Machine Learning

Part 1: Gradient Descent

Zengchang Qin (PhD)

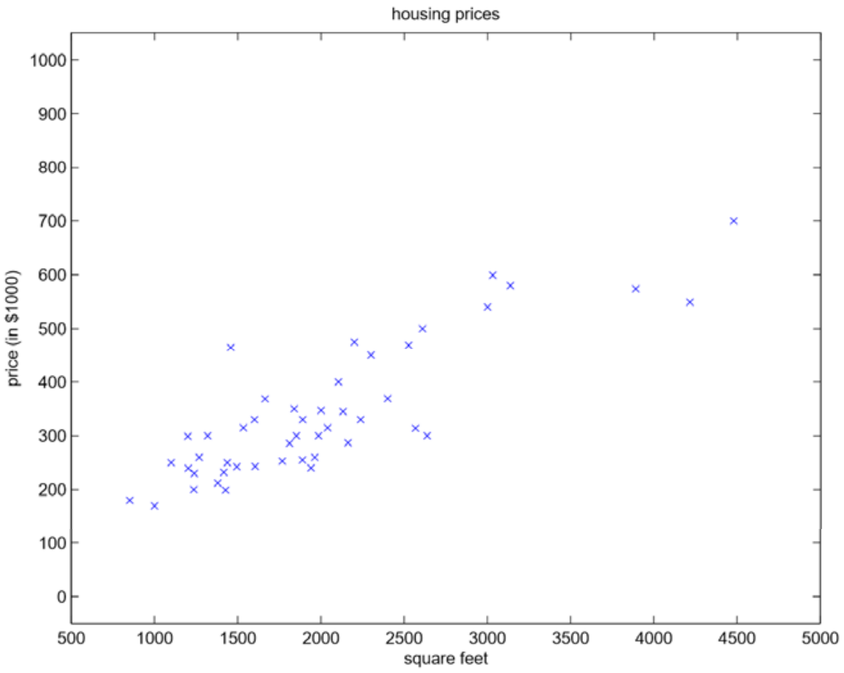
First Machine Learning Problem

Housing Price in Portland

In **Andrew Ng's** Lecture, there is a dataset giving the living areas and prices of 47 houses from Portland, Oregon. We are looking for a function gives the pattern of inputs-outputs.

Living area (feet^2)	Price (1000\$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	⋮

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



Multi-dimensional Attributes (Features)

A pair $(x^{(i)}, y^{(i)})$ is called a **training** example, $x \in R^d$ is called the **feature** and y is called the target or label of the example.

To perform **supervised learning**, we must decide how we're going to represent functions/hypotheses h.

Living area (feet^2)	#bedrooms	Price (1000$s)
2104	3	400
1600	3	330
2400	3	369
1416	2	232
3000	4	540
⋮	⋮	⋮

Machine Learning Problem

Given a database $D = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots (x^{(N)}, y^{(N)})\}$

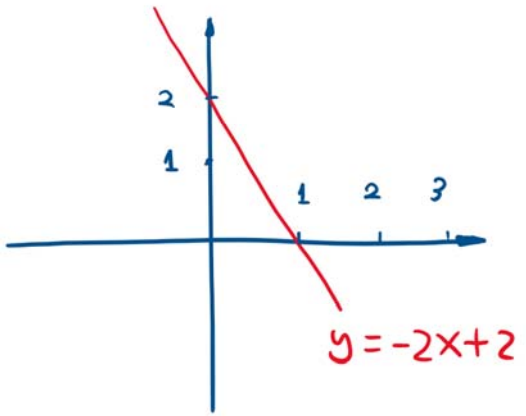
Where $x^{(i)}$ is a Vector in n-dimensional space and $y^{(i)}$ is a scalar

i.e.: $x_i \in R^n$, $y_i \in R$

We hope to learn $f(x^{(i)}) \rightarrow y^{(i)}$.

This is a typical machine learning problem

If our hypothesis of relation function is a linear model.



$y = ax + b$

$y = \theta_1 x + \theta_0$

if we set $x_0 = 1$

$y = \theta_1 x_1 + \theta_0 x_0$

$= \theta^T x$

$\theta^T = (\theta_0, \theta_1)$

$X = \begin{pmatrix} x_0 \\ x_1 \end{pmatrix}$

In general form

$y = \theta^T x \quad \begin{cases} \theta \in R^{n+1} \\ x \in R^{n+1} \end{cases}$

Notation

Size (feet^2)	Number of bedrooms	Number of floors	Age of home (years)	Price ($1000)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
...	...	...	...	...

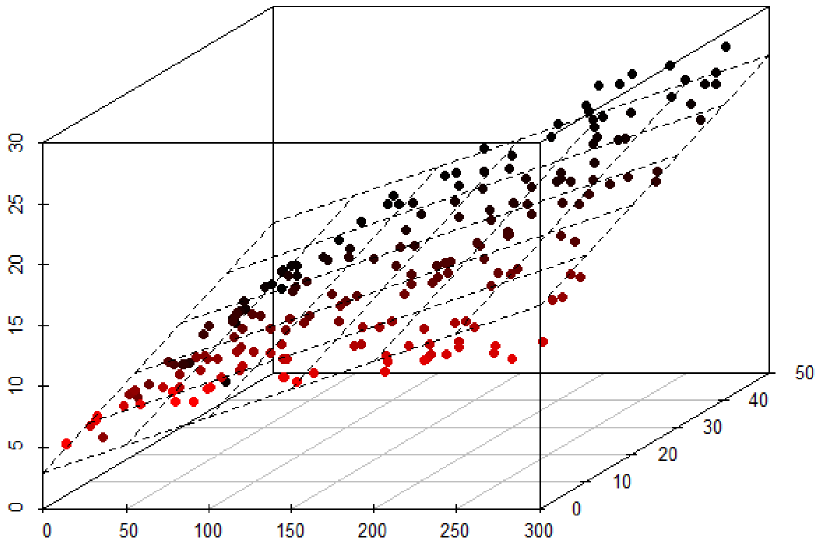
Notation:

n = number of features

$x^{(i)}$ = input (features) of i^{th} training example.

$x_j^{(i)}$ = value of feature j in i^{th} training example.

Multi-dimensional Linear Regression



From one dimension to 2 dimensional case:

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

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

$$h(x) = \sum_{i=0}^{n} \theta_i x_i = \theta^T x$$

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

Error Function