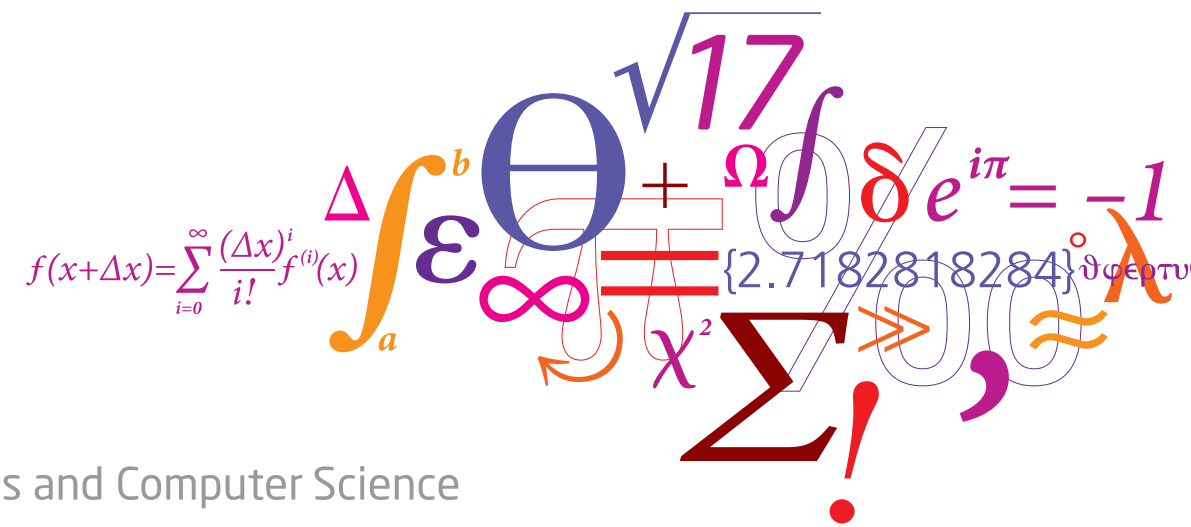
02450: Introduction to Machine Learning and Data Mining

Recap

Morten Mørup

DTU Compute, Technical University of Denmark (DTU)



DTU Compute
Department of Applied Mathematics and Computer Science

Lecture Schedule

1. Introduction
   7 October: C1

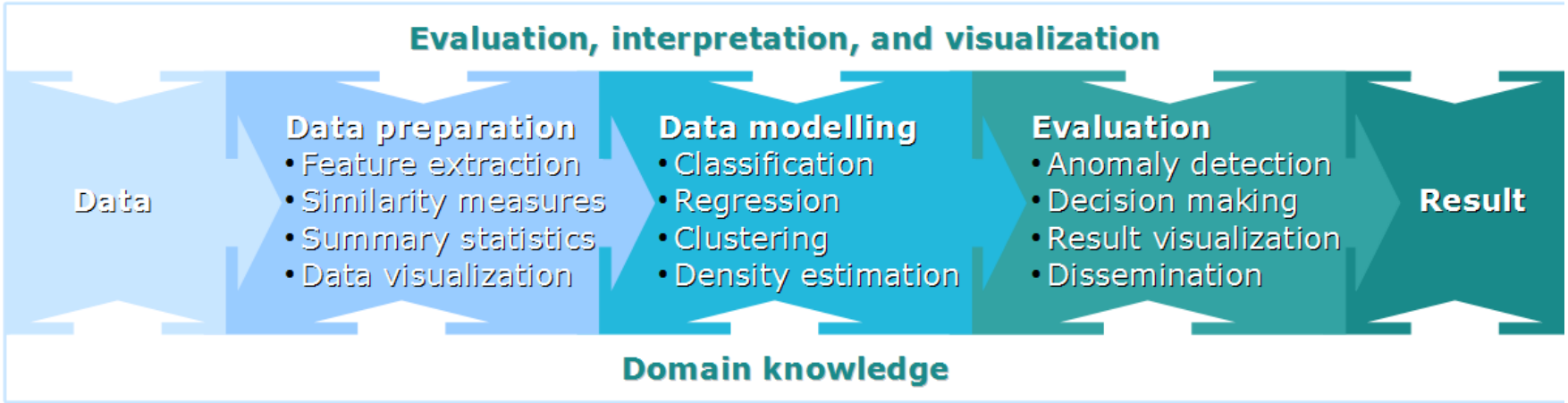
   Data: Feature extraction, and visualization
2. Data, feature extraction and PCA
   7 October: C2, C3
3. Measures of similarity, summary statistics and probabilities
   7 October: C4, C5
4. Probability densities and data Visualization
   7 October: C6, C7

   Supervised learning: Classification and regression
5. Decision trees and linear regression
   8 October: C8, C9
6. Overfitting, cross-validation and Nearest Neighbor
   8 October: C10, C12
7. Performance evaluation, Bayes, and Naive Bayes
   9 October: C11, C13
8. Artificial Neural Networks and Bias/Variance
   9 October: C14, C15
9. AUC and ensemble methods
   10 October: C16, C17

   Unsupervised learning: Clustering and density estimation
10. K-means and hierarchical clustering
    10 October: C18
11. Mixture models and density estimation
    11 October: C19, C20
12. Association mining
    11 October: C21

    Recap
13. **Recap**
    11 October: C1-C21

Piazza online help: `https://piazza.com/dtu.dk/fall2019/october2019`



Learning Objectives

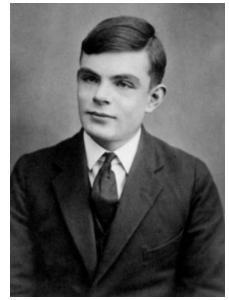
- Remember key aspects taught in the course

What is Machine Learning (ML)

Alan Turing (1946)

We are not in a position to answer if a machine can think because the terms machine and think are undefined. Rather we should ask if a machine can imitate a human (the Turing test).

