Introduction

25 November 2020 09:21 PM

WEEK 1

MACHINE LEARNING

→ As per Arthur Samuel (1959)

"Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed"

→ As per 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 of T as measured by P, improves with experience E."

★ EXAMPLE

For a game of checkers,

→ The experience E would be the experience of having the program play tens of thousands of games itself
→ The task T would be the task of playing checkers
→ The performance measure P would be the probability that it wins the next game of checkers against some new opponent

→ Machine Learning can be broadly classified into 2 categories

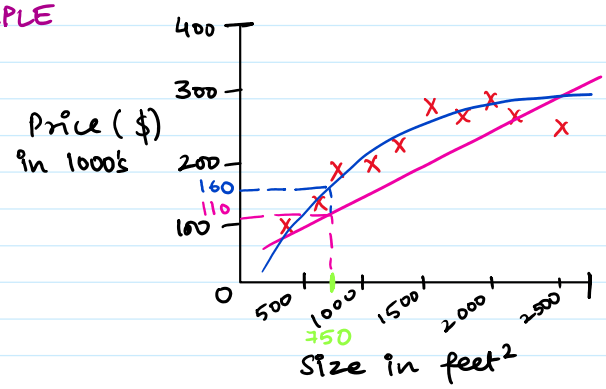
SUPERVISED LEARNING | UNSUPERVISED LEARNING

→ There are others as well but not as popular :
- Reinforcement Learning
- Recommender Systems

SUPERVISED LEARNING

→ Supervised learning is probably the most common type of machine learning

★ EXAMPLE



House Price Prediction

↳ Data has been collected from city of Portland, Oregon ←
↳ Plot of the data looks like this

Q Let's say you have a friend with a house of 750 sq feet and they are hoping to sell the house. They want to know how much they can get for the house.

→ How can a learning algorithm help here ?

get for the house.

→ How can a learning algorithm help here ?

↳ One thing the learning algorithm might do is put a **straight line** through the data.
↳ Shows 110 K for the house

↳ What if instead of fitting a straight line, we post a **polynomial function**
↳ Shows 160K for the house

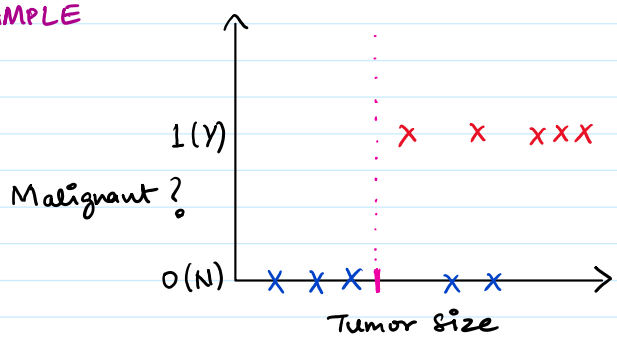
→ Each of the above fits would be a fine example of a learning algorithm
→ This is an example of "Supervised Learning Algorithm".

→ The term "Supervised" refers to the fact that we gave the algorithm a data set in which the 'right answers' were given.
↳ We gave it a data set of houses in which for every example in this data set, we told the algorithm what is the right price
↳ And the task of the algorithm was just to produce more of such right answers

- To define a bit of terminology, this is called a Regression problem

predict continuous valued output (example : price)

★ EXAMPLE



Breast Cancer (malignant, benign)

From medical records, predict whether breast cancer is malignant or benign

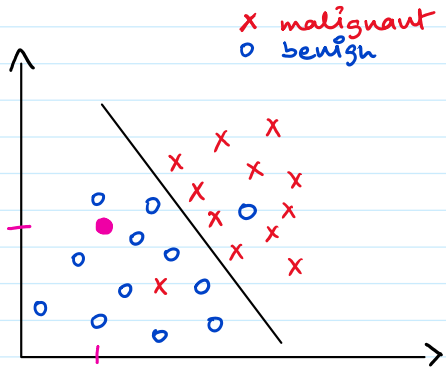
Q For this tumor size, what is the chance that the tumor is malignant v/s benign ?

- To define a bit of terminology, this is an example of a Classification problem

discrete valued output (0 or 1) ← classes / labels
<we can have more than 2 classes>

- In this example, we have only used 1 feature or attribute to predict whether a tumor is malignant or benign ~ Tumor size
↳ We can have more such features

★ EXAMPLE



Breast Cancer (malignant, benign)

This time we have more features

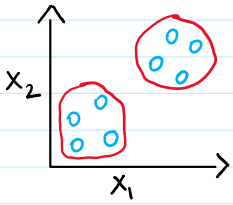
Q For this particular point, we have to tell what is the chance for the tumor to be malignant or benign.

Tumor size

→ In a dataset like this, what a learning algorithm might do is fit a straight line to the data to try to separate out the malignant tumors from the benign ones
↳ After doing so, we can say that the point in question has a higher chance of being benign

⊡ UNSUPERVISED LEARNING

→ In Unsupervised Machine Learning, we are given data which has no classes or labels



Given this dataset, the unsupervised learning algorithm might decide that the data lives in 2 different clusters

→ The problem always is — here is some data, can you find some structure in the dataset?
→ Applications — Google News, Gene Array categorization, Market Segmentation etc.

→ We derive structure from the data but we don't necessarily know the effects of the variables

→ With unsupervised learning, there is no feedback based on the prediction results