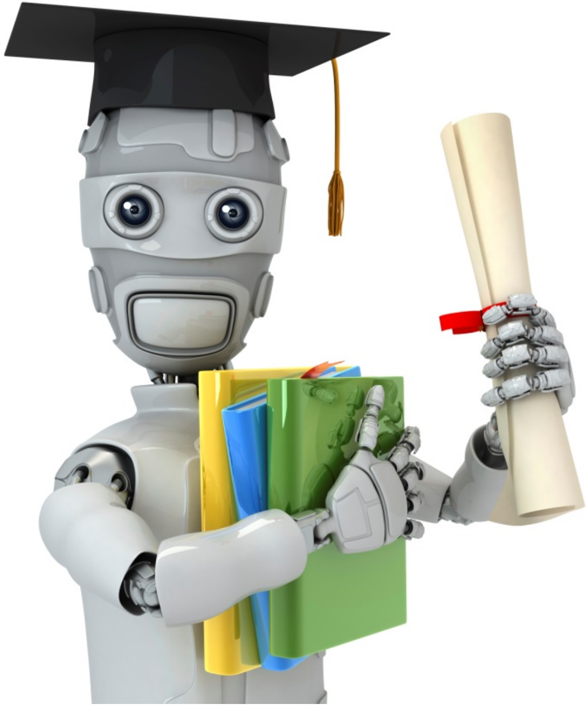


Machine Learning

Linear Regression with multiple variables

Multiple features

Multiple features (variables).

Size (feet^2)	Price (\$1000)
x	y
2104	460
1416	232
1534	315
852	178
…	…

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Multiple features (variables).

Size (feet^2)	Number of bedrooms	Number of floors	Age of home (years)	Price (\$1000)
x_1	x_2	x_3	x_4	y
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
…	…	…	…	…

Row ~~Feature~~ Vector

학습데이터 개수

$m = 47$

Column Vector

$$x^{(2)} = \begin{bmatrix} 1416 \\ 3 \\ 2 \\ 40 \end{bmatrix}$$

$x^{(2)}_3 = 2$

필요하면 Transpose로 Row화

Notation:

주어진 입력변수의 개수

- n = number of features

$n = 4$ ≒ dimension or attribute

- $x^{(i)}$ = input (features) of i^{th} training example.
- $x_j^{(i)}$ = value of feature j in i^{th} training example.

Andrew Ng

Hypothesis:

Previously: $h_\theta(x) = \cancel{\theta_0 + \theta_1 x}$

상수 — 각각의 변수에 대한 가중치

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$$

E.g. $h_\theta(x) = 80 + 0.1 x_1 + 0.01 x_2 + 3 x_3 - 2 x_4$

age

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

피처갯수가 많은 시나리오

For convenience of notation, define $\boxed{x_0 = 1.}$ $\quad (x_0^{(i)} = 1)$

$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n+1} \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \in \mathbb{R}^{n+1}$$

$$\underbrace{\begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_n \end{bmatrix}}_{\theta^T} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$(n+1) \times 1$ matrix $\qquad x$

$\theta^T x$

$$h_\theta(x) = \theta_0 \underbrace{x_0}_{=1} + \theta_1 x_1 + \cdots + \theta_n x_n$$

$$= \boxed{\theta^T x}$$

내적의 곱

Multivariate linear regression.

Multiple features (variables).

$\dfrac{x_1 - 852}{1252}$

Size (feet^2)	Number of bedrooms	Number of floors	Age of home (years)	Price (\$1000)
x_1	x_2	x_3	x_4	y
2104	5	1	45	460
$\boxed{1416}$	$\boxed{3}$	$\boxed{2}$	$\boxed{40}$	232
1534	3	2	30	315
852	2	1	36	178
…	…	…	…	…

