Week 1

July 17, 2021

Learning Objectives

- Deep learning's rise is being driven by several major factors.
- Deep learning in supervised learning.
- The major types of models (such as CNNs and RNNs) and when they should be used.
- When is deep learning appropriate to use?

Why is ML so prevalent?

- It is a branch in artificial intelligence.
- We aim to build a machine that can "think."
- Machines may be programmed to do mathematical tasks.
- It is impossible (for us) to create AI based on set of rules.
- We want robots to discover such rules on their own, to learn from the data.

Examples

Machine learning has exploded in popularity as a result of the huge amount of data generated and colected in recent years.

Database mining sources

- Data from the internet (click-stream or click through data). Mine to better understand users. This is common in Silicon Valley.
- Medical records. More and more data is being saved electronically.
- Biological data. Gene sequences, ML algorithms improve our understanding of the human genome.
- Engineering info. Data from sensors, log reports, photos etc.

Applications that we cannot program by hand

- Autonomous helicopter.
- Handwriting recognition.
- Natural language processing (NLP).
- Computer vision.

Self customizing programs

- Netflix

- Amazon

- iTunes genius

What is machine learning?

Arthur Samuel (1959)

- "Field of study that gives computers the ability to learn without being explicitly programmed".

- Samuels created a checkers software and had it play 10,000 games against itself. Work out which board positions were good and bad depending on wins/losses.

Tom Michel (1999)

- "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".

- The checkers example: E = 10000s games, T is playing checkers, P if you win or not.

Several types of learning algorithms

- Supervised learning: Teach the computer how to do something, then let it use it's new found knowledge to do it.

- Unsupervised learning: Let the computer learn how to do something, and use this to determine structure and patterns in data.

- Reinforcement learning.

- Recommender systems.

Supervised learning - introduction

- Probably the most prevalent form of machine learning type.

- Example: How can we predict housing prices? Ans: Collect house pricing data and examine how it relates to size in feet.

House price prediction
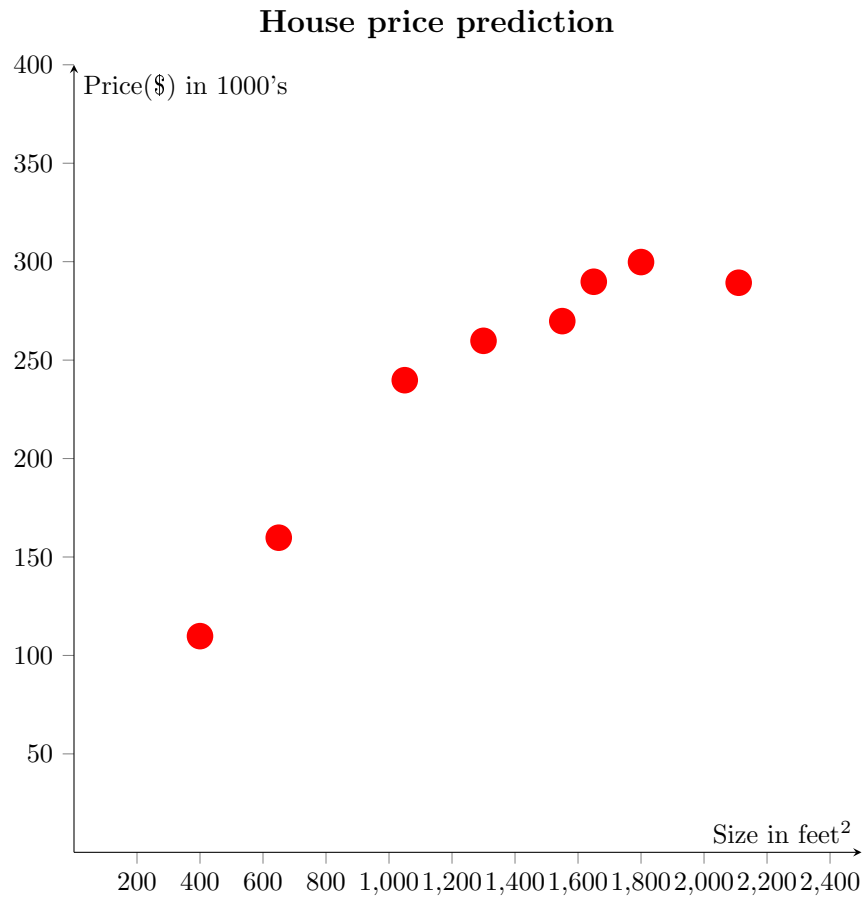
Example problem: "Given this data, a friend has a house 750 square feet - how much can they be expected to get?"

What approaches can we use to solve this?

- Straight line through data. Maybe $150,000.

- Second order polynomial. Maybe $200,000.

- One point we'll go over later is whether to use a straight or curved line.

- Each of these techniques is a method of carrying out supervised learning.

We also call this a regression problem

- Predict continuous valued output (price).

- No real discrete delineation.

3

## Define breast cancer as malignant or benign based on tumour size
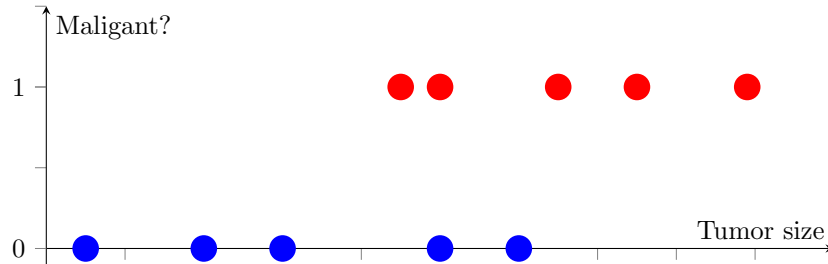
- Can you estimate the prognosis based on the tumor's size?

- This is an example of a classification problem.

- Classify data into one of two categories - malignant or not - with no in-betweens.

- In classification problems, the output can only have a discrete number of potential values.

You may have many attributes to consider.

We also call this a regression problem

- Predict continuous valued output (price).

- No real discrete delineation.
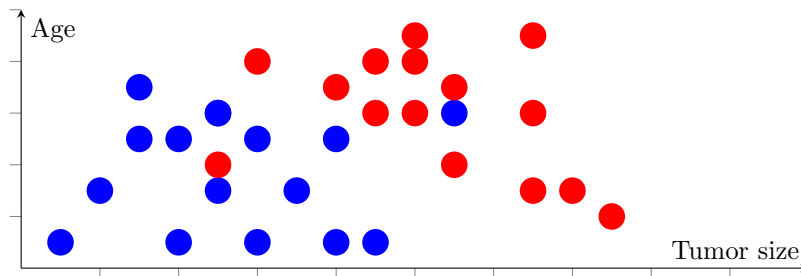
- You can try to establish different classes based on that data by drawing a straight line between the two groups.

- Defining the two groups with a more sophisticated function (which we'll go over later)

- Then, when you have someone with a certain tumor size and age, you can ideally utilize that information to assign them to one of your classes.

4

Unsupervised learning - introduction

- Second major type.

- We use labeled datasets (as opposed to unlabeled).

Clustering algorithm

- Google news. Groups news stories into cohesive groups.

- Genomics.

- Microarray data. Have a group of individuals. On each measure expression of a gene. Run algorithm to cluster individuals into types of people.

- Organize computer clusters. Identify potential weak spots or distribute workload effectively.

- Social network analysis. Customer data.

- Astronomical data analysis. Algorithms give amazing results.
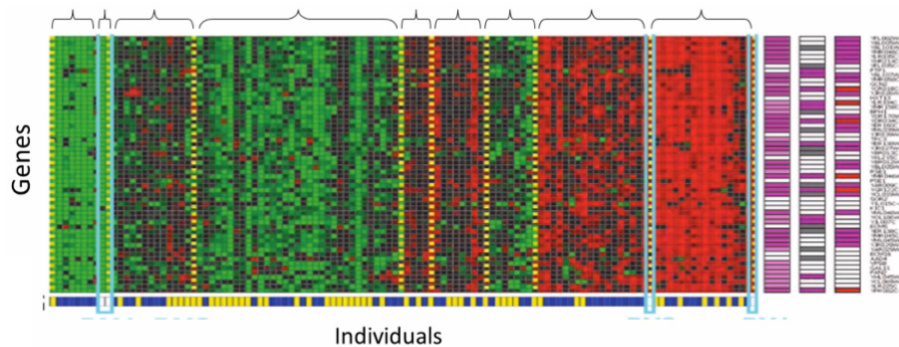
Cocktail party problem

- Depending on where your microphone is, record slightly different versions of the conversation.

- Give the recordings to the algorithm.

- It should be able to figure out that there are two audio sources.

$$[W, s, v] = svd((repmat(sum(x.*x, 1), size(x, 1), 1).*x)*x');$$