Multivariate Linear Regression

Hypothesis: $h_\theta(x) = \theta^T x = \theta_0 x_0 + \theta_1 x_1 \cdots \theta_n x_n$ $\quad x_0 = 1$

Parameters: θ, an $n+1$ dimensional vector

Gradient Descent
New Algorithm: $(n \geq 1)$

$$\theta_j := \theta_j - \alpha \underbrace{\frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x_j^{(i)}}_{\frac{\partial}{\partial \theta_j}\left(J(\theta)\right)}$$

Notation

$x_j^{(i)}$: value of feature j, in the i^{th} sample

$x^{(i)}$: column vector for i^{th} sample

m = # of training samples

n = # of features

Feature Scaling

Idea: Make sure features are on a similar scale

E.G. x_1 = Size (0 - 2,000 ft²) $\Longrightarrow x_1 = \frac{\text{Size}}{2{,}000}$

x_2 = # of beds (1-5) $\Longrightarrow x_2 = \#/5$

$0 \leq x_1, x_2 \leq 1$

General Rule: Get every feature into: $-1 \leq x_i \leq 1$ (or close)

This will help gradient descent converge!

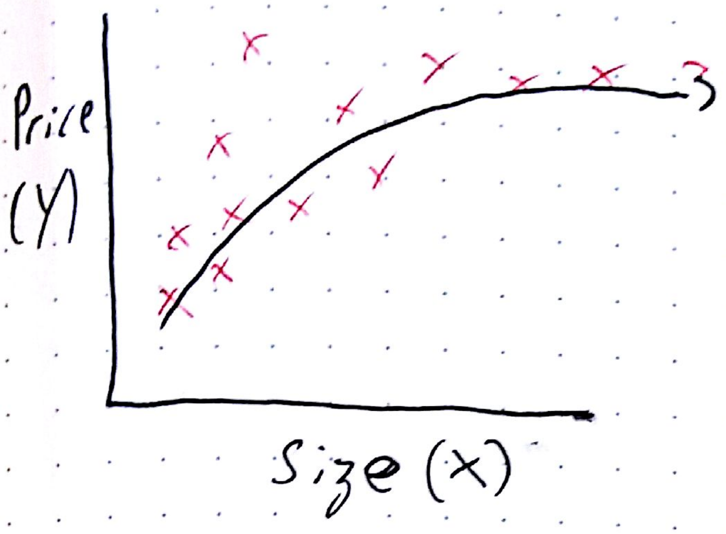
Mean Normalization

$x_1 = \frac{\text{Size} - 1000}{2{,}000}$

$x_2 = \frac{\#\text{beds} - 2}{5}$ [max - min]

Replace x_i with $x_i - \mu_i$ to have about zero mean.

Features & Polynomial Regression



$$\theta_0 + \theta_1 x + \theta_2 x^2$$

or

$$\theta_0 + \theta_1 x + \theta_2 \sqrt{x}$$

Normal Equation

Example: $m = 4$

x_0	Size	# of Beds	# of floors	Age home	Price
1	2104	5	1	45	460
1	1416	3	2	40	232
1	1534	3	2	30	315
1	852	2	1	36	178

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \end{bmatrix} \qquad y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$$

$m \times (n+1)$

m-dimensional Vector

$$\Rightarrow \theta = (X^T X)^{-1} X^T y$$