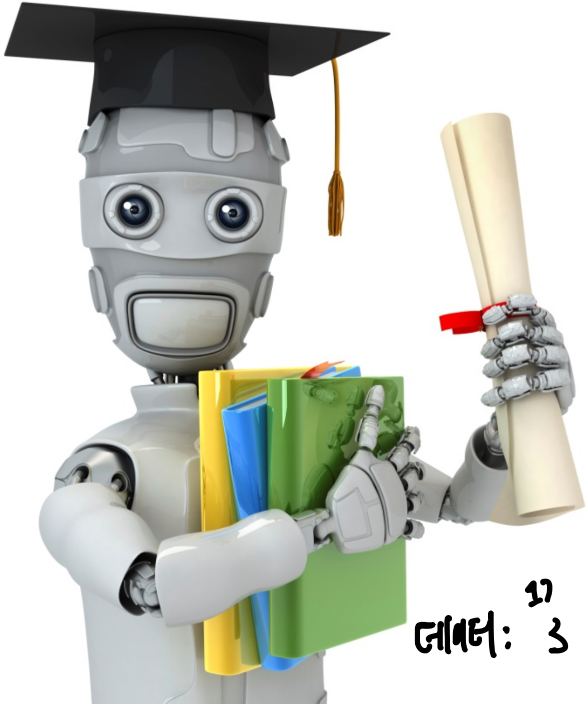
$m = 47$

Notation:

- n = number of features $n = 4$
- $x^{(i)}$ = input (features) of i^{th} training example.
- $x_j^{(i)}$ = value of feature $\underline{j}$ in $\underline{i}^{th}$ training example.

$$x^{(2)} = \begin{bmatrix} 1416 \\ 3 \\ 2 \\ 40 \end{bmatrix}$$

$$x_3^{(2)} = 2$$



Linear Regression with multiple variables

Gradient descent in practice I: Feature Scaling

Machine Learning

데이터: 1) 3 2) $-1 \to 0$ 3) $7 \to 1$ $0 \sim 1$ Norm.

$y = ax + b$

$0 = a \cdot -1 + b$

$1 = a \cdot 7 + b$

$1 = 8a$

$a = \frac{1}{8}$

$b = \frac{1}{8}$

Feature Scaling

— 각각의 단위가 다르다! m² 수 방갯수 Range

• Outlier에 취약함

Idea: Make sure features are on a similar scale.

E.g. x_1 = size (0-2000 feet²) ← 1000 단위에서 논다

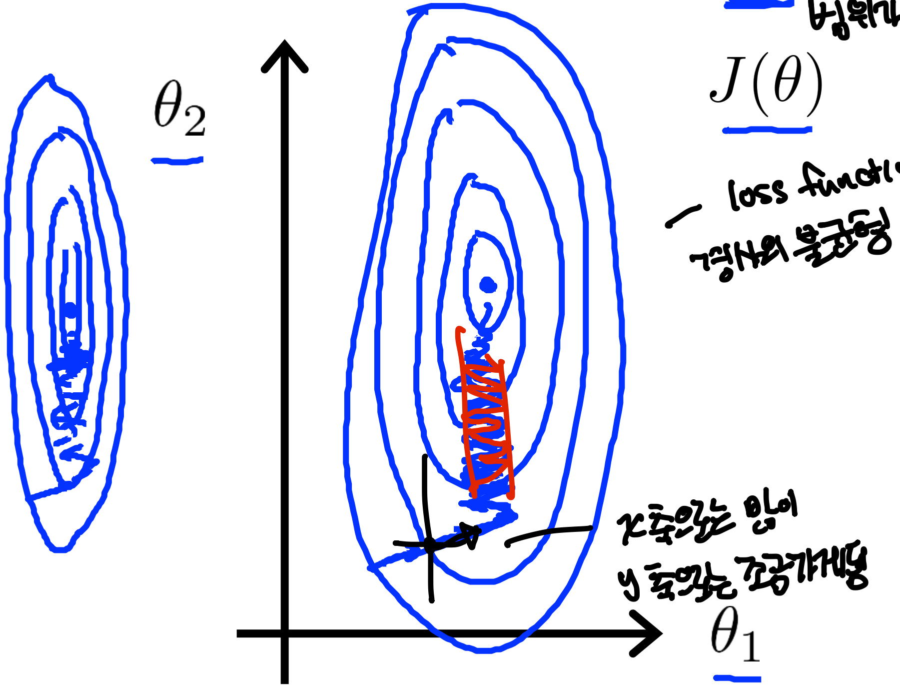
x_2 = number of bedrooms (1-5) ← 가중치값이 비슷한 영향 끼치려면 범위가 훨씬 크다

