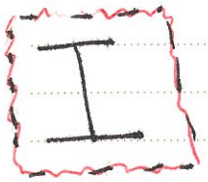


INTRODUCTION

Lecture 1: Welcome

→ The ranking which Google uses for search results is a machine learning algorithm. The reason, probably, is that it's impossible to hard-code all possible searches people might do.

→ Spam filters are also learning algorithms. Again, it would be impossible to list all possible spam emails one could receive.

→ Machine learning grew out of work in Artificial Intelligence (AI).

⇒ There are many tasks for which we need machines to learn on their own.

→ Some modern application of Machine learning:

(i) Database mining: growth of automation/web has resulted in much larger datasets which need to be understood. Ex) Web click data, medical records, biology, ...

(ii) Applications that cannot be programmed by hand:
→ handwriting recognition
→ Natural Language Processing (NLP)

(iii) Self-customizing programs: Netflix Product recommendations

Again, with millions of users, it's impossible to write code for each one to recommend new shows they might enjoy.

(iv) Understanding the human brain! How is it that we humans learn? (read AI)

Lecture 2: What is Machine Learning?

Some definitions of Machine Learning:

(i) Arthur Samuel (1959): Field of study that gives computers the ability to learn without being explicitly programmed.

↳ He built a Checkers-playing program where the computer would learn from playing tens of thousands of games against itself. This was the world's first self-learning program.

(ii) Tom Mitchell (1998): A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

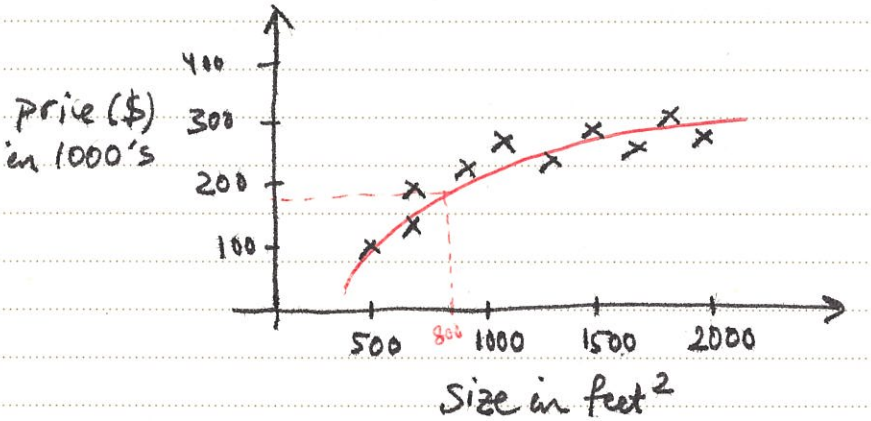
- Two main types of Machine learning algorithms

Supervised learning Unsupervised learning

Lecture 3: Supervised Learning

→ probably the most common type of machine learning problem.

→ Simple example: suppose we have the following data for price of houses:

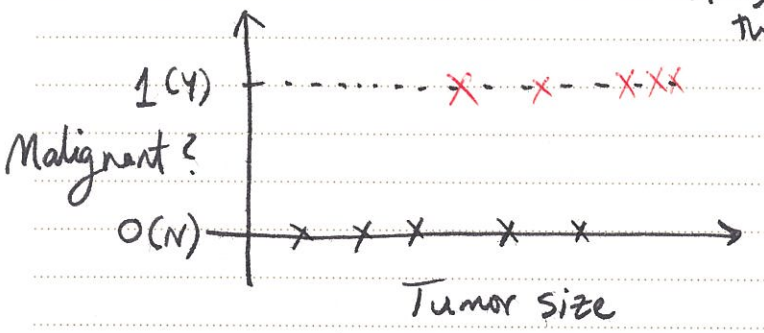


Question: what is the price of a house which is 800 feet²? Well, the learning algorithm could fit a curve through the data points!

* Supervised learning: the name comes from the fact that the training sample (i.e. our data) contains the "right answers". In other words, for a set of inputs, we know what the output should be.

→ Regression: predict continuous valued outputs. (as opposed to discrete values).

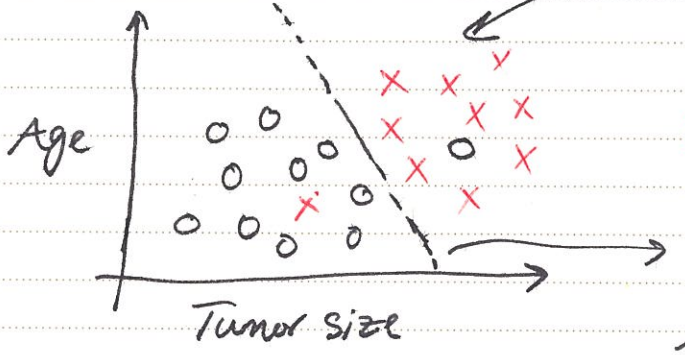
→ Example 2) Breast cancer. We have data on size of a bunch of breast tumors, as well as if they're malignant or not.



Question: if we have a tumor with a new size, what is the probability that it's malignant?

This is an example of a classification problem

Discrete-valued output. (could have multiple discrete outcomes obviously)



O ≡ benign
X ≡ malignant

The learning algorithm may decide that this line in the Age-Tumor size plane best separates benign & malignant tumors.

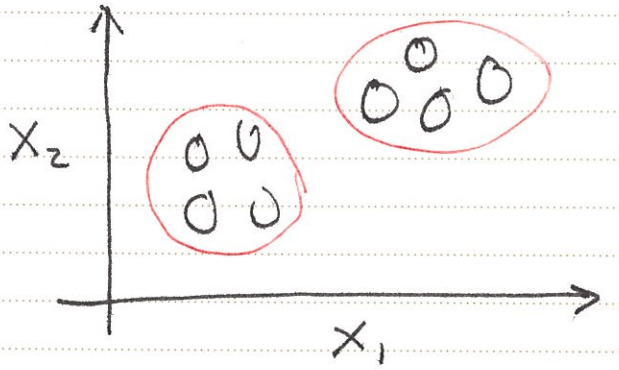
→ In this last example we had two features (age & tumor size). Andrew claims that there are problems for which an infinite number of features is necessary (I hope at least countably infinite...) & that there are learning algorithms that can deal with this!

(This algorithm is called the Support Vector Machine)

Lecture 4: Unsupervised Learning

→ In supervised learning, the training sample tells us the correct output for a set of inputs (e.g. a tumor whose size and age we know is malignant -- inputs: age & size of tumor; output: malignant). Then we seek to predict the output when we get new inputs.

→ Unsupervised learning is different in that we are not given the right answers. It's more like, here's a data set, find some structure in it:



Ex) A learning algorithm might tell us that most of the data is concentrated in two different regions. This sort of algorithm is called a Clustering Algorithm

→ A real-life example of a machine-learning algorithm which uses clustering is Google News!

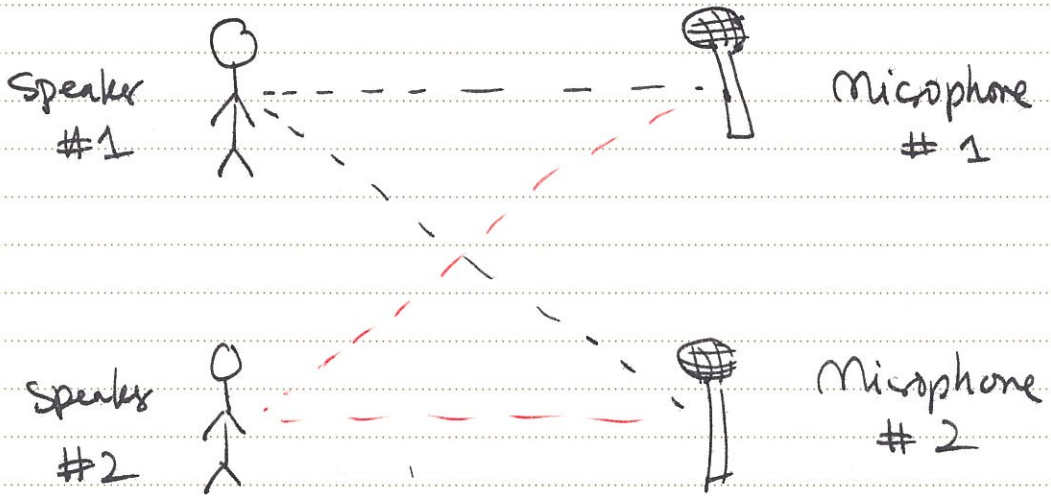
Google News looks at thousands of news & it groups them similar news together. For example, it would recognize that these articles from these different sources are about the same topic & it would cluster them together:

① THE SOURCE: BP Kills Macondo, But Its Legacy Lives On.
② CNN: Well is dead, but much Gulf Coast work remains.
③ Guardian: BP oil spill costs nearly $10 bn.

(This is remarkable!)

2) Another example of an unsupervised learning algorithm:

The Cocktail Party Problem



Speaker 1 & 2 are talking at the same time. In microphone 1, the sound of speaker 1 is recorded more loudly than speaker 2, simply because speaker 1 is closer to microphone 1. And vice versa for microphone 2.

2) If the volume offsets are pronounced enough, a human would easily recognize by listening to the recordings of microphones 1 & 2, that there are two speakers & even to separate out what the two speakers are saying. The amazing thing is that there are unsupervised-learning algorithms that are capable of doing the same! Based on the volume offsets, they recognize the structures of the two speaker sounds & are able to separate them out!

* On programming environments

→ The recommended language is Octave, which is a free open source software.

→ Apparently it is much quicker to build learning algorithms in Octave. It is only after a working prototype written in Octave exits, that people rewrite the code in, say, C++ to make it more efficient.

→ This is standard practice in Silicon Valley.