Proposed we should consider machines that were able to learn like children.



Alan Turing
(1912-1954)

Arthur Samuel (1959)

***Machine learning:* "Field of study that gives computers the ability to learn without being explicitly programmed"**

Samuels wrote a checkers playing program,
had the program play 10000 games against itself and
work out which board positions were good and bad
depending on wins/losses.

https://commons.wikimedia.org/wiki/File:This_is_the_photo_of_Arthur_Samuel.jpg

Arthur Samuel
(1901-1990)

Tom Mitchell (1999) (http://www.cs.cmu.edu/~tom/)

***Well posed learning problem:* "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."**

For checkers we have:

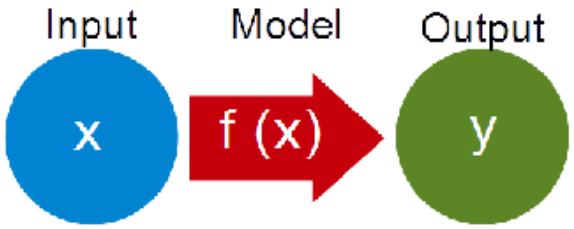
E = 10000 games

T = Playing checkers

P = If you win or not

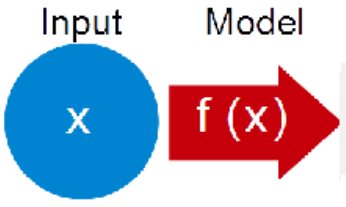
What can ML be used for

Supervised learning



y: different classes ⇒ classification problem
y: continuous values ⇒ regression problem

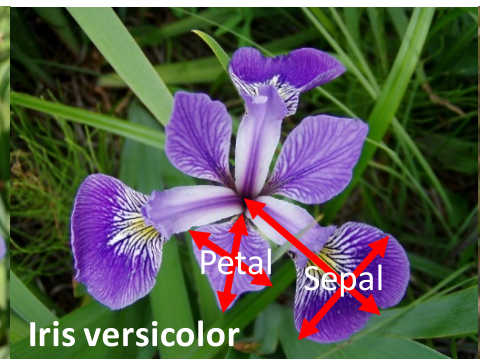
Unsupervised learning



Find patterns in data (explorative analysis)

Density estimation
Association mining
Clustering

Example: Fisher's Iris data

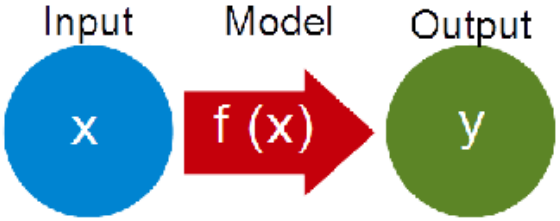


https://en.wikipedia.org/wiki/Iris_flower_data_set#/media/File:Kosaciec_szczecinkowaty_Iris_setosa.jpg
https://en.wikipedia.org/wiki/Iris_flower_data_set#/media/File:Iris_versicolor_3.jpg
https://en.wikipedia.org/wiki/Iris_flower_data_set#/media/File:Iris_virginica.jpg

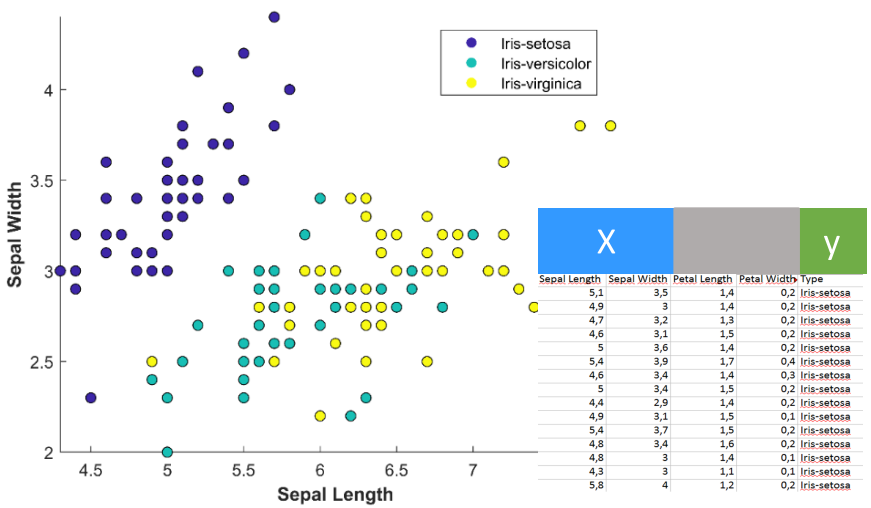
Ronald Fisher
(1890 - 1962)

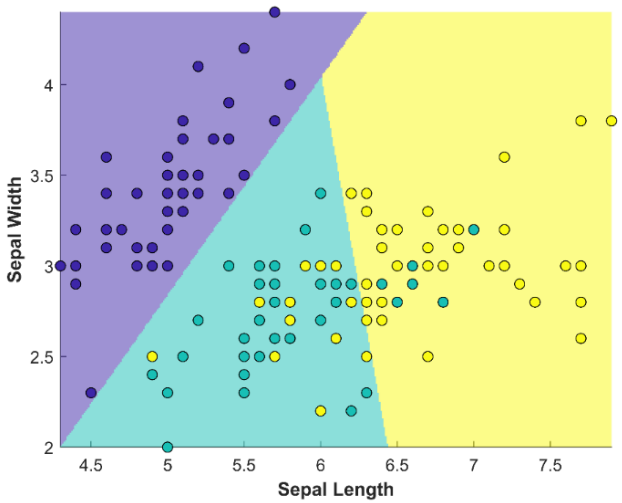
A	B	C	D	E
Sepal Length	Sepal Width	Petal Length	Petal Width	Type
5,1	3,5	1,4	0,2	Iris-setosa
4,9	3	1,4	0,2	Iris-setosa
4,7	3,2	1,3	0,2	Iris-setosa
4,6	3,1	1,5	0,2	Iris-setosa
5	3,6	1,4	0,2	Iris-setosa
5,4	3,9	1,7	0,4	Iris-setosa
4,6	3,4	1,4	0,3	Iris-setosa
5	3,4	1,5	0,2	Iris-setosa
4,4	2,9	1,4	0,2	Iris-setosa
4,9	3,1	1,5	0,1	Iris-setosa
5,4	3,7	1,5	0,2	Iris-setosa
4,8	3,4	1,6	0,2	Iris-setosa
4,8	3	1,4	0,1	Iris-setosa
4,3	3	1,1	0,1	Iris-setosa
5,8	4	1,2	0,2	Iris-setosa