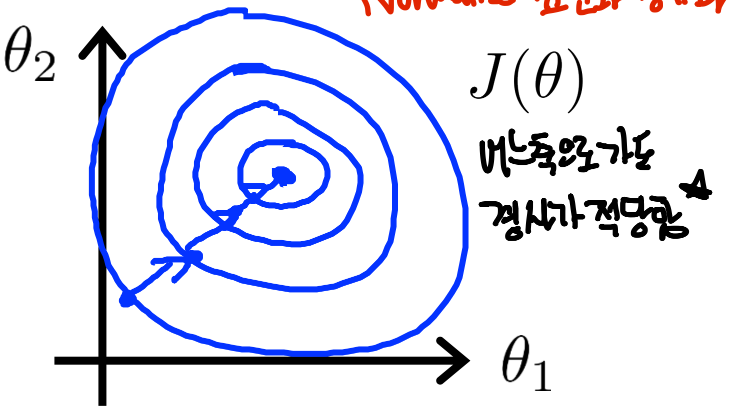
$$x_1 = \frac{\text{size (feet}^2\text{)}}{2000}$$ ~ 최대값으로 나눈다

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

비슷한자리수로 변환

$$0 \le x_1 \le 1 \sim \underbrace{0 \le x_2 \le 1}_{\text{Normalize 표준화 정규화}}$$





Feature Scaling

평균 2000
분산 600^2
x

$\frac{x-2000}{600}$ → 평균 : 0, 분산 : 1

e.g. $(1800-2000)^2+(2200-2000)^2+\cdots+\cdots$ ~ 가우시안분포

최대값 및 최소값 고려

→ 평균과 분산에 의한 Normalization 더 많이 사용

Get every feature into approximately a $\boxed{-1 \le x_i \le 1}$ range.

$x_0 = 1$

$0 \le x_1 \le 3$ ✓

$-2 \le x_2 \le 0.5$ ✓

$-100 \le x_3 \; \boxed{100}$ ✗

$-0.0001 \le x_4 \le \boxed{0.0001}$ ✗

-3 to 3 ✓

$-\frac{1}{3}$ to $\frac{1}{3}$ ✓

$x_0 = 1$

Hypothesis: $h_\theta(x) = \theta^T x = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$

Parameters: $\cancel{\theta_0, \theta_1, \ldots, \theta_n}$ $\quad \theta$ $\quad$ n+1 - dimensional vector

Cost function:

$$\cancel{J(\theta_0, \theta_1, \ldots, \theta_n)} = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$J(\theta)$

x_1	x_2	y
1	2	5
3	-1	7

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 = \theta^T x = [\theta_0 \theta_1 \theta_2] \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix}$

$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} \quad x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}$

234 $\rightarrow$ $J(\theta_0 \theta_1 \theta_2) = \frac{1}{2 \cdot 3} \left((\theta_0 \cdot 1 + \theta_1 \cdot 1 + \theta_2 \cdot 2 - 5)^2 + (\theta_0 \cdot 1 + \theta_1 \cdot 3 + \theta_2 \cdot (-1) - 7)^2 \ldots\right.$

Gradient descent:

Repeat $\{$

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

$\}$

$\frac{dJ(\theta_0, \theta_1, \theta_2)}{d\theta_0} = \frac{1}{3}\left((\theta_0 \cdot 1 + \theta_1 \cdot 1 + \theta_2 \cdot 2 - 5) \cdot 1 \ldots\right.$

(simultaneously update for every $j = 0, \ldots, n$)

Gradient Descent

Previously (n=1):

Repeat {

$$\theta_0 := \theta_0 - \alpha \underbrace{\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})}_{\frac{\partial}{\partial\theta_0}J(\theta)}$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x^{(i)}$$

$x^{(i)}$

(simultaneously update θ_0, θ_1)

}

New algorithm $(n \geq 1)$: 피처갯수 1+

Repeat {

$\frac{\partial}{\partial\theta_j} J(\theta)$

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

(simultaneously update θ_j for $j = 0, \ldots, n$)

