# Week 2.1 Multivariate Linear Regression

## Multiple Features

- 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
- Hypothesis:

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

- for convenience of notation, define $x_0 = 1$

$$X = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}$$

$$\Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \cdots \\ \theta_n \end{bmatrix}$$

$$h_\theta(x) = \Theta^T X$$

## Gradient Descent for Multiple Variables

- Cost Function:

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

- Gradient descent:
  Repeat {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \ldots, \theta_n)$$

  } simultaneously

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

# Gradient Descent in Practice I - Feature Scaling

- **Idea: Make sure features are on a similar scale**
- E.g. More like ellipses
  - $x_1$: size (0-2000 square feet)
  - $x_2$: number of bedrooms (1-5)
- More like circles, converge much faster
  - $x_1$: size/2000
  - $x_2$: number of bedrooms/5
- **Mean normalization**
- Replace $x_i$ with $x_i - \mu_i$ (except $x_0$)

# Gradient Descent in Practice II - Learning Rate

- Making sure gradient descent is working correctly ($J(\theta)$ should decrease after every iteration)
- Example automatic convergence test:
  - Declare convergence if $J(\theta)$ decreases by less than $10^{-3}$ in one iteration
- Use smaller $\alpha$

# Features and Polynomial Regression

## Housing prices prediction

- $h_\theta(x) = \theta_0 + \theta_1 * frontage + \theta_2 * depth$
- maybe $area = frontage * depth$ matters

## Polynomial Regression

- cubic model: $h_\theta(x) = \theta_0 + \theta_1 * size + \theta_2 * size^2 + \theta_3 * size^3$
- $x_1 = (size)$
- $x_2 = (size)^2$
- $x_3 = (size)^3$
- choice of features (cubes come back down):
- $h_\theta(x) = \theta_0 + \theta_1 * size + \theta_2 * size^2$
- $h_\theta(x) = \theta_0 + \theta_1 * size + \theta_2 \sqrt{size}$