur dioxide  
 7 - total sulfur dioxide  
 8 - density  
 9 - pH  
 10 - sulphates  
 11 - alcohol  
 12 - quality (score between 1 and 10)
- Output variable (based on sensory data):
- Learning algorithm
- Other possible tasks
- Binary classification (supervised)
- Multiclass classification (supervised)
- Feature selection
- Anomaly detection (unsupervised)
- Clustering (unsupervised)

## Human Activity Recognition Using Smartphones – example

- Recorded from 30 volunteers within an age bracket of 19-48 years wearing a smartphone (Samsung Galaxy S II) on the waist
  - Each person performed six activities (WALKING, WALKING UPSTAIRS, WALKING DOWNSTAIRS, SITTING, STANDING, LAYING) .
  - Accelerometer and gyroscope captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz
  - The experiments have been video-recorded to label the data manually
  - The obtained dataset has been randomly partitioned into two sets, where 70% of the subjects was selected for the training data and 30% for the test data
  - The signals were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). From each window, a vector of features was obtained by calculating variables from the time and frequency domain
  - The acceleration was separated by a Butterworth low-pass filter into body acceleration and gravity
  - The gravitational force is assumed to have only low frequency components, therefore a filter with 0.3 Hz cutoff frequency was used

# Human Activity Recognition Using Smartphones – example

Data Set Characteristics:	Multivariate, Time-Series	Number of Instances:	10299
Attribute Characteristics:	N/A	Number of Attributes:	561
Associated Tasks:	Classification, Clustering	Missing Values?	N/A

- Domain
  - Human activity represented by
  - $\mathcal{D} \subset \{\mathbf{x} : \mathbf{x} \in \mathbb{R}^{561}\}$
- Data samples
  - sensory (the output) variables
  - $\mathcal{X} \subset \mathcal{D}, \mathbf{x}_n \in \mathbb{R}^{561}, n = 1, \dots, N = 10299$
- Target function
  - Assign activity label
  - Multiclass classification (supervised)
  - $f : \mathcal{X} \rightarrow \mathcal{Y}, f(\mathbf{x}) = y, \mathbf{x} \in \mathcal{X}, y \in \mathbb{N}, y \in [1, 6], y \in \{1, \dots, 6\}$
- Hypothesis class  $\mathcal{H} \subset \{h : \mathcal{X} \rightarrow \mathcal{Y}\}$
- Error function  $\hat{R} = \frac{1}{n} \sum_i \dots$
- Learning algorithm
- Other possible tasks
  - Binary classification (supervised)
  - Feature selection
  - Clustering (unsupervised)

## Part 1 summary

- Components of learning
- Tasks, types of learning
- Types of data
- ML problem formulation
- Example applications