Introduction

Prerequisites: basic linear algebra (matrices and vectors, eigenvectors), basic statistics and probability theory, basic programming skills (Python).

Online resources:

- UCI Machine learning repository, which contains a large collection of standard datasets for testing learning algorithms: **archive.ics.uci.edu/ml**
- examples of recent work in machine learning, start by taking a look at the conferences NIPS: **old.nips.cc** and ICML. Some other related conferences include UAI, AAAI, IJCAI.
- platform for data mining competitions: **www.kaggle.com**

Goals of the course:

1. Convey my own excitement;
2. Teach how to apply algorithms;
3. Give a background for research in machine learning.

Final project:

- You can form study groups (max is 2 students per group)
- Two options for project:
  - investigate some aspect of machine learning;
  - apply machine learning algorithm to the problem you are interested in.

Course organization:

Machine Learning is one of the most interesting interdisciplinary area. It touches many different industries and you will find the stuff we take in this class very useful. During the last 15-20 years we have obtained much more capabilities for machine learning. Some examples of machine learning problems are:

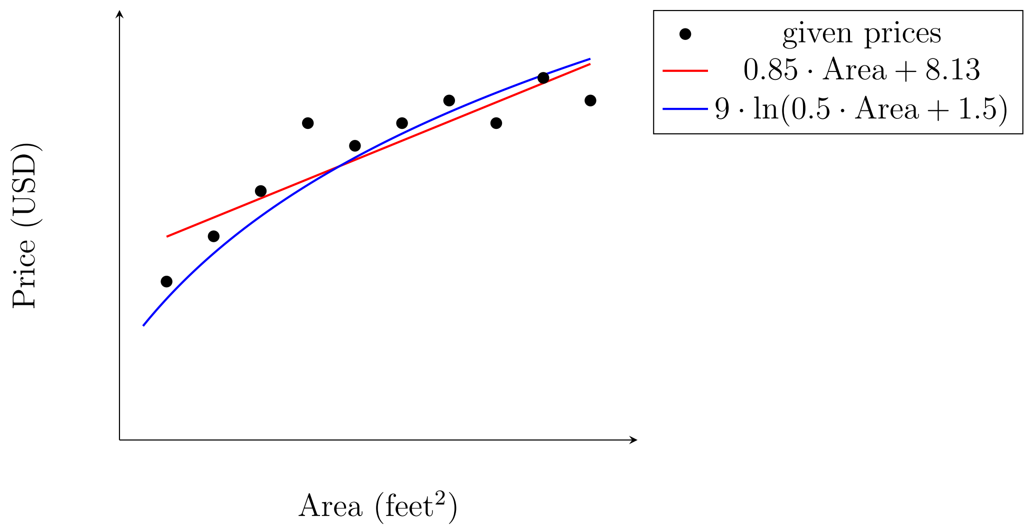
- digits recognition (amount on cheques)

- database mining (patient medical records)
- recommender systems (YouTube movies)

To solve all these problems we need learning algorithms. The course is divided in 4 sections:

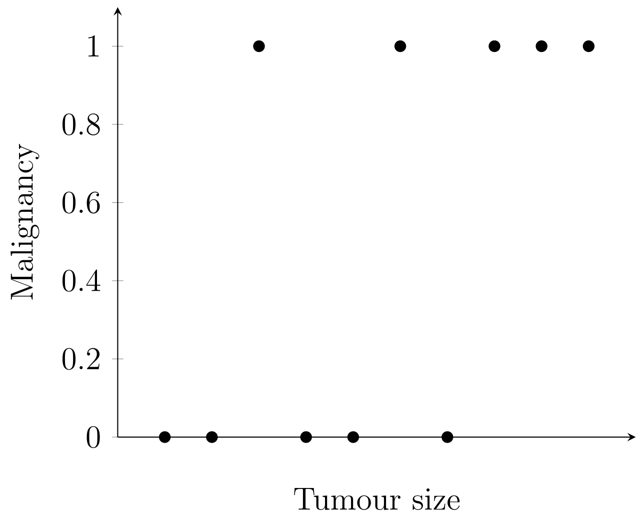
- **Supervised learning**

  For example, we can collect housing prices (in USD) and areas (in feet^2). The simplest way to visualise is to put them on the graph:

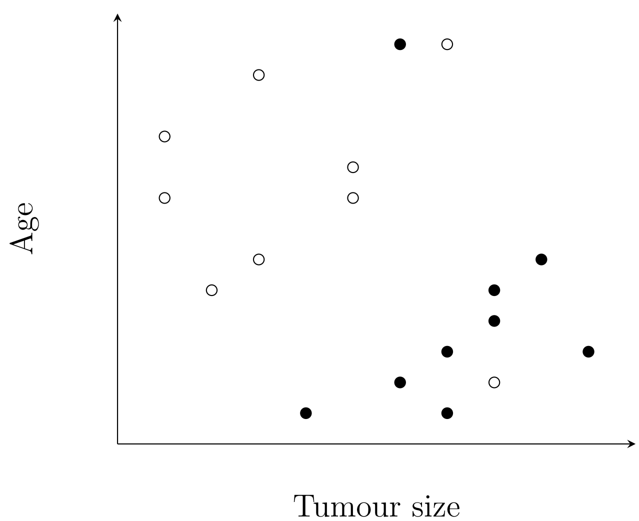
  

  Now we can try to find some linear or more complicated function to predict the price based on the given area. We call it **regression problem**.

  For another example we consider the problem where the tumour malignancy should be predicted based on the tumour size. The output for this problem is not continuous, but rather discrete: 0 means that the tumour is malignant, 1 - benign. This is a **classification problem.** Then the picture looks different:



More general example is when number of inputs is bigger than one: you may want to predict malignancy based on tumour size and age:



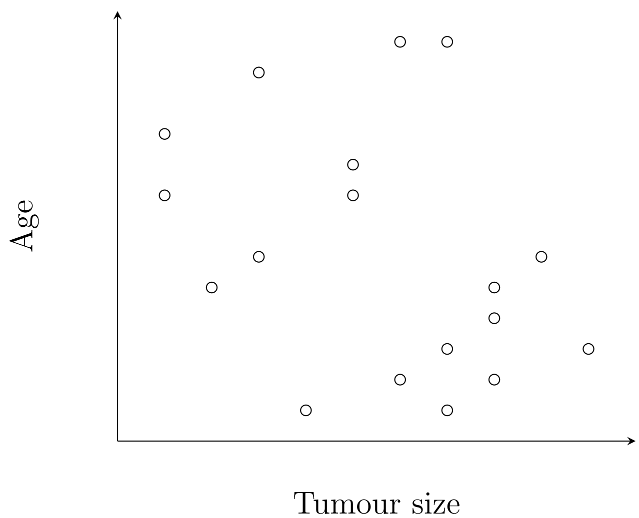
The general idea is that we should split classes by some curve. In many problems the data could have dimensions bigger than 3 and it is impossible to show it on the graph. But there are a lot of fascinating algorithms (like support vector machine) that can handle these complicated cases.

- **Learning theory**

  This part requires a lot efforts of mathematicians. We try to understand how algorithms work, what is the guarantee that the algorithm will work? The example for the previous problem is: how much training data do we need to predict prices?

- **Unsupervised learning**

  In this section we should find interesting structures in the data, for example, clusters.

  Some areas that involve the unsupervised learning technique are geophysics, image processes, social network analysis, market segmentation. Another example of unsupervised learning is cocktail party problem: we should split voices of different people based on audio records. The solution of this problem could be obtained by the independent component analysis (ICA).

- **Recommender systems**

  In this part we have the set of users and set of objects (for example, set of movies from YouTube) with ratings given by users to the objects. If we have missed ratings, we should complete them, i.e. how would user rate the given movie? The example of the algorithm that solves this kind of problems is factorization machine.