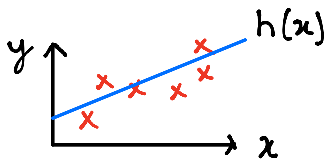
$h_\theta(x) = \underbrace{\theta_0 + \theta_1 x}_{\text{linear ex.}}$ θ_i : Parameter

Linear regression w/ 1 variable



minimize θ_0, θ_1

$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$

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

$J(\theta_0,\theta_1)$ - cost function
squared error function

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

Parameters : θ_0, θ_1

Cost Function :

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

Goal: $\underset{\theta_0,\theta_1}{\text{minimize}}\ J(\theta_0,\theta_1)$

Gradient Descent (minimize J)

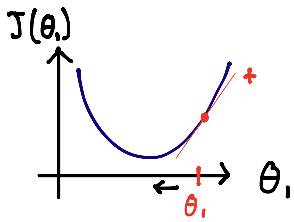
"Batch"
uses all training examples

repeat until convergence {

$$\theta_j := \theta_j - \underset{\text{learning rate}}{\alpha} \underbrace{\frac{\partial}{\partial\theta_j} J(\theta_0,\theta_1)}_{\text{derivative}}$$

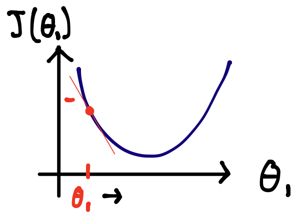
}

$$\left.\begin{aligned}
\text{temp0} &:= \theta_0 - \alpha\frac{\partial}{\partial\theta_0}J(\theta_0,\theta_1)\\
\text{temp1} &:= \theta_1 - \alpha\frac{\partial}{\partial\theta_1}J(\theta_0,\theta_1)\\
\theta_0 &:= \text{temp0}\\
\theta_1 &:= \text{temp1}
\end{aligned}\right\}\ \text{update simultaneously}$$



$$\theta_1 := \theta_1 - \alpha \underbrace{\frac{d}{d\theta_1} J(\theta_1)}_{+\text{ slope}}$$

$$\theta_1 := \theta_1 - \alpha (+)$$



$$\frac{d}{d\theta} J(\theta_1) \leq \theta$$

$$\theta_1 := \theta_1 - \alpha (-)$$

Small α $\rightarrow$ too slow

Large α $\rightarrow$ overshoot min. & may fail to converge or even diverge

---

$$\theta_j := \theta_j - \alpha \underbrace{\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)}, \quad J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$= \frac{\partial}{\partial \theta_j} \left[\frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)})^2 - y^{(i)}\right)^2 \right]$$

$$= \frac{\partial}{\partial \theta_J} \left[\frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^2 \right]$$

$$\theta_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) \Rightarrow \boxed{\theta_0 := \theta_0 - \alpha \cdot \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)}$$

$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) \Rightarrow \boxed{\theta_1 := \theta_1 - \alpha \cdot \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x^{(i)}}$$

Cost Function for linear regression : "convex function"

- only 1 global min