Linear Regression with Multiple Variables

- Multiple features

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

- Gradient descent for multiple variables

Hypothesis:

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

Cost function:

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

Gradient descent:

Repeat {

$$\theta_j := \theta_j - \alpha\frac{\partial J(\theta)}{\partial \theta_j}$$

$$= \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$$

}

- Feature scaling

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

Replace x_i with $x_i - \mu_i$ to make features have approximately zero mean.

- Learning rate

How to choose learning rate α: Declare convergence if $J(\theta)$ decreases by less than 10^{-3} in one iteration.