}

$x_0^{(i)} = 1$

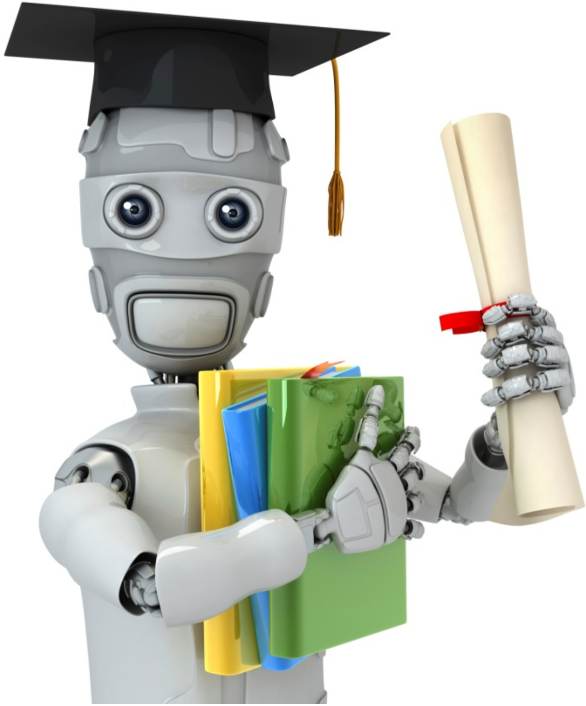
0.1

$$\theta_0 := \theta_0 - \alpha \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_0^{(i)}$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_1^{(i)}$$

$$\theta_2 := \theta_2 - \alpha \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_2^{(i)}$$

...



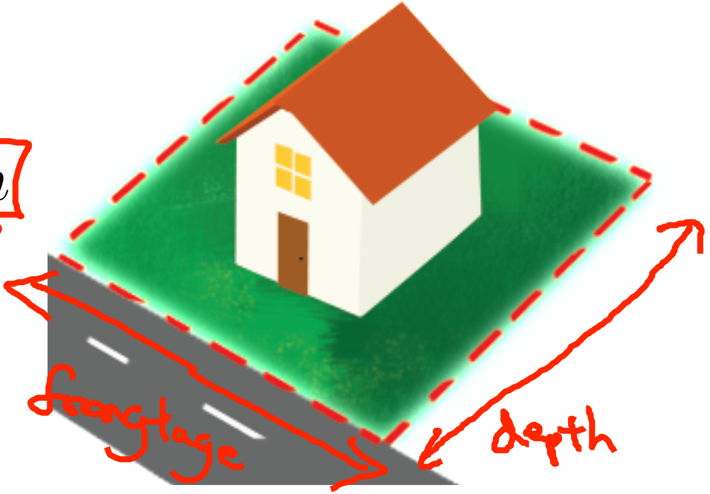
Machine Learning

Linear Regression with multiple variables

Features and polynomial regression

Housing prices prediction

$$h_\theta(x) = \theta_0 + \theta_1 \times \underbrace{\boxed{frontage}}_{\underline{x_1}} + \theta_2 \times \underbrace{\boxed{depth}}_{\underline{x_2}}$$



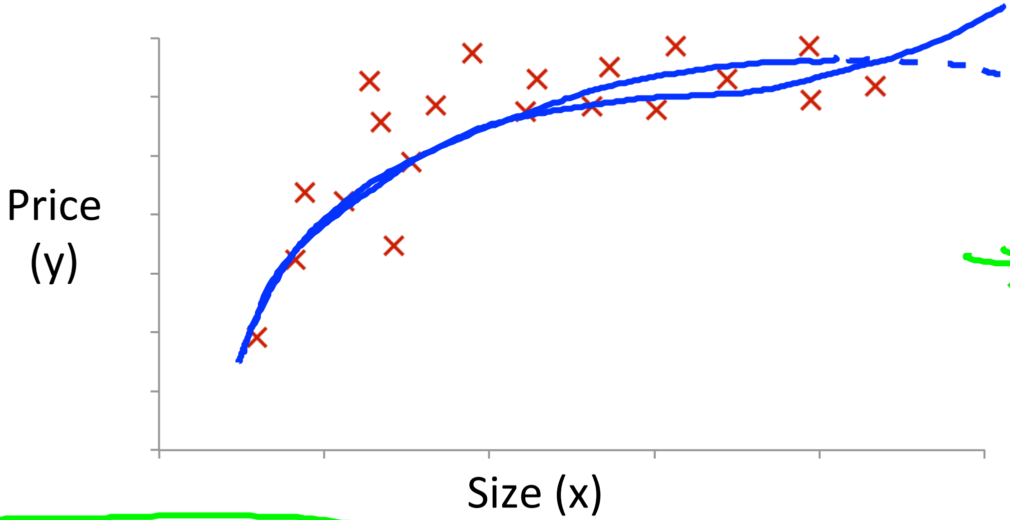
Area

$$x = \underline{frontage * depth}$$

$$h_\theta(x) = \theta_0 + \theta_1 x$$

land area

Polynomial regression



2차식 [illegible]

