Machine Learning by Andrew NG Lecture Notes

Vinod Kumar

November 7, 2016

Introduction

What is Machine Learning

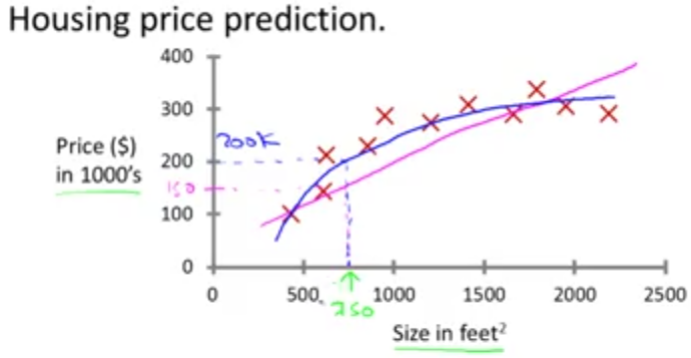
Field of study that gives computers the ability to learn without being explicitly programmed.

Why Machine Learning

- Many programs can't be hardcoded.For example Autonomous helicopter,handwriting recognition, most of Natural Language Processing (NLP), Computer Vision etc.
- User based recommendations can't be done without machine learning.As each user requires a different program,it is not quite possible to write so many programs for so many people.Ex Amazon,Netflix etc

Types Of Machine Learning

Supervised Machine Learning :-



An Example: Consider the data set as described in the image above.To find the price of a new house,from it's size we can try to fit a straight through the data and find out the price of the house based on this straight line.We can also use a quadratic or a second order polynomial for the same problem to get a different answer.This is an example of Supervised Learning.

The term supervised learning refers to the fact that we use a data set in which the right answers were given.In the above example we used a data set of houses in which for every example in this data set, we told it what is the right price and the task of the algorithm was just to produce more of these right answers for new instances.To be a bit more precise the above example is what we call a regression problem which is a type of supervised problem where we try to predict a continuous value output.

In case the data set is discrete then the problem is a classification.One example is when you try to predict whether a tumour is malignant or benign according to the tumour size using a data set with tumour sizes and the type of tumour.

Unsupervised Learning :-
The term unsupervised learning refers to the fact that we gave the algorithm a data set in which we only specify the features of the instances and not the type of the instance.The algorithm should find the structure in the data and draw conclusions from it.One way to do is to divide the given instances in the data sets into clusters and when a new instance is given the learning algorithm can predict which cluster this new instance belongs to.This is known as clustering algorithm.

An Example: Google news collects many new stories on the web and groups them together into different categories using this algorithm.This is done so that all stories about a single incident are grouped together.Also this technique is used in social network analysis where given knowledge about which friends you email the most or given your facebook friends or your Google+ circles, we can automatically identify which are cohesive groups of friends.

An interesting observation The problems from the above examples can also be done by supervised learning if given a proper data set ,i.e. even supervised learning can make clusters out of labelled data similar to the

ones made by unsupervised learning.This is one example where supervised learning and unsupervised learning can be used interchangeably.In fact this is true in most cases.The type of algorithm is mostly chosen according to the type of data available.

Model Representation of a Linear Regression

Consider the "price of the house" example described under Supervised Algorithm.We will use the same problem with a bigger data set.The data set used for learning in an algorithm is called as training set.In practice the objective of the learning algorithm is to find the hypothesis of the problem.In this case a hypothesis function $h_\theta(x)$ that can map the size of a house to it's price using the given data. For our example we will find the hypothesis by fitting a straight line through our training set.

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Cost Function

There can be many values for θ_0 and θ_1.To find the best possible hypothesis we use those values which give the least possible cost function.The cost function generally refers to how close the approximations are to the original values.In this case the cost function is

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)$$

where x_i and y_i represent the original size and price of the houses,$h_\theta(x_i)$ refers to the hypothesis function.