Multi feature variables.

size. NO. bedroom NO. years → price ?

n → number of features.

$x^{(i)}$ — ith. sample (row)

$x_j^{(i)}$ — feature j in sample i.

Hypothesis.

m — total NO. of training samples.

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots \dots \theta_n x_n.$$

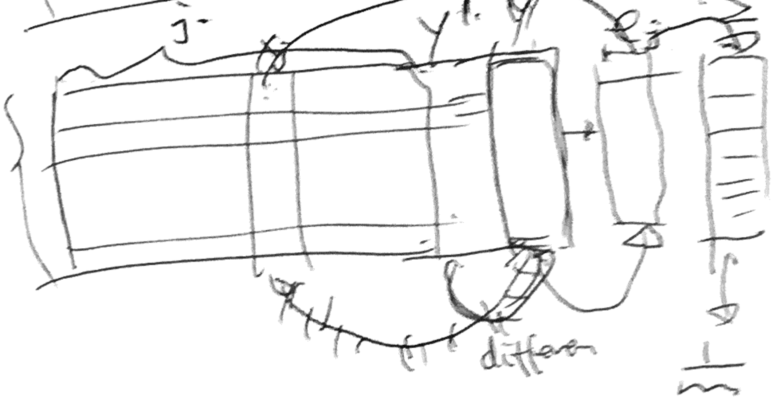
$x_0 = 1.$

$$X = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}$$

$$h_\theta(x) = \theta^T x.$$

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

$$\boxed{\theta_j = \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j} \quad \text{for all } j}$$



Gradient.

$$\theta_j = \theta_j - \alpha \sum_{i=0}^{m} e^i x_j^i$$

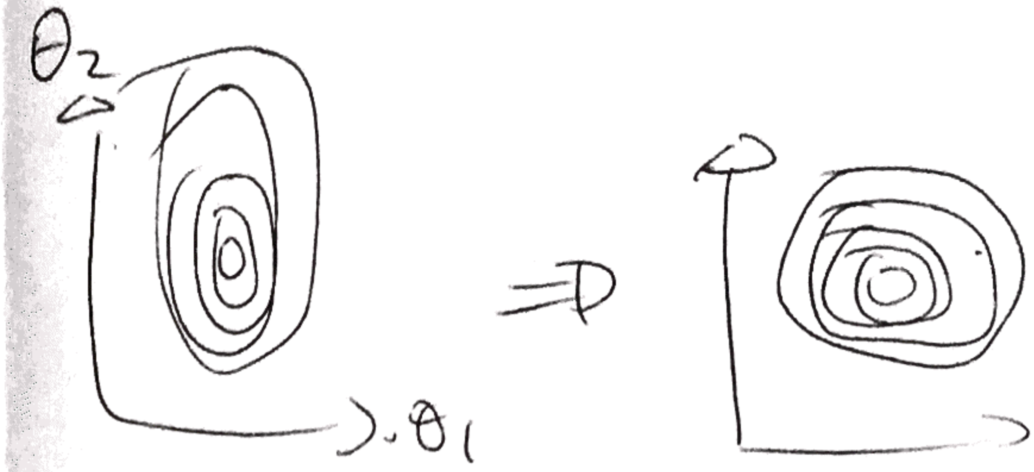
what's multi-variable hyposis function?
what's cost function?
How to update θ_j — feature weight for feature j?

Feature Scaling.

x_1 = size feets.

x_2 = No of bedrooms. 1–5.

skewed cost function.



$-1 \le x_i \le 1$

3 ✓

-3

-100   100 ✗

-0.0001   0.0001 ✗

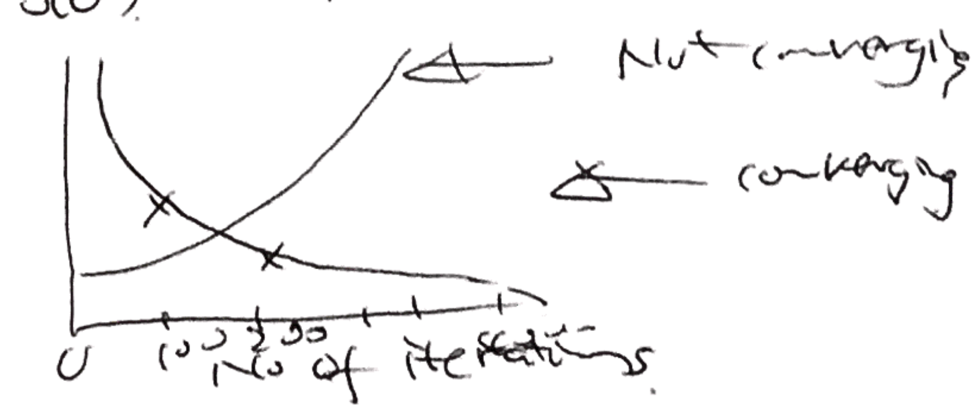
Mean Normalization.

$$x_i = \frac{x_i - \bar{x}_i}{\text{Range}(x_i)}$$ (Not for $x_0 = 1$).

(Range = $X_{max} - X_{min}$)
(or standard deviation)

Learning Rate.

$J(\theta)$ cost function



use smaller α

slower converge.   fast converge or not converge

α: smaller ← → bigger.

0.001   0.003   0.01   0.03

0.1   0.3   1.