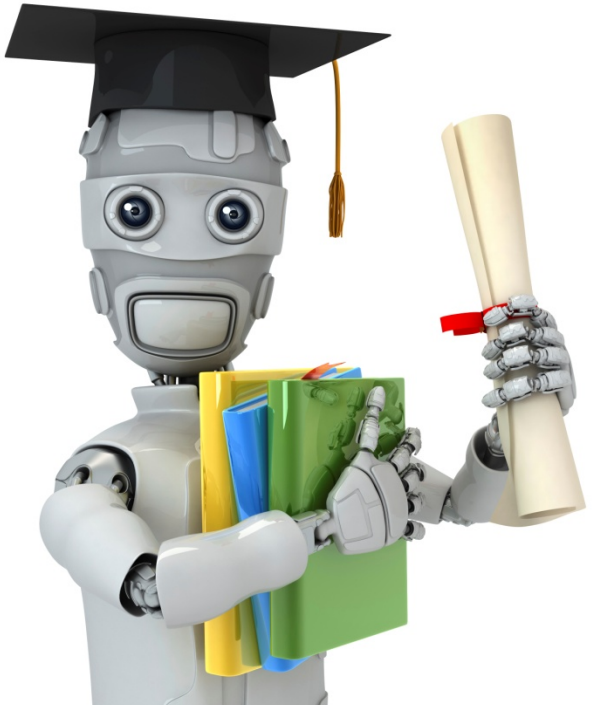


Machine Learning

Linear Regression with multiple variables

Gradient descent in practice I: Feature Scaling

Feature Scaling

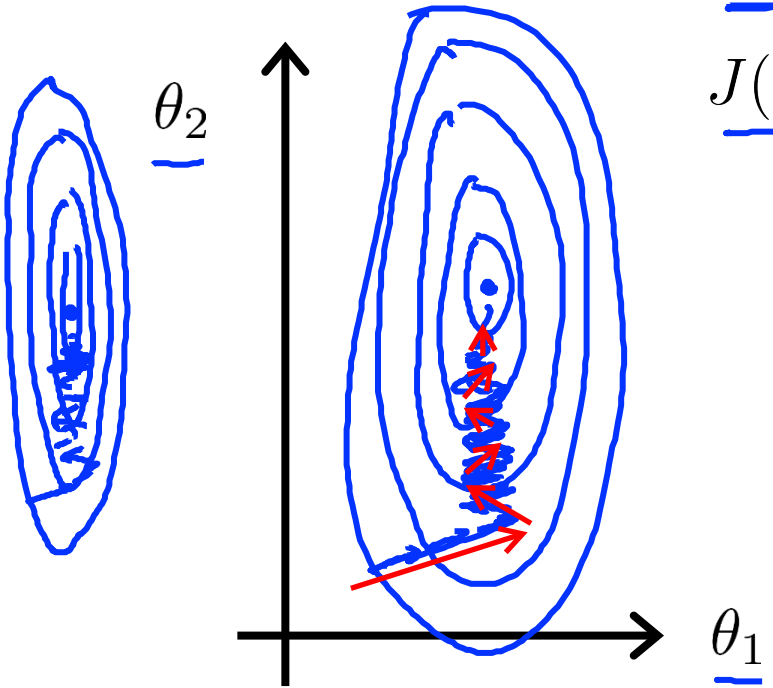
ao mudarmos a escala o GD converge + rapido

with similar range values GD can converge more quickly

Idea: Make sure features are on a similar scale.

E.g. x_1 = size (0-2000 feet²)

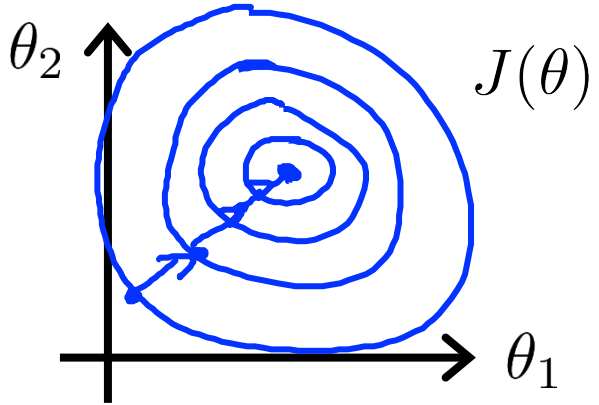
x_2 = number of bedrooms (1-5)



$$x_1 = \frac{\text{size (feet}^2\text{)}}{2000}$$

$$x_2 = \frac{\text{number of bedrooms}}{5}$$

$0 \le x_1 \le 1$ $\quad 0 \le x_2 \le 1$



Feature Scaling

Get every feature into approximately a $-1 \leq x_i \leq 1$ range.

$x_0 = 1$

$0 \leq x_1 \leq 3$ ✓

$-2 \leq x_2 \leq 0.5$ ✓

$-100 \leq x_3 \leq 100$ ✗

$-0.0001 \leq x_4 \leq 0.0001$ ✗

-3 to 3 ✓

$-\frac{1}{3}$ to $\frac{1}{3}$ ✓

Mean normalization

Replace x_i with $x_i - \mu_i$ to make features have approximately zero mean (Do not apply to $x_0 = 1$).

E.g. $x_1 = \frac{size - 1000}{2000}$

$x_2 = \frac{\#bedrooms - 2}{5}$ 4

Average size = 100

1-5 bedrooms

$-0.5 \leq x_1 \leq 0.5$, $-0.5 \leq x_2 \leq 0.5$

$x_1 \leftarrow \frac{x_1 - \mu_1}{s_1}$ $\quad$ $x_2 \leftarrow \frac{x_2 - \mu_1}{s_2}$

μ_1: average value of x in training set

s_1: range (max - min) (or standard deviation)