

LECTURE 3

HOW TO MINIMIZE COST

Sung Kim <hunkim+ml@gmail.com>
<http://hunkim.github.io/ml>

Acknowledgement

01. Andrew Ng's ML Class

- <https://class.coursera.org/ml-003/lecture>
- [http://www.holehouse.org/mlclass\(note\)](http://www.holehouse.org/mlclass(note))

02. Convolutional Neural Networks for Visual Recognition

- <http://cs231n.github.io>

03. TensorFlow

- <https://www.tensorflow.org>
- <https://github.com/aymericdamien/TensorFlow-Examples>

Hypothesis and Cost

$$H(x) = W + b$$

$$\text{cost}(W, b) = \frac{1}{m} \sum_{i=1}^m (H(x^{(i)}) - y^{(i)})^2$$

Simplified Hypothesis

$$h(x) = wx$$

$$\text{cost}(W) = \frac{1}{m} \sum_{i=1}^m (wx^{(i)} - y^{(i)})^2$$

What Cost (W) Looks Like?

$$\text{cost}(W) = \frac{1}{m} \sum_{i=1}^m (Wx^{(i)} - y^{(i)})^2$$

X	Y
1	1
2	2
3	3

W=1, cost(W)=?

What Cost (W) Looks Like?

$$\text{cost}(W) = \frac{1}{m} \sum_{i=1}^m (Wx^{(i)} - y^{(i)})^2$$

X	Y
1	1
2	2
3	3

$W=1, \text{cost}(W)=0$

$$\frac{1}{3}((1 \times 1 - 1)^2 + (1 \times 2 - 2)^2 + (1 \times 3 - 3)^2)$$

$W=0, \text{cost}(W)=4.67$

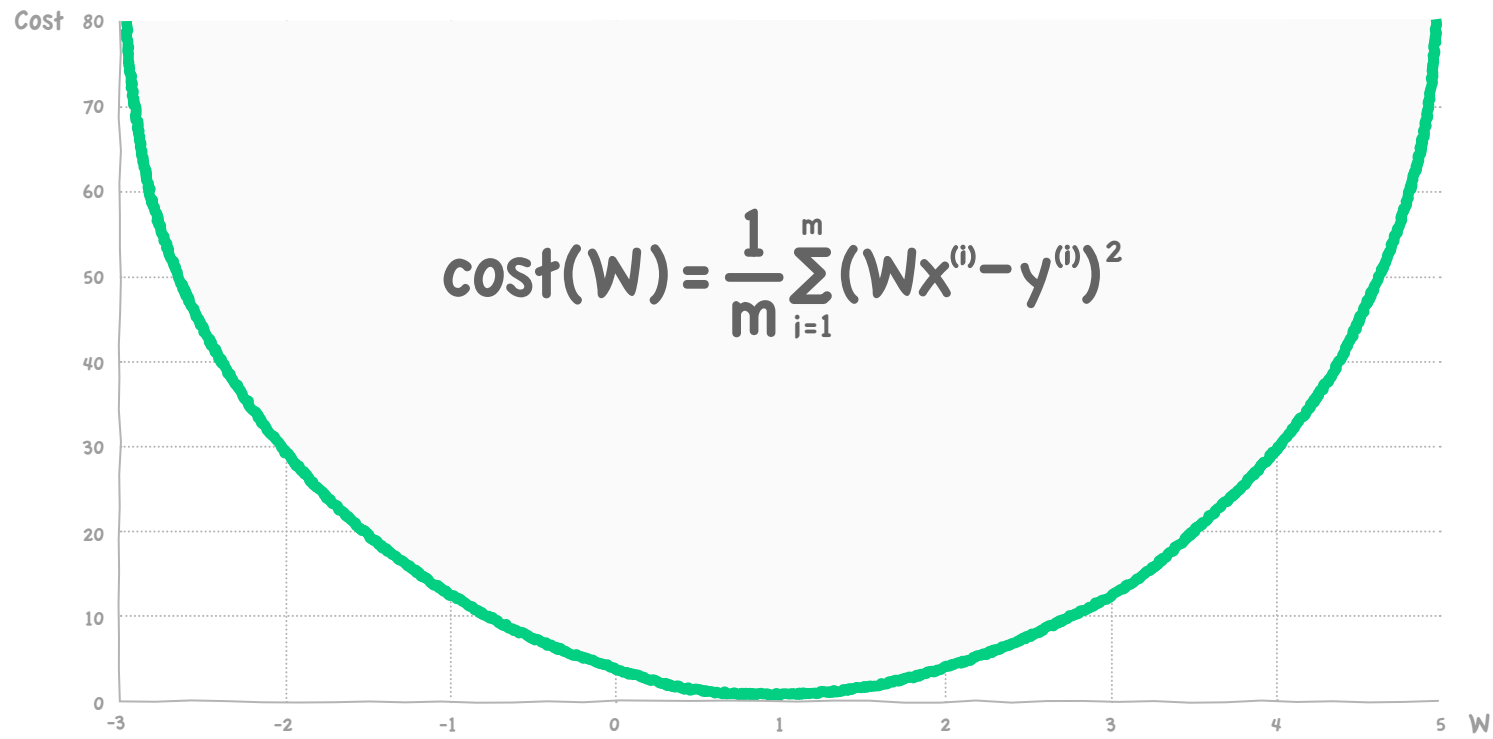
$$\frac{1}{3}((0 \times 1 - 1)^2 + (0 \times 2 - 2)^2 + (0 \times 3 - 3)^2)$$

$W=2, \text{cost}(W)=?$

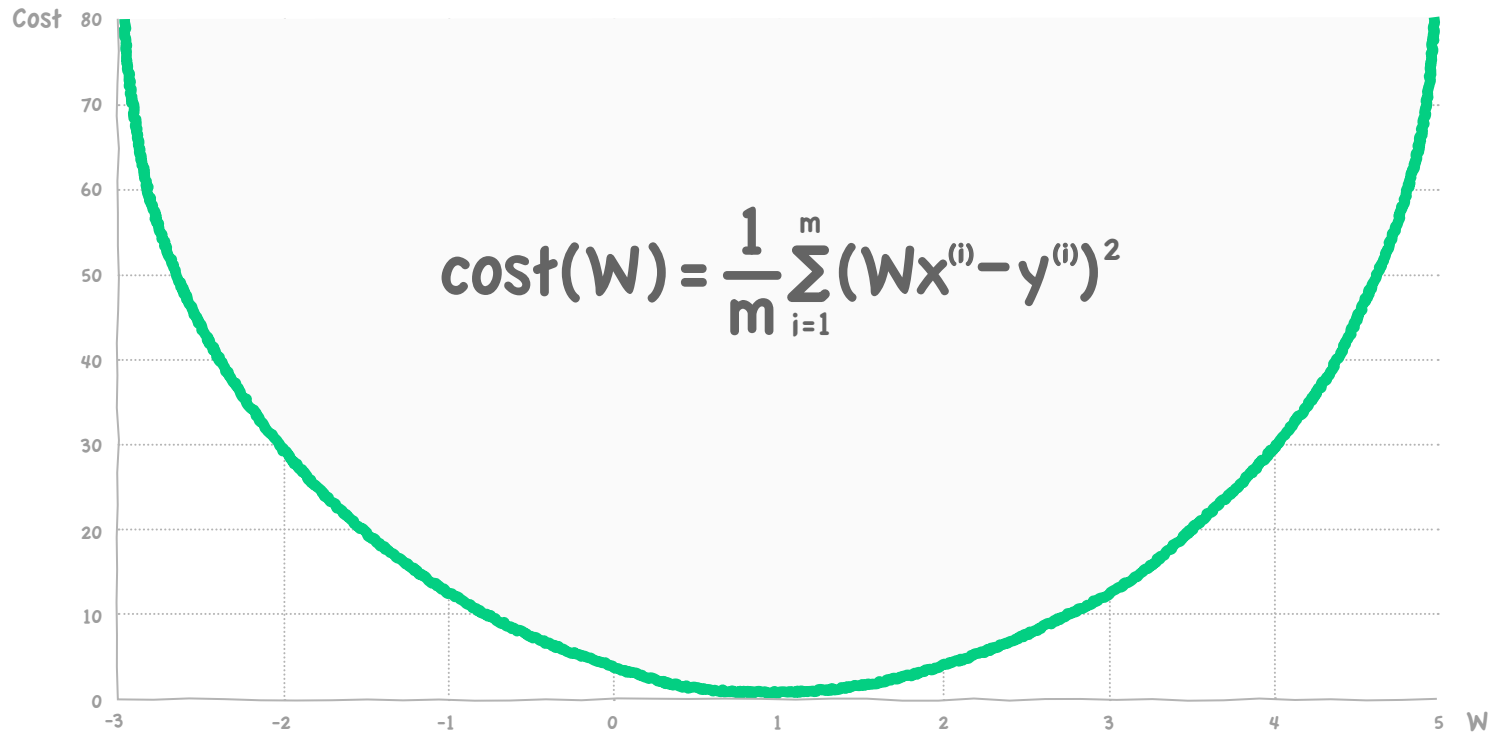
What Cost (W) Looks Like?

- $W=1, \text{cost}(W)=0$
- $W=0, \text{cost}(W)=4.67$
- $W=2, \text{cost}(W)=4.67$

What Cost(W) Looks Like?



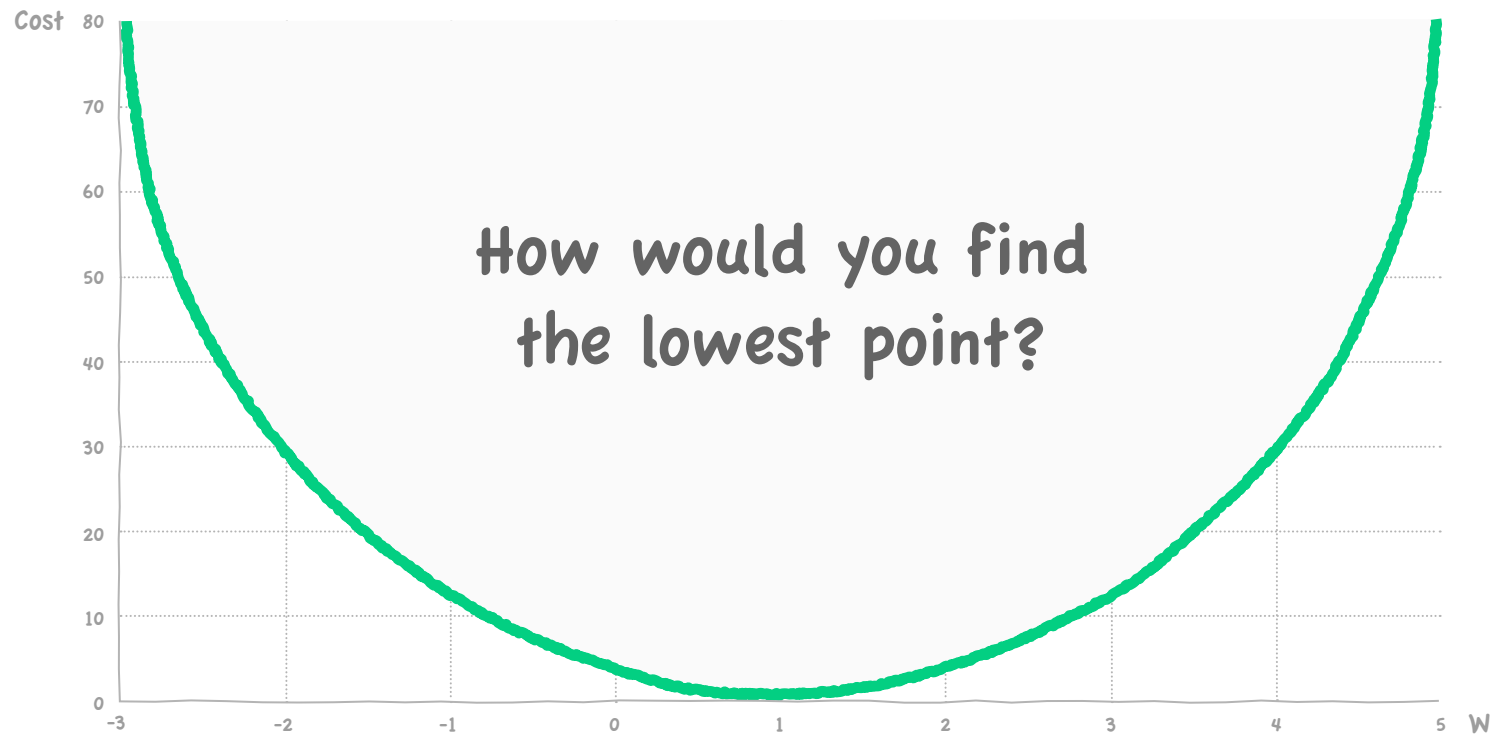
How to Minimize Cost?



Gradient Descent Algorithm

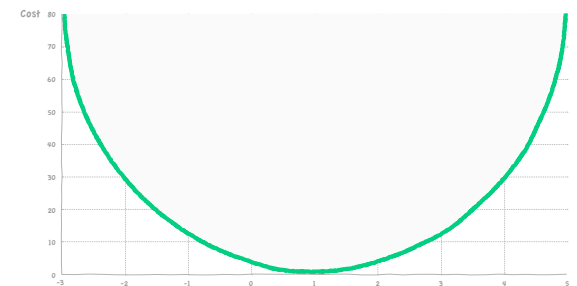
- Minimize cost function
- Gradient descent is used many minimization problems
- For a given cost function, $\text{cost}(W, b)$, it will find W, b to minimize cost
- It can be applied to more general function: $\text{cost}(w_1, w_2, \dots)$

How It Works?



How It Works?

01. Start with initial guesses
 - Start at 0,0 (or any other value)
 - Keeping changing W and b a little bit to try and reduce $\text{cost}(W, b)$
02. Each time you change the parameters, you select the gradient which reduces $\text{cost}(W, b)$ the most possible
03. Repeat
04. Do so until you converge to a local minimum
05. Has an interesting property
 - Where you start can determine which minimum you end up



Formal Definition

$$\text{cost}(W) = \frac{1}{m} \sum_{i=1}^m (Wx^{(i)} - y^{(i)})^2$$



$$\text{cost}(W) = \frac{1}{2m} \sum_{i=1}^m (Wx^{(i)} - y^{(i)})^2$$

Formal Definition

$$\text{cost}(W) = \frac{1}{2m} \sum_{i=1}^m (Wx^{(i)} - y^{(i)})^2$$

$$W := W - \alpha \frac{\partial}{\partial W} \text{cost}(W)$$

Formal Definition

$$W := W - \alpha \frac{\partial}{\partial W} \frac{1}{2m} \sum_{i=1}^m (Wx^{(i)} - y^{(i)})^2$$

$$W := W - \alpha \frac{1}{2m} \sum_{i=1}^m 2(Wx^{(i)} - y^{(i)})x^{(i)}$$

$$W := W - \alpha \frac{1}{m} \sum_{i=1}^m (Wx^{(i)} - y^{(i)})x^{(i)}$$

Partial Derivative Calculator

Partially differentiate functions step-by-step

Limits

► Integrals

▼ Derivatives

- First Derivative
- Second Derivative
- Third Derivative
- Higher Order Derivatives
- Derivative at a point
- **Partial Derivative**
- Implicit Derivative

► Derivative Applications

► Series

► ODE

► Laplace Transform

► Taylor/Maclaurin Series

full pad »

x^2	x^\square	\log_\square	$\sqrt{\square}$	$\sqrt[\square]{\square}$	\leq	\geq	$\frac{\square}{\square}$	\cdot	\div	x°	π
$(\square)'$	$\frac{d}{dx}$	$\frac{\partial}{\partial x}$	\int	\int_\square^\square	\lim	Σ	∞	θ	$(f \circ g)$	H_2O	$\begin{pmatrix} \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \end{pmatrix}$

$$\frac{\partial}{\partial W}(WX - Y)^2$$

Go

Graph » Examples » Practice »

Solution

Show Steps ▼

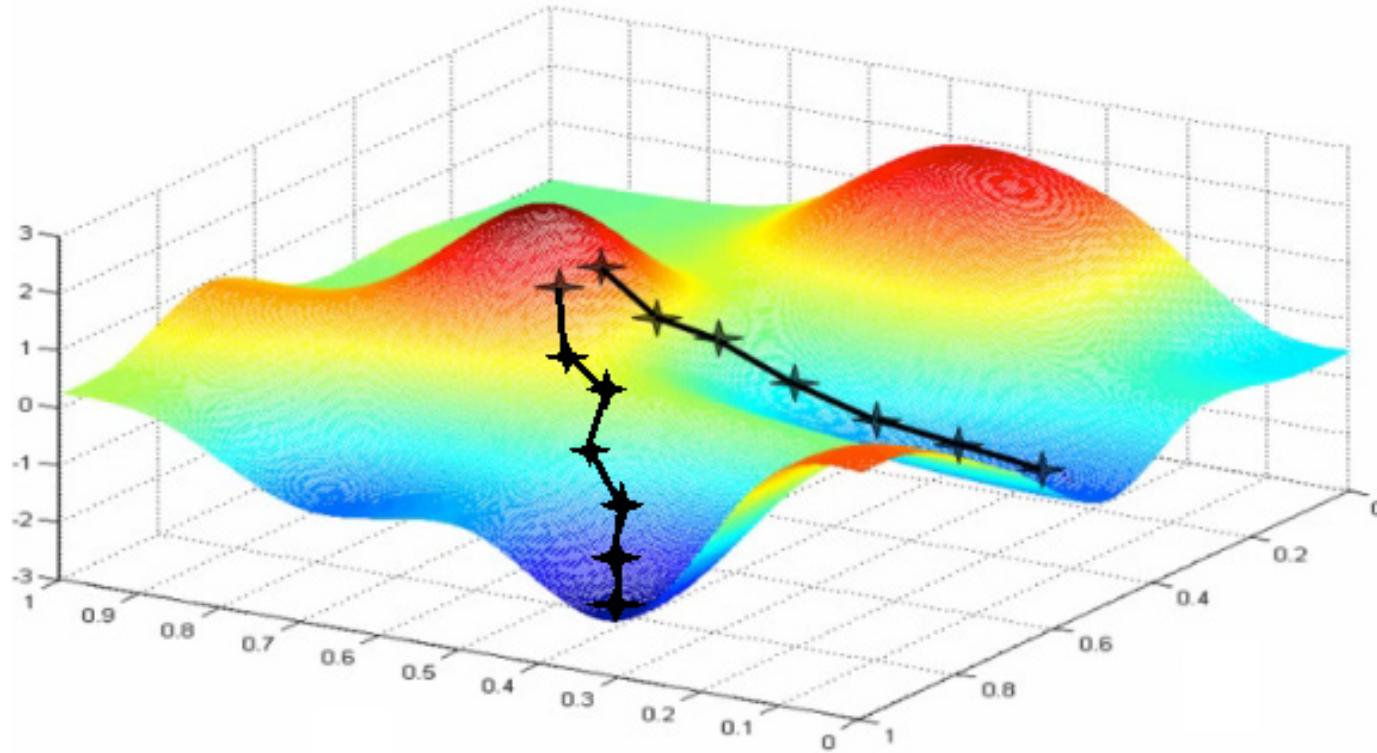
$$\frac{\partial}{\partial W}((WX - Y)^2) = 2(WX - Y)X$$

Gradient Descent Algorithm

$$W := W - \alpha \frac{1}{m} \sum_{i=1}^m (Wx^{(i)} - y^{(i)}) x^{(i)}$$

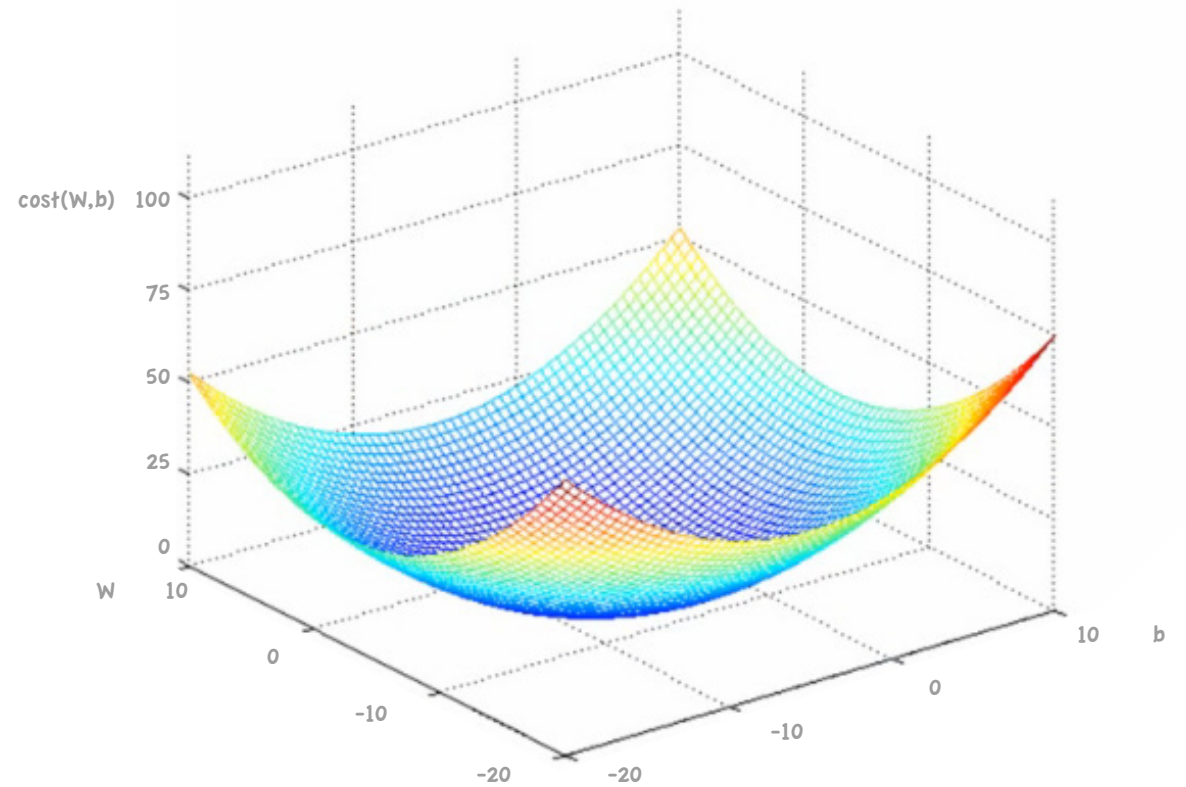


Convex Function



Convex Function

$$\text{cost}(W,b) = \frac{1}{m} \sum_{i=1}^m (\text{H}(x^{(i)}) - y^{(i)})^2$$



NEXT LECTURE

MULTIVARIABLE LOGISTIC REGRESSION