Statistical Learning and Machine Learning

Lecture 1 - Introduction

August, 2025

Course's Team

Lecturers:
Naveed ur Rehman (naveed.rehman@ece.au.dk)
Teaching Assistants:
Daniel Østerballe (202205835@post.au.dk)
Purbak Sengupta (202310899@post.au.dk)
Sanzida Akter (202403885@post.au.dk)

Course information I

10 ECTS course of 14 week duration

Contact days:

- Monday (12:00 - 16:00): 2h of lecture (12-14) followed by 2h of exercises
- Wednesday (12:00 - 16:00): same format as on Monday

Problem set for exercise sessions:

- Programming exercises on GitHub (Monday 14:00-16:00)
- Theoretical exercises + Prog. exercises (Wednesday 14:00-16:00)
- Hybrid exercises focusing on your understanding of the topics (Wednesday 14:00-16:00)

Textbook: C. M. Bishop, "Pattern Recognition and Machine Learning", New York, Springer, 2006.

Course information II

Weekly Reports:

- You will need to submit a report summarizing your solutions on the programming exercises each week.
- **Must submit and pass at least 9 reports** (out of 12) to sit in the exam.

Assessment:

- Solely based on the **Final Exam**
- Submission and approval of minimum number of reports to sit in the exam

Feedback on weekly assignments:

- Binary decision (Pass or Fail)
- You will get concrete feedback on how to improve your assignment hand-in, in case you 'fail' the assignment

Brightspace

- Familiarize yourself with the course's page in Brightspace
- You will find curriculum, lecture material, report submission and sample exam questions

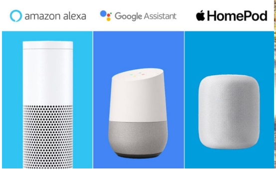
brightspace.au.dk

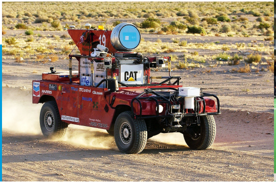
Objectives of today's lecture

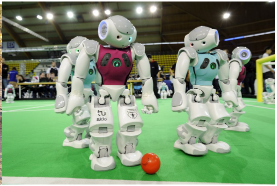
- To introduce artificial intelligence, machine learning and statistical learning
- To explain different types of machine learning problems
- Underscore the importance of the field via real-life applications
- To summarize the course content

Artificial Intelligence (AI)

- "AI refers to the ability of machines, or computers, to perform tasks that typically require human intelligence".
- AI enables machines to mimic certain aspects of human thinking and problem-solving, making them able to perform tasks smartly using a wide range of technologies e.g., machine learning, rule-based systems, robotics, NLP, and computer vision.







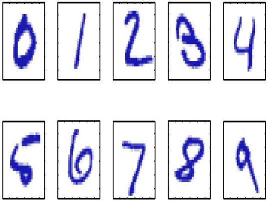
Virtual assistants

Self driving cars

Robo cup

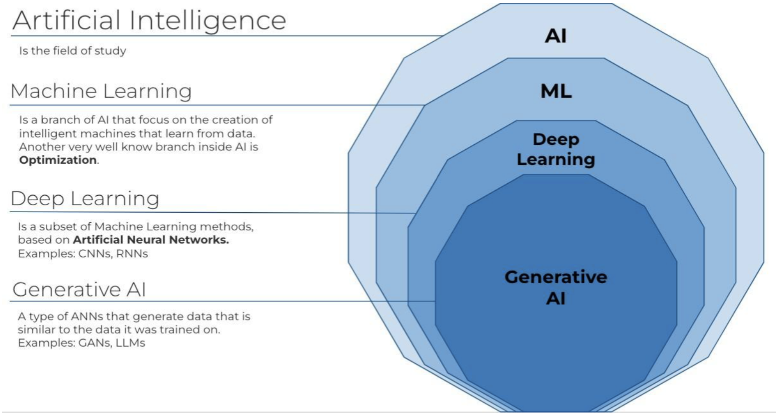
Machine Learning (ML)

- The main idea is to “give machines access to data and let them learn and improve for themselves”
- Relevant to problems which are hard to describe or model e.g., face detection
- solution: data-driven approach as opposed to model-based approach