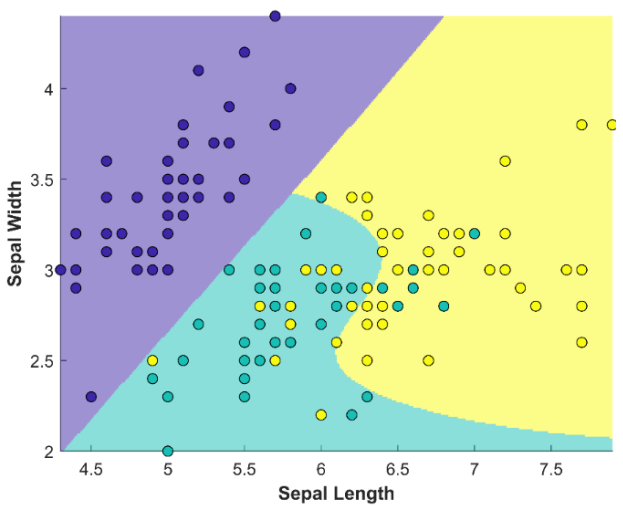
Supervised learning

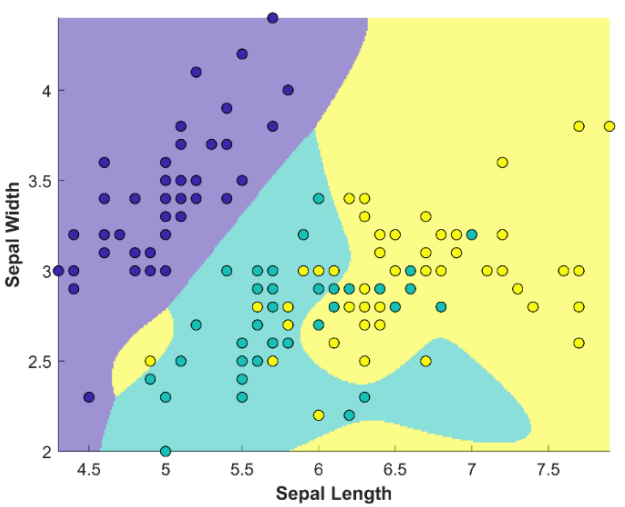


y: different classes ⇒ classification problem

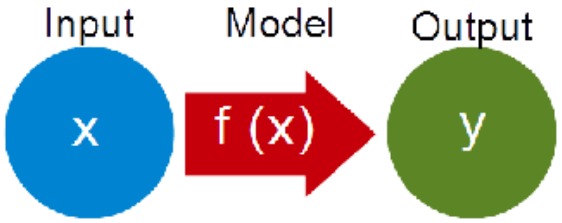




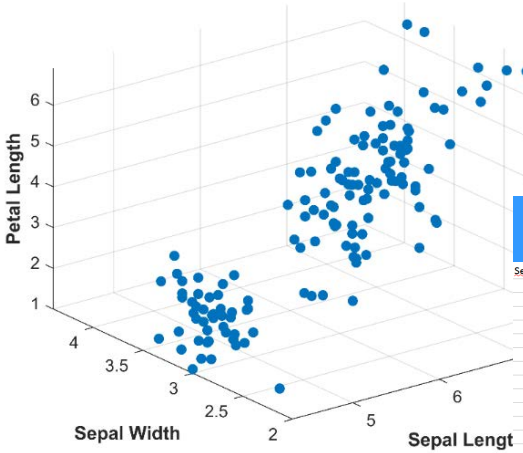




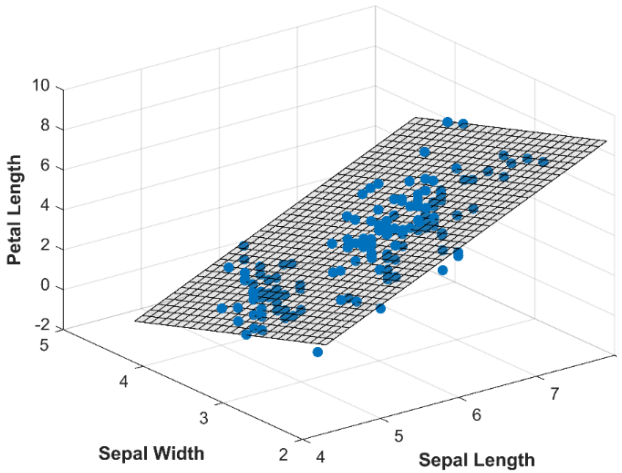
Supervised learning

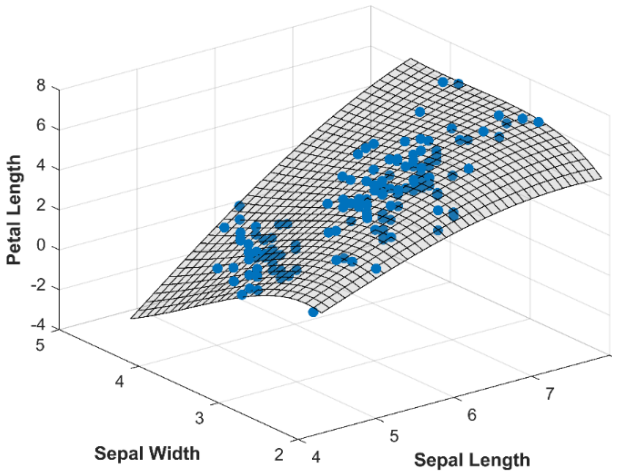


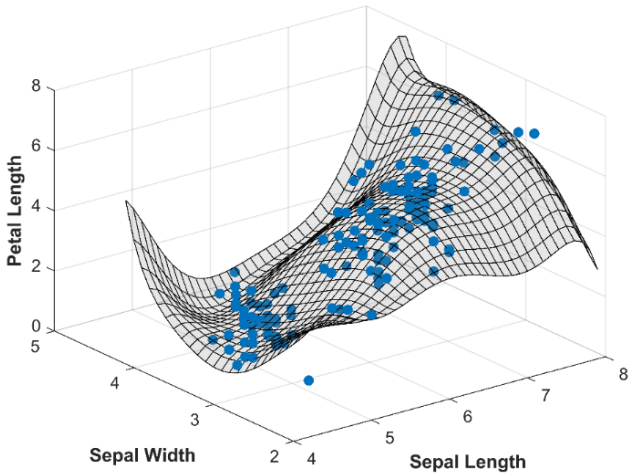
y: continuous values ⇒ regression problem



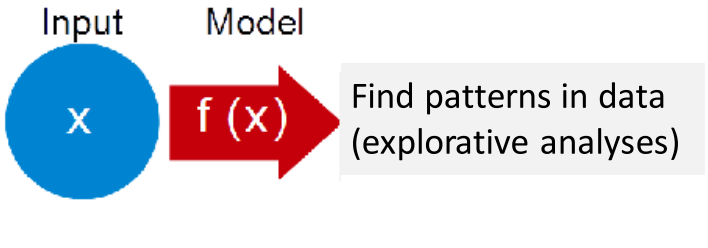
x			y
Sepal Length	Sepal Width	Petal Length	Petal Width	Type
5,1	3,5	1,4	0,2	Iris-setosa
4,9	3	1,4	0,2	Iris-setosa
4,7	3,2	1,3	0,2	Iris-setosa
4,6	3,1	1,5	0,2	Iris-setosa
5	3,6	1,4	0,2	Iris-setosa
5,4	3,9	1,7	0,4	Iris-setosa
4,6	3,4	1,4	0,3	Iris-setosa
5	3,4	1,5	0,2	Iris-setosa
4,4	2,9	1,4	0,2	Iris-setosa
4,9	3,1	1,5	0,1	Iris-setosa
5,4	3,7	1,5	0,2	Iris-setosa
4,8	3,4	1,6	0,2	Iris-setosa
4,8	3	1,4	0,1	Iris-setosa
4,3	3	1,1	0,1	Iris-setosa
5,8	4	1,2	0,2	Iris-setosa

