

Machine Learning

Shubham Chauhan

Class 1

31 Oct 2017

Outline

Introduction

How to predict?

Regression

Introduction

There are two broad types of techniques that exist in ML :

- ▶ Supervised Learning: Given a set of data

$$D = (x^n, y^n), n = 1, \dots, N$$

the task is to learn the relationship between the input x and output y such that, when given a novel input x^* the predicted output y^* is accurate. The pair (x^*, y^*) is not in D but assumed to be generated by the same unknown process that generated D . To specify explicitly what accuracy means one defines a loss function $L(y^{pred}, y^{true})$ or, conversely, a utility function $U = -L$.

Introduction

- ▶ Unsupervised Learning : Given a set of data $D = x^n, n = 1, \dots, N$ in unsupervised learning we aim to find a plausible compact description of the data. An objective is used to quantify the accuracy of the description. In unsupervised learning there is no special prediction variable so that, from a probabilistic perspective, we are interested in modeling the distribution $p(x)$. The likelihood of the model to generate the data is a popular measure of the accuracy of the description.

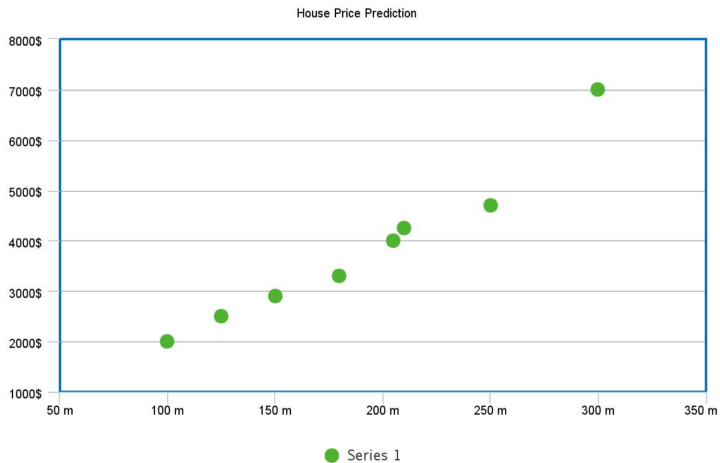
Introduction

- ▶ Data in real life problems are generally collected through surveys
- ▶ Surveys may have random human errors
- ▶ Most methods we will be using deal with expectations as they minimize the effect of error in our predictions
- ▶ Data Cleansing by finding outliers

Examples

For this class we will consider variation of cost of the house with the area of the house In this example we want to find a pattern or curve which this dataset follows, hence predict the price for any value of area.

Introduction



How to predict?

- ▶ Curve fitting is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points, possibly subject to constraints. - Wikipedia
- ▶ Thus we need a criteria to compare two curves on a dataset
- ▶ We describe an error function $E(f, D)$ which takes a curve f and dataset D as input and returns a real number
- ▶ Error function must be such that it can capture how bad the prediction is

Better fit?



Regression

Definition :

General Regression problem: Determine a function f^* such that $f^*(x)$ is the best predictor for y , with respect to D :

$$f^* = \operatorname{argmin} E(f, D)$$

Where $f \in F$, F denotes the class of functions over which the error minimization is performed