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

$\rightarrow \boxed{\rightarrow \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3}$

$$\boxed{h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3}$$
$$= \theta_0 + \theta_1 (size) + \theta_2 (size)^2 + \theta_3 (size)^3$$

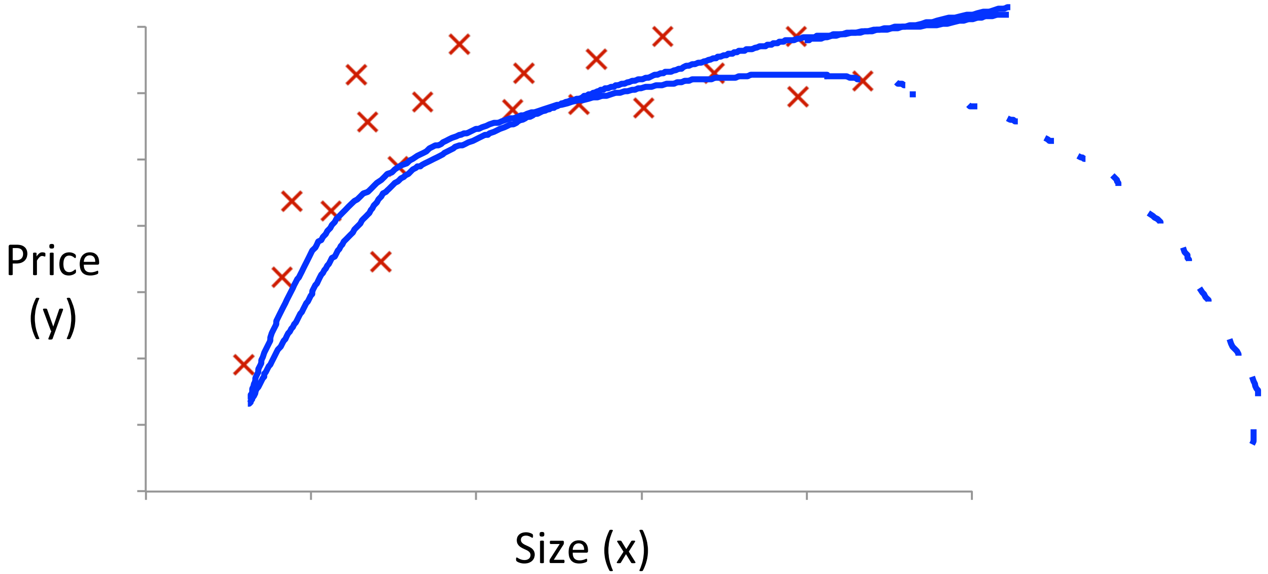
$$\left.\begin{aligned} \rightarrow x_1 &= (size) \\ \rightarrow x_2 &= (size)^2 \\ \rightarrow x_3 &= (size)^3 \end{aligned}\right\}$$

Size: 1 - 1000

size²: 1 - 1,000,000

Size³: 1 - 10⁹

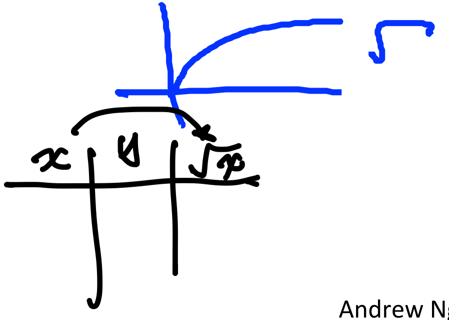
Choice of features



$$\rightarrow \quad h_\theta(x) = \theta_0 + \theta_1(size) + \theta_2(size)^2$$

[illegible] ✓

$$\rightarrow \quad \boxed{h_\theta(x) = \theta_0 + \theta_1(size) + \theta_2\sqrt{(size)}}$$



Andrew Ng