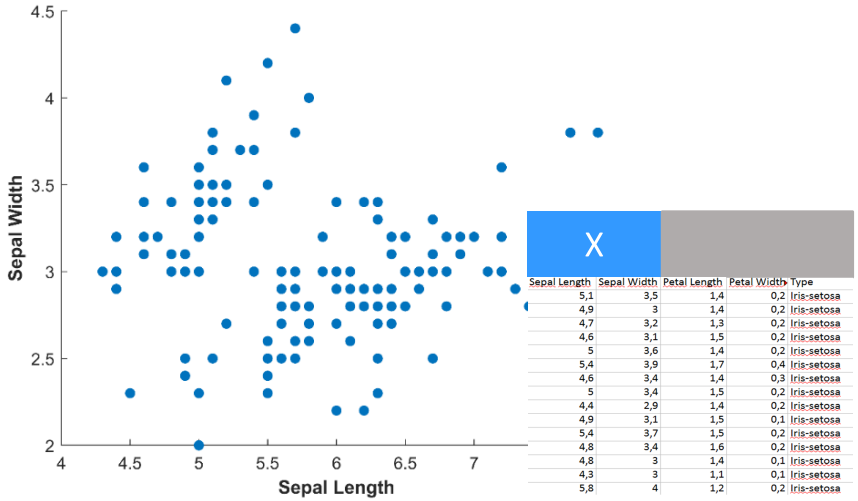


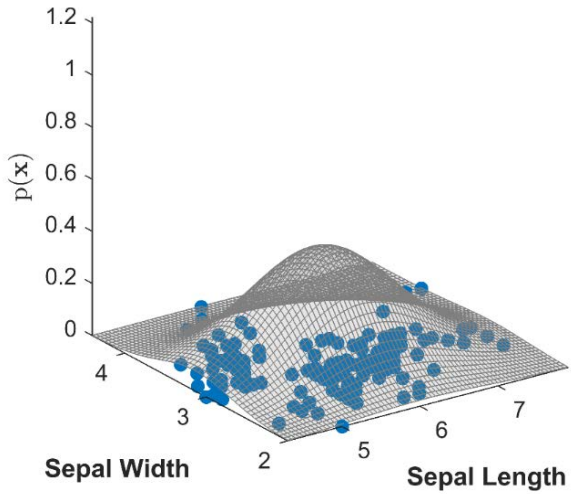


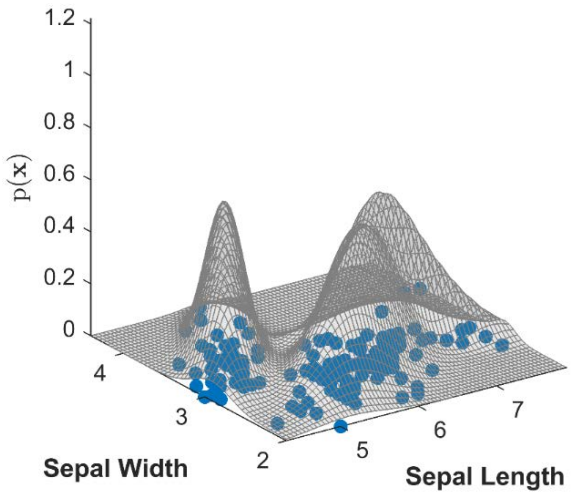


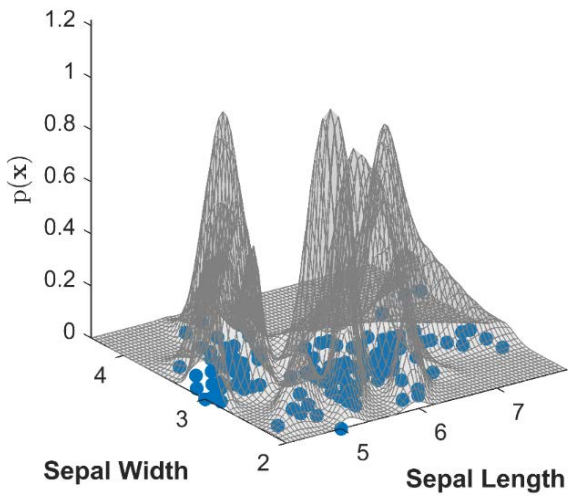
Unsupervised learning





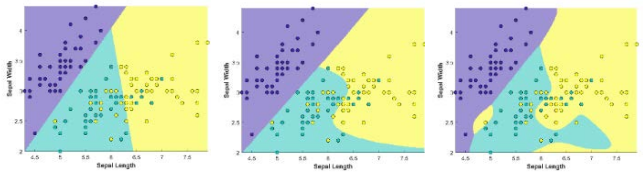




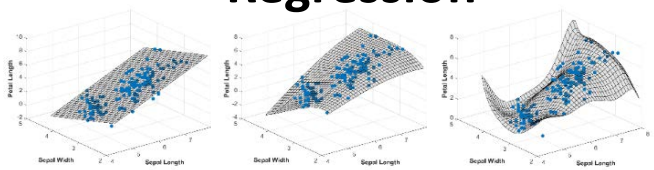


Control of model complexity

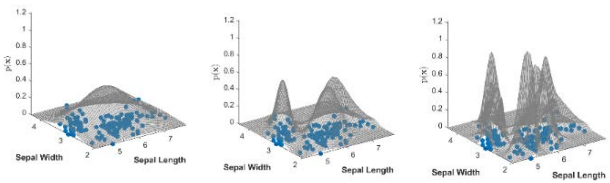
Classification



Regression

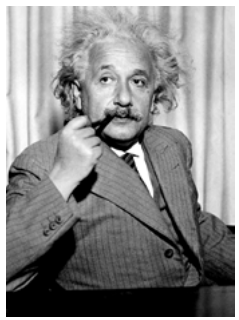


Density estimation



Lex Parsimoniae,
Law of parsimony
Given two models with same predictive performance, the simpler model is preferred over the more complex model (paraphrased)

William of Ockham
(1288-1347)



"Everything should be made as simple as possible, but not simpler"

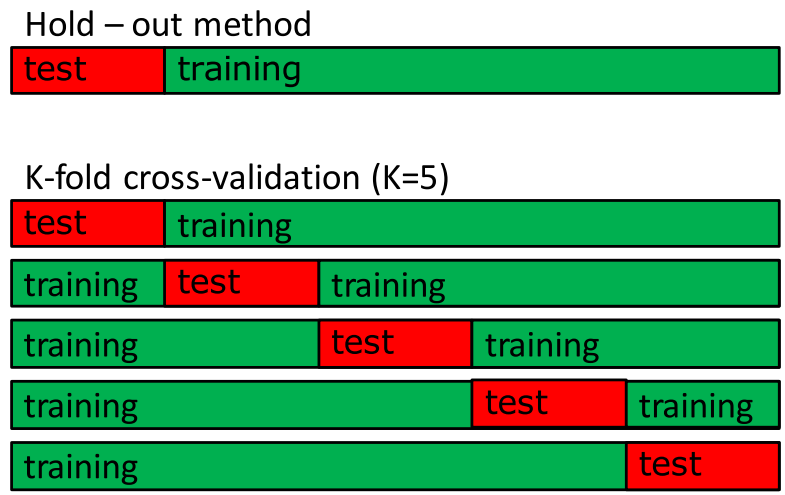
Albert Einstein
(1879 - 1955)

Machine learning in one sentence:

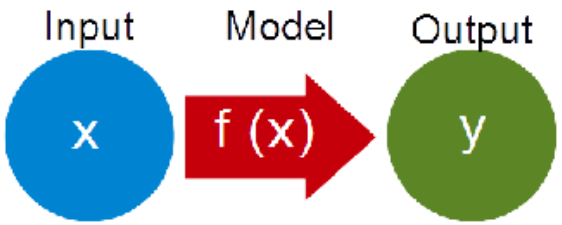
The aim of machine learning is to minimize the generalization error.

Generalisation error:
The extend to which a machine learning method on average will fail when evaluated on an infinite amount of (test) data.

