

Lecture 1.2

Machine Learning is Math.

We have the following:

$$x = (x_1, x_2, \dots, x_n)$$

where each entry in x is a feature.

Y is a response (what we are trying to predict) and Z is the *mapping* from *features* to *response*.

Machine learning is an approximation of Z .

Problem Space: Property

We got the following housing market problem space and we are trying to decide the features and the responses of this problem.

x_i = sqft, location, number of bedrooms/bathrooms, year built

Y = interest rate, price, property tax

So, each row will be a **sample** and the columns will be **features**.

Problem Space: Washing Machines

x_i = bed size, door: top or front, time, weight, washing cycle

Y = how long going to last, price, profit

All of these are good, but some of them are kind of useless because how is the washing cycle supposed to determine how long is it going to last?

Supervised Learning

This finds patterns in *fully observed* data, then try to predict *partially observed* data.

Unsupervised Learning

Find *hidden structure* in data when Y cannot be formally observed.