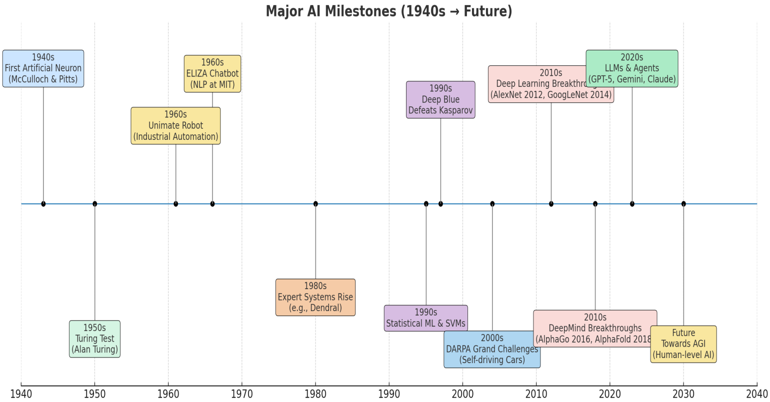
Face detection | Text recognition | Netflix recommendation

AI vs ML



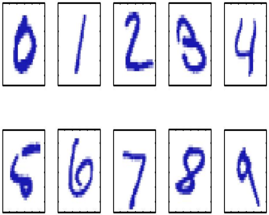
AI vs ML vs DL (Source: PXL Vision)

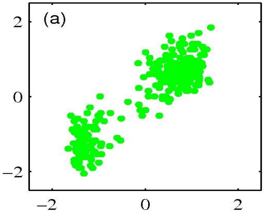
Timeline of AI/ML breakthroughs



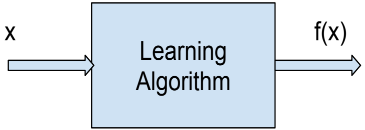
Goals of Learning

- The two main goals of learning are:
  1. **Data Understanding**: means extracting useful features or patterns from data to improve our understanding of the data and underlying process(es) e.g., clustering, dimensionality reduction
  2. **Prediction**: making important decisions about system of interest e.g., hand-written digit recognition; wind power forecasting
- **Challenges**: data may be complex/nonstationary, multi-modal, high-dimensional, noisy and depending on the application may be prohibitively large or inadequately small.





More about Prediction

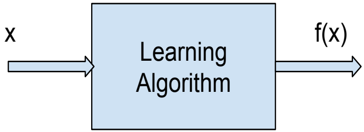


- x: input e.g., handwritten digits or corresponding features
- $f(x)$: learning output or prediction e.g., one of the 9 labels corresponding to digits 0-9 Other examples: recognizing a face from a sequence of images, video; time series forecasting.

Prediction can be the next step after data understanding (feature extraction) whereby we use the patterns/features from data to make important decisions about our system of interest.

Prediction: Classification vs Regression

- **Classification**: problems deal with assigning input data to a specified category. Here, the output $f(x)$ belong to a discrete set of values or labels. Examples are
  - Classifying email as spam or not
  - Classification of hand written digit
- **Regression**: are prediction problems in which the output $f(x)$ corresponds to a continuous range of values (continuous variable) e.g., wind power prediction.



Whether the following problems are classification or regression problems?

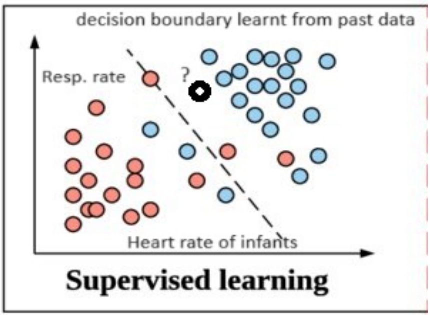
1. Whether one or more advertisements on a website will be clicked or not
2. Predicting prices of a house in a city
3. Whether an online order placed by a customer should be blocked or not based on his/her credit exposure
4. Dividing a customer-base into multiple groups (frequent buyers; seldom buyers etc) based on their spending habits
5. Predicting the salary of an employee based on his/her age, highest education, years of experience etc
6. Forecasting of electrical energy consumption in Aarhus for next-day (24 hrs ahead)

Whether the following problems are classification or regression problems?