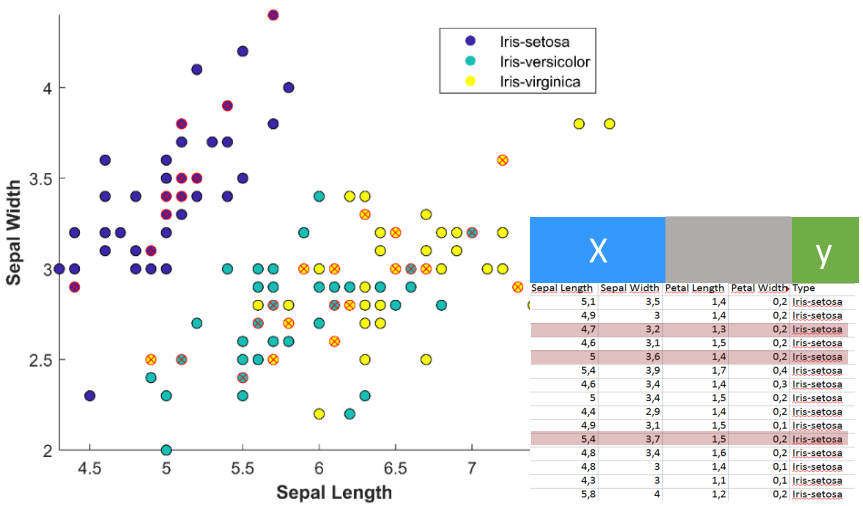
Cross-validation:
A framework for quantifying the generalization error from the available data.

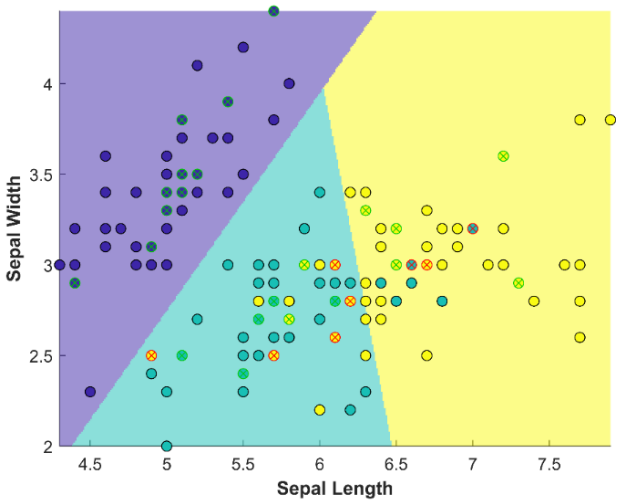


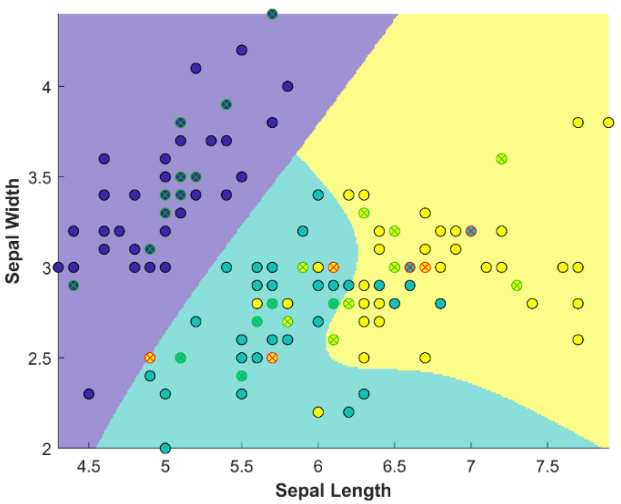
Supervised learning

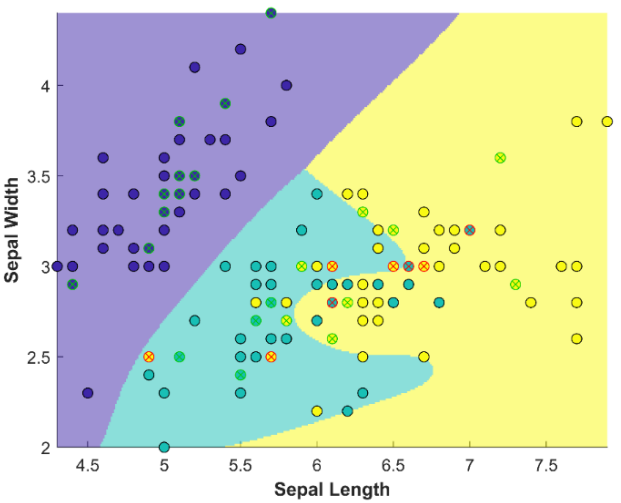


y: different classes ⇒ classification problem

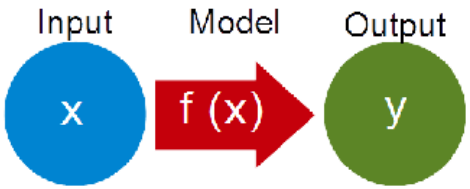




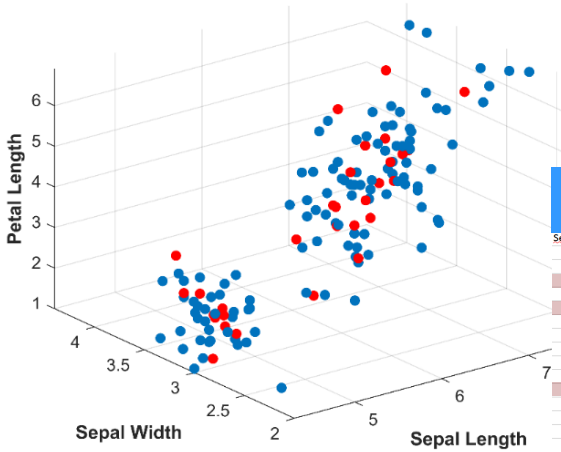




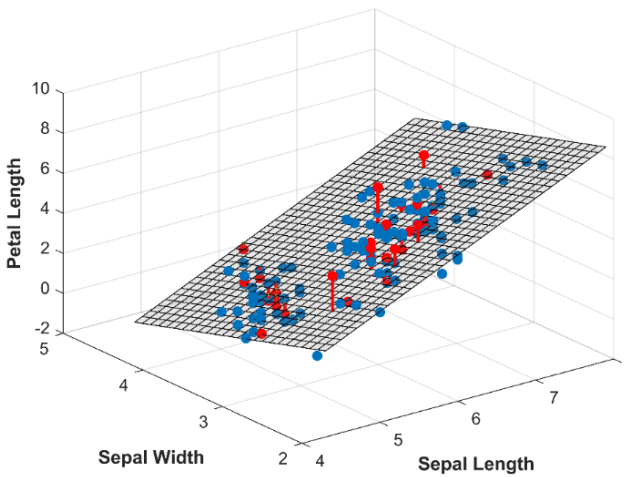
Supervised learning

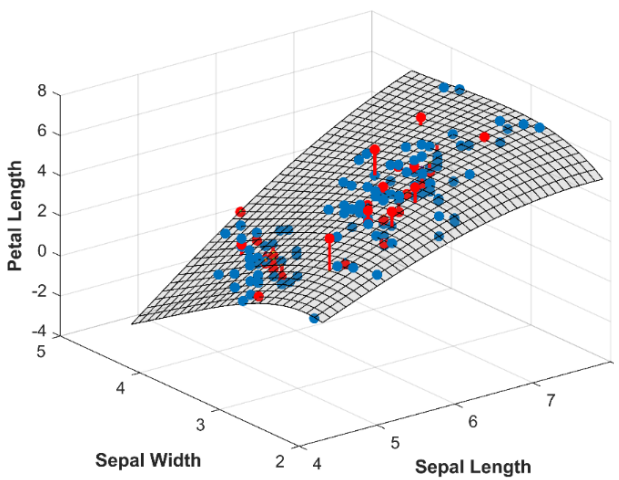


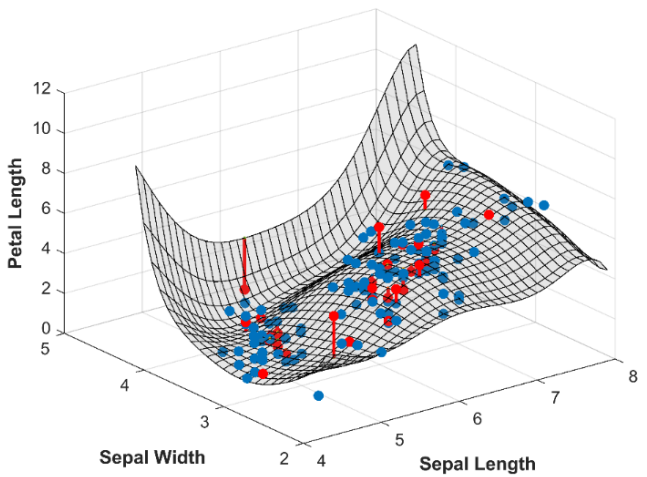
y: continuous values ⇒ regression problem



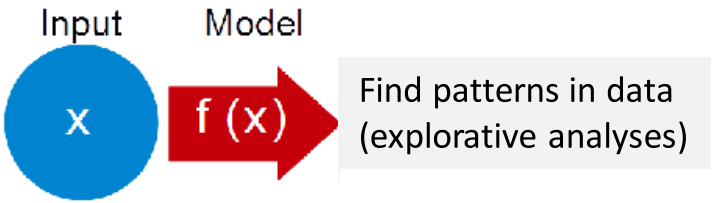
x				y
Sepal Length	Sepal Width	Petal Length	Petal Width	Type
5,1	3,5	1,4	0,2	Iris-setosa
4,9	3	1,4	0,2	Iris-setosa
4,7	3,2	1,3	0,2	Iris-setosa
4,6	3,1	1,5	0,2	Iris-setosa
5	3,6	1,4	0,2	Iris-setosa
5,4	3,9	1,7	0,4	Iris-setosa
4,6	3,4	1,4	0,3	Iris-setosa
5	3,4	1,5	0,2	Iris-setosa
4,4	2,9	1,4	0,2	Iris-setosa
4,9	3,1	1,5	0,1	Iris-setosa
5,4	3,7	1,5	0,2	Iris-setosa
4,8	3,4	1,6	0,2	Iris-setosa
4,8	3	1,4	0,1	Iris-setosa
4,3	3	1,1	0,1	Iris-setosa
5,8	4	1,2	0,2	Iris-setosa



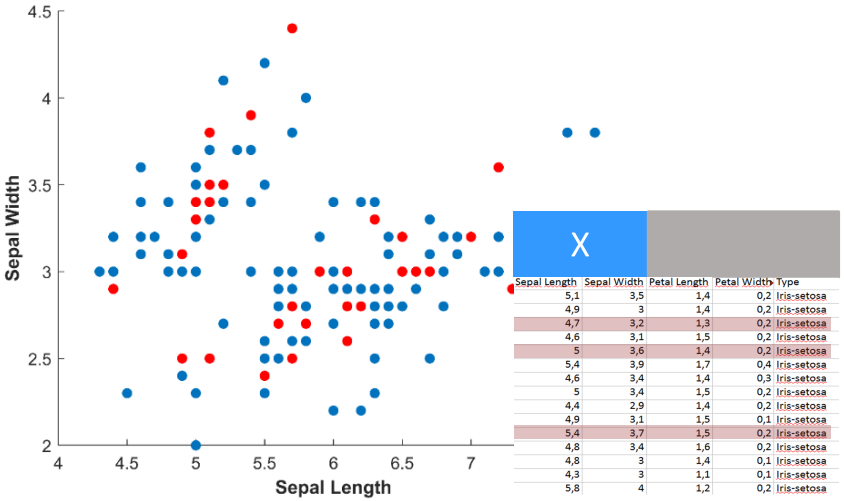


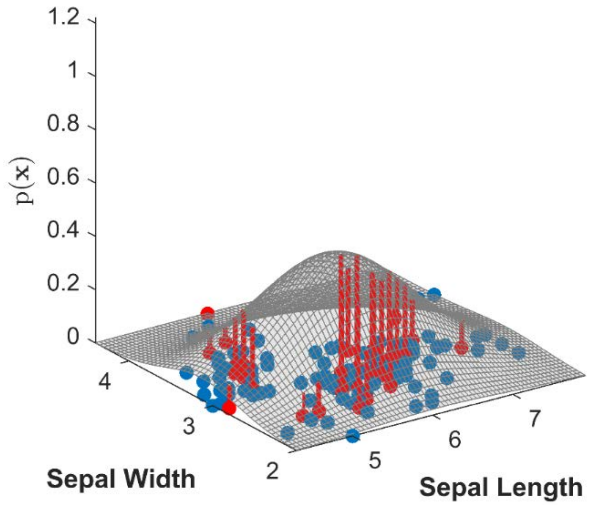


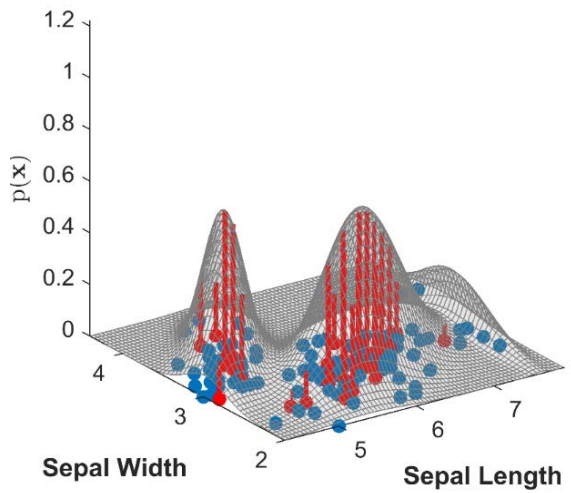
Unsupervised learning

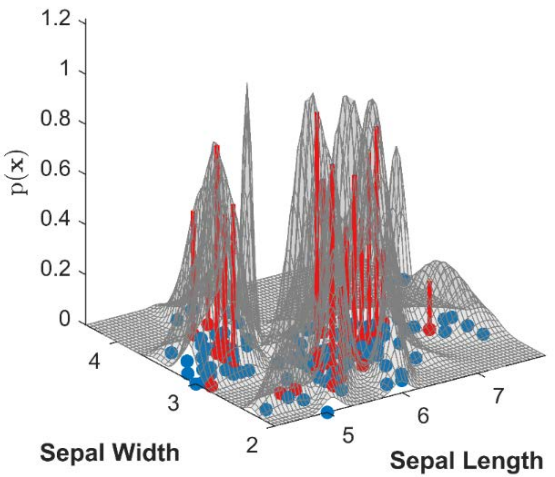


Find patterns in data
(explorative analyses)

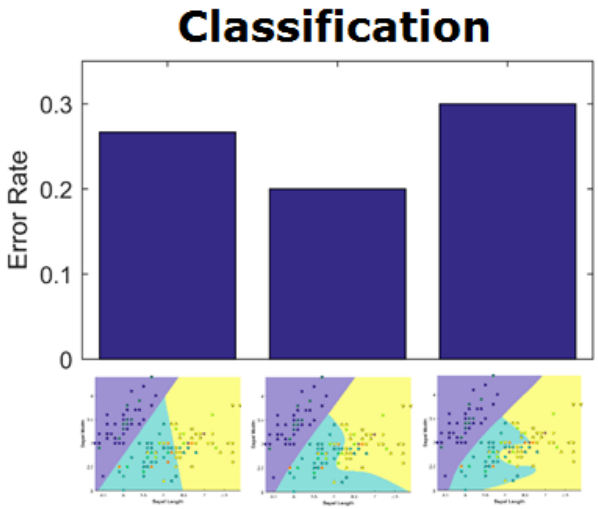


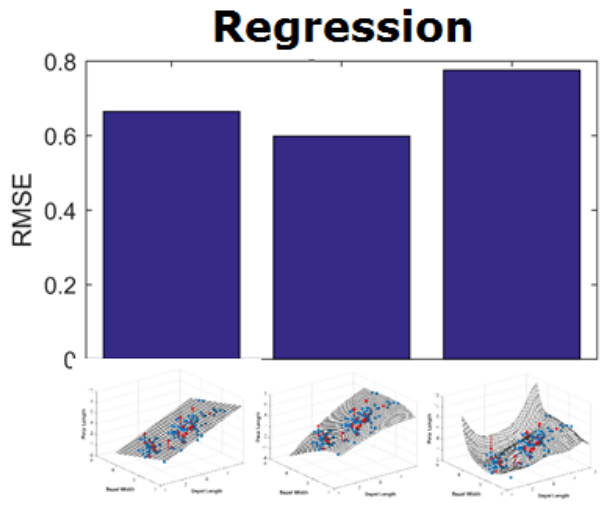


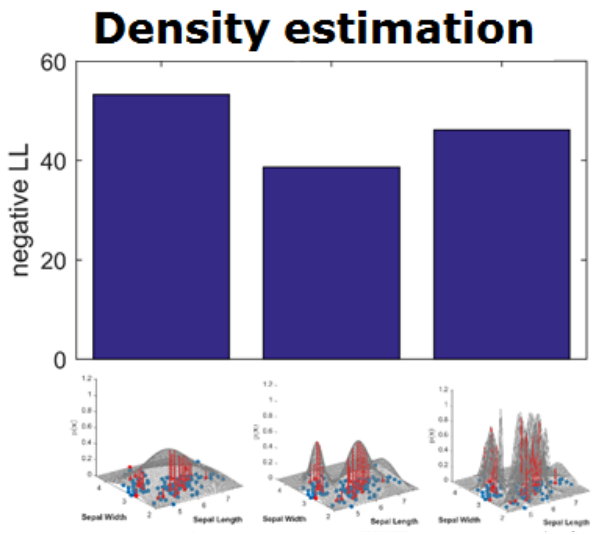




Experience = Amount of data (number of observations)
Task = Classification/Regression/Density estimation
Performance =How well we can predict classes/output values/where data occur







Machine learning workflow as taught in this course