LECTURE 3

HOW TO MINIMIZE COST

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

NAVER | Clova

Acknowledgement

01. Andrew Ng's ML Class
  - https://class.coursera.org/ml-003/lecture
  - 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

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

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

Simplified Hypothesis

$$H(x) = Wx$$

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

What Cost(W) Looks Like?

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

X	Y
1	1
2	2
3	3

$W=1, cost(W)=?$

What Cost(W) Looks Like?

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

X	Y
1	1
2	2
3	3

$W=1, cost(W)=0$

$$\frac{1}{3}((1\times1-1)^2 + (1\times2-2)^2 + (1\times3-3)^2)$$

$W=0, cost(W)=4.67$