1. Whether one or more advertisements on a website will be clicked or not (classification)
2. Predicting prices of a house in a city (regression)
3. Whether an online order placed by a customer should be blocked or not based on his/her credit exposure (classification)
4. Dividing a customer-base into multiple groups (frequent buyers; seldom buyers etc) based on their spending habits (classification)
5. Predicting the salary of an employee based on his/her age, highest education, years of experience etc (regression)
6. Forecasting of electrical energy consumption in Aarhus for next-day (24 hrs ahead) (regression)

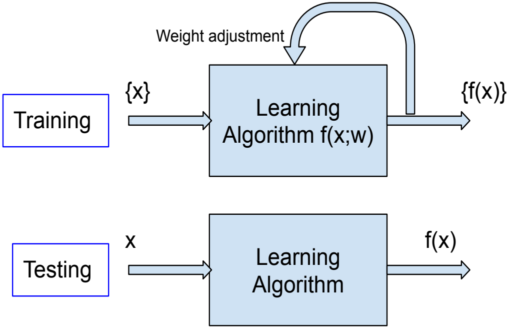
Types of Learning Problems I

- **Supervised Learning**



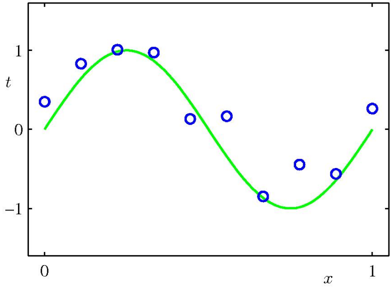
Types of Learning Problems II

1. Labeled training data includes input along with corresponding correct output
2. The training data is used to teach model to yield the desired outputs



Types of Learning Problems III

3. Concrete example: curve fitting



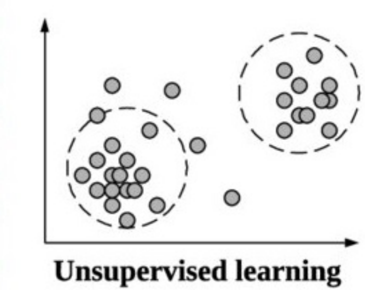
Green curve (target): $sin(2\pi x)$.
Blue circles (observations): possibly noisy

Types of Learning Problems IV

4. Input x may be raw input or relevant features from input e.g., scaled and translated images of digits; average image intensity over rectangular regions of images for face recognition
5. **Examples**:
   1. Classification problems (email spam detection, hand-written digit classification, delivering content that matches users interests in social media, streaming and online shopping platforms)
   2. Regression Problems (stock index prediction; wind power forecasting)
6. **Algorithms**:
   1. Linear Regression
   2. Neural networks (Multilayer Perceptron, Convolutional NN, Deep NN)
   3. Random Forest

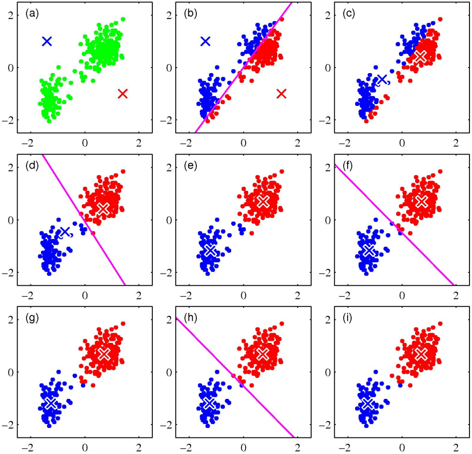
Types of Learning Problems V

- **Unsupervised Learning**: No known target values are given which means there is nothing to predict from data. Instead, the goal is to understand and analyze data based on its hidden patterns/structures e.g., data mining
  - Clustering algorithms process raw, unclassified data into groups containing similar information or structures e.g. K-means clustering



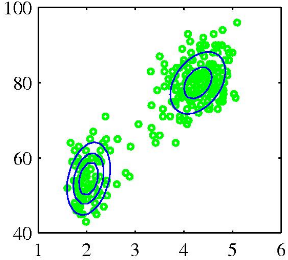
Types of Learning Problems VI

- Illustration of K-means clustering:



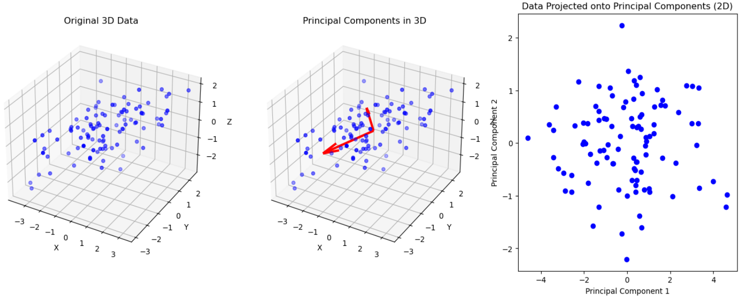
Types of Learning Problems VII

- Density Estimation: involves the estimation of probability density function of observed data.
  1. Parametric density estimation (selecting a common distribution and estimating its parameters from data) vs non-parametric density estimation(techniques to fit a model to the arbitrary data distribution)
  2. Example: Finding the mean and variance of the Gaussian distribution from data. How? Using maximum likelihood and Bayesian approaches.



Types of Learning Problems VIII

- Dimensionality Reduction: Transforming complex high-dimensional data into lower dimensional spaces. Why?
  1. Algorithms: Principal Component analysis (PCA), singular value decomposition (SVD)



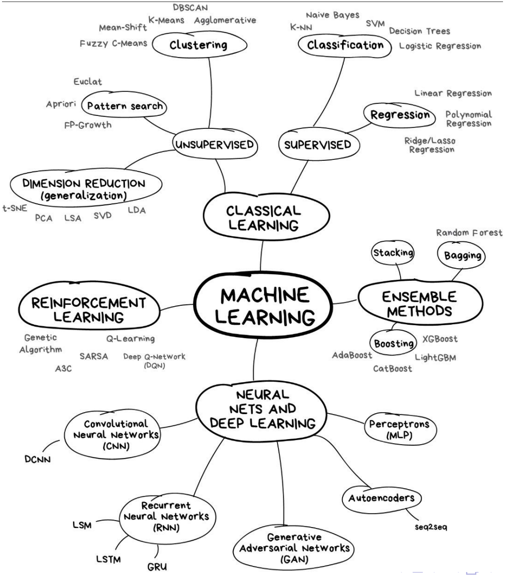
Whether the following problems are supervised or unsupervised problems?

1. Using prior information about spams, filtering out an incoming email into Spam or Inbox (normal)
2. Grouping customers through their purchasing behaviour without prior data of those groups
3. Handwritten digit recognition using past examples of handwritten digits and their true labels
4. Using electroencephalogram EEG (brain data) to classify different brain states (i.e., *relaxed* or *engaged* in cognitive activity); no previous examples of EEG data and brain states is available

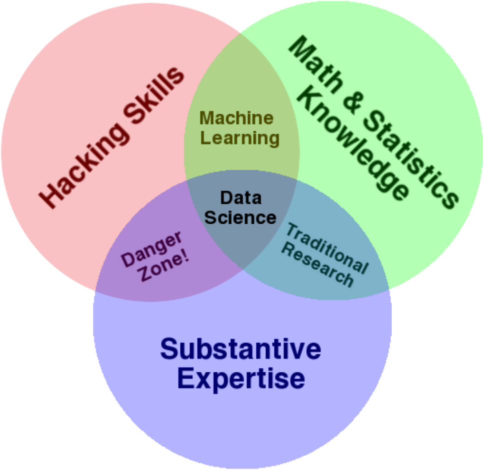
Whether the following problems are supervised or unsupervised problems?

1. Using prior information about spams, filtering out an incoming email into Spam or Inbox (normal) (supervised)
2. Grouping customers through their purchasing behaviour without prior data of those groups (unsupervised)
3. Handwritten digit recognition using past examples of handwritten digits and their true labels (supervised)
4. Using electroencephalogram EEG (brain data) to classify different brain states (i.e., *relaxed* or *engaged* in cognitive activity); no previous examples of EEG data and brain states is available (unsupervised)

Overview of machine learning methods



Data Science vs Machine learning



Statistical and Machine Learning in Action I

Automatic spam detection system

- Training data: 4601 emails sent to an individual (named George, at HP labs, before 2000). Each is labeled as spam or email.
- goal: build an automatic spam filter
- input features: relative frequencies of 57 of the most commonly occurring words and punctuation marks in these email messages.

	george	you	your	hp	free	hpl	!	our	re	edu	remove
spam	0.00	2.26	1.38	0.02	0.52	0.01	0.51	0.51	0.13	0.01	0.28
email	1.27	1.27	0.44	0.90	0.07	0.43	0.11	0.18	0.42	0.29	0.01

Hastie, T., Tibshirani, R., Friedman. J., (2009) The elements of statistical learning. 2nd ed. New York: Springer

Statistical and Machine Learning in Action II

Identifying risk factors for prostrate cancer from clinical measurements

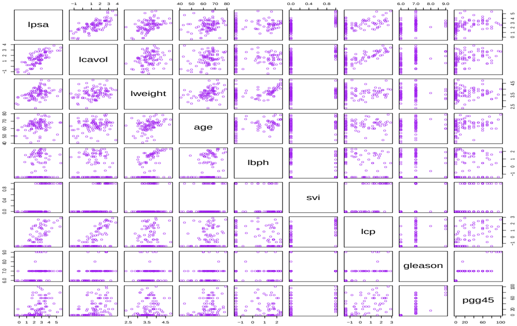


FIGURE 1.1. *Scatterplot matrix of the prostate cancer data. The first row shows the response against each of the predictors in turn. Two of the predictors,* `svi` *and* `gleason`*, are categorical.*

Hastie, T., Tibshirani, R., Friedman. J., (2009) The elements of statistical learning. 2nd ed. New York: Springer

Statistical and Machine Learning in Action III

Predict the identity of handwritten image of a digit

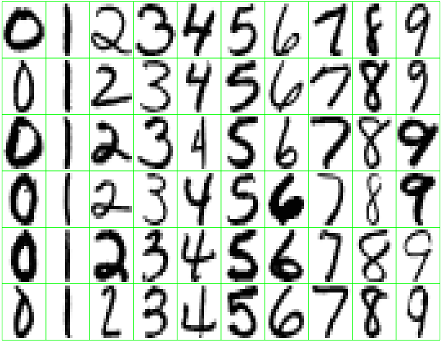
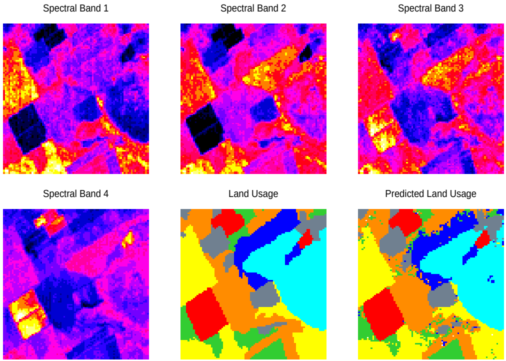


FIGURE 1.2. *Examples of handwritten digits from U.S. postal envelopes.*

Hastie, T., Tibshirani, R., Friedman. J., (2009) The elements of statistical learning. 2nd ed. New York: Springer

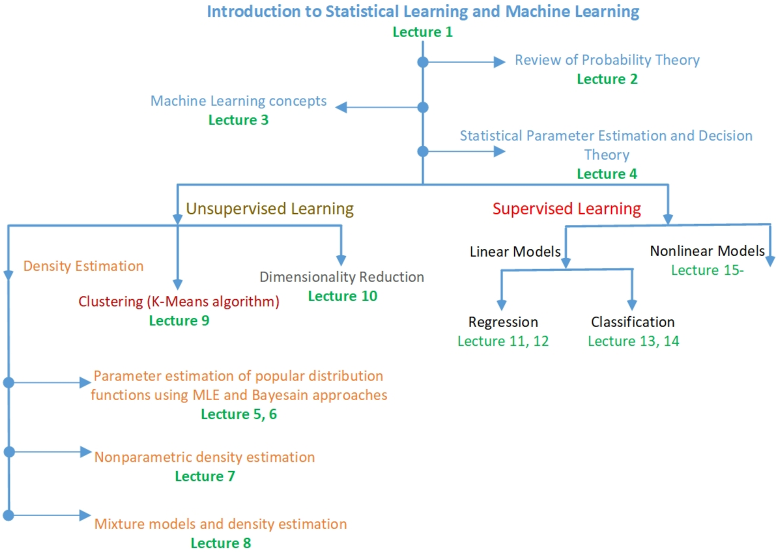
Classify a pixel in LANDSAT image by usage



usage$\in$ red soil, cotton, vegetation stubble, mixture, gray soil, damp and very damp stgray soil

Hastie, T., Tibshirani, R., Friedman. J., (2009) The elements of statistical learning. 2nd ed. New York: Springer

Course overview



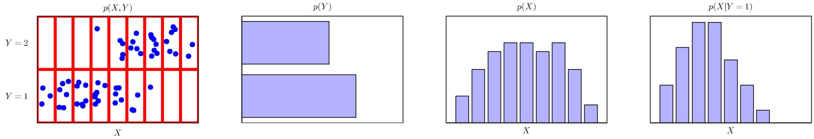
Course contents: Week 1

Introduction:

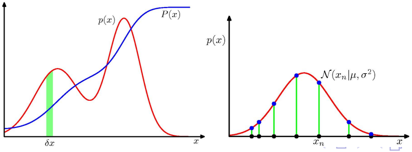
- This lecture...

Probability Theory:

- Basics of probabilities



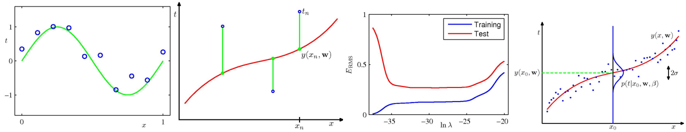
- Probability densities:



Course contents: Week 2

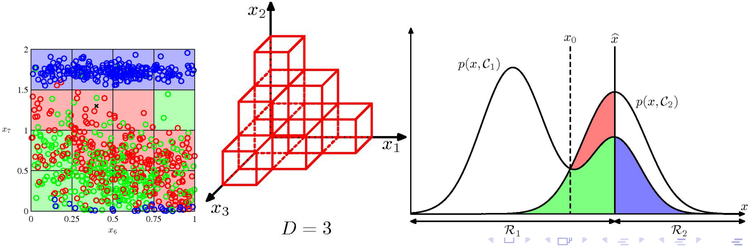
Polynomial curve fitting:

- linear regression based on polynomial representation of data points:



Decision theory:

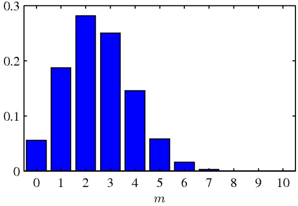
- Curse of dimensionality, Minimizing the misclassification error, Discriminant functions:



Course contents: Week 3

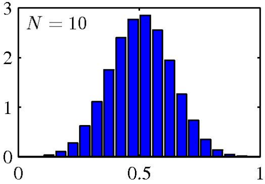
Probability distributions:

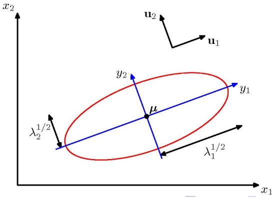
- Binary variables, Binomial distribution:



Probability distributions:

- Multinomial distribution and Gaussian distribution:

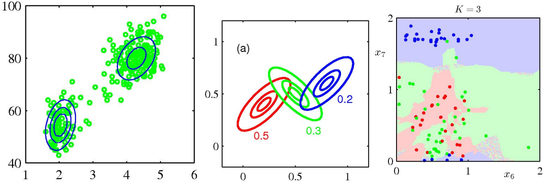




Course contents: Week 4

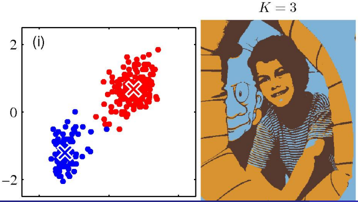
Probability distributions:

- Mixture of Gaussians, kernel density estimation:



Mixture models:

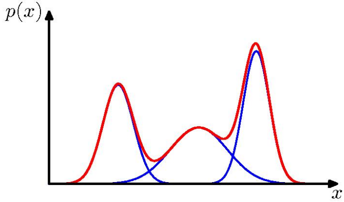
- Clustering, Mixture of Gaussians:

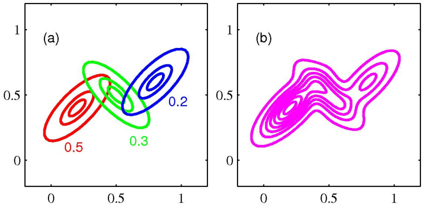


Course contents: Week 5

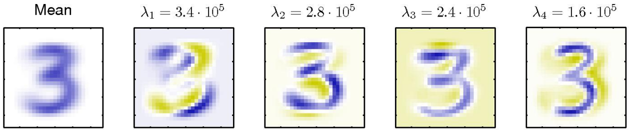
Expectation Maximization:

- Mixture of Gaussians, K-Means clustering:



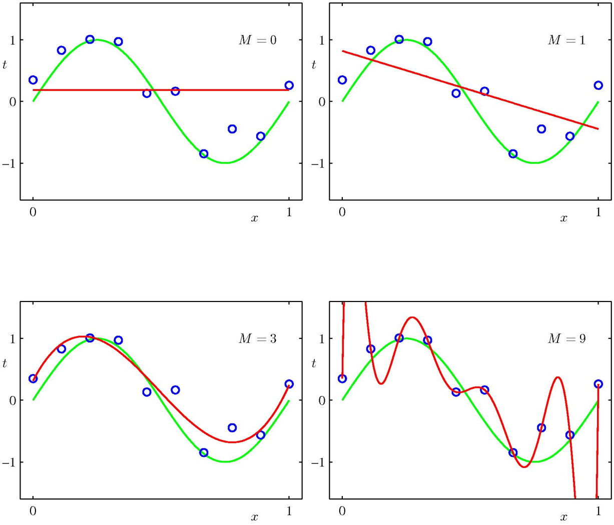


Principal Component Analysis:

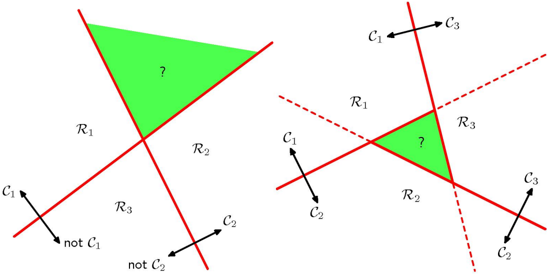


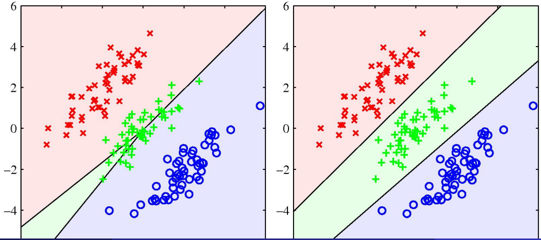
Course contents: Week 6

Linear models for regression:



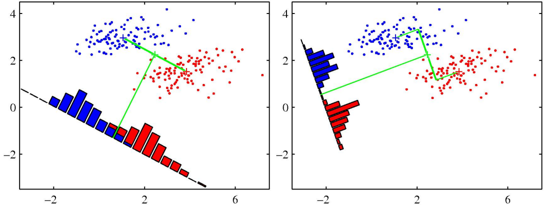
Linear models for classification:

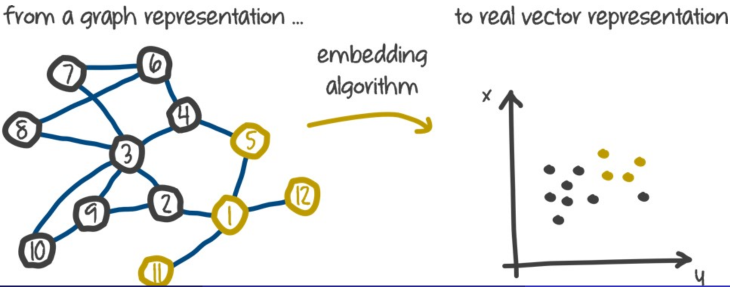




Course contents: Week 8

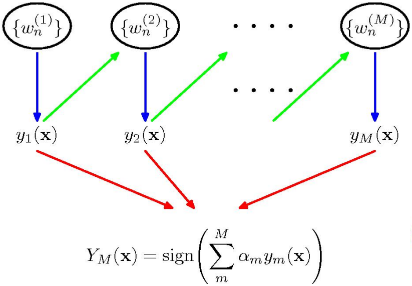
Linear subspace learning:

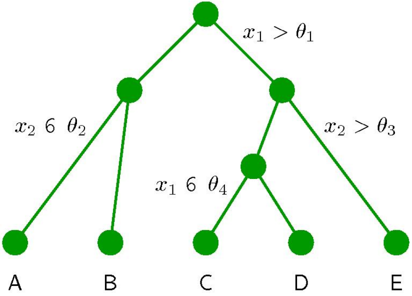




Course contents: Week 9

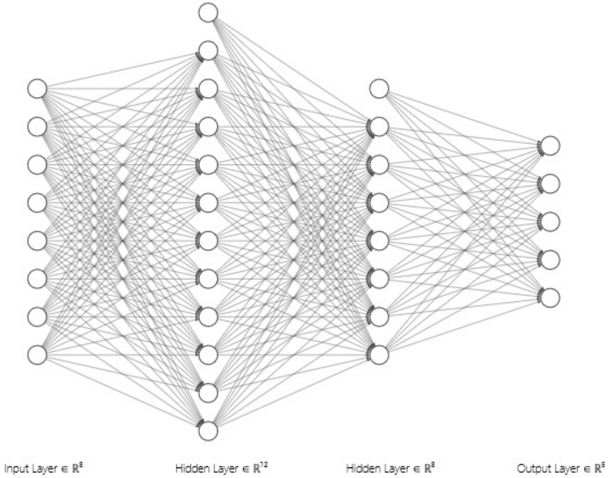
Combining machine learning models:





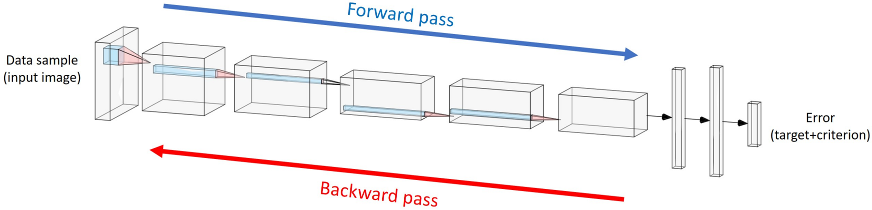
Course contents: Week 10

Neural networks:

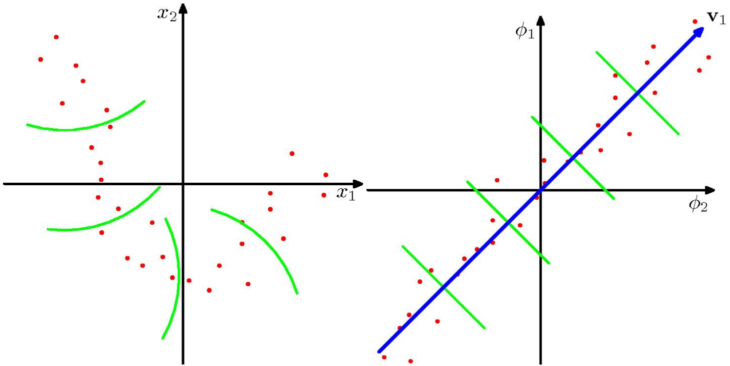


Course contents: Week 11

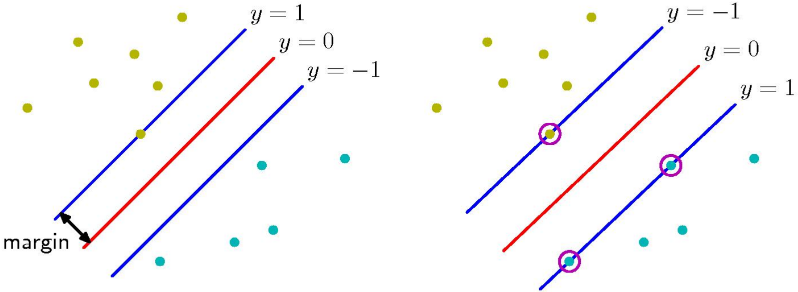
Convolutional neural networks



Kernel methods:



Sparse kernel methods:



Course contents: Week 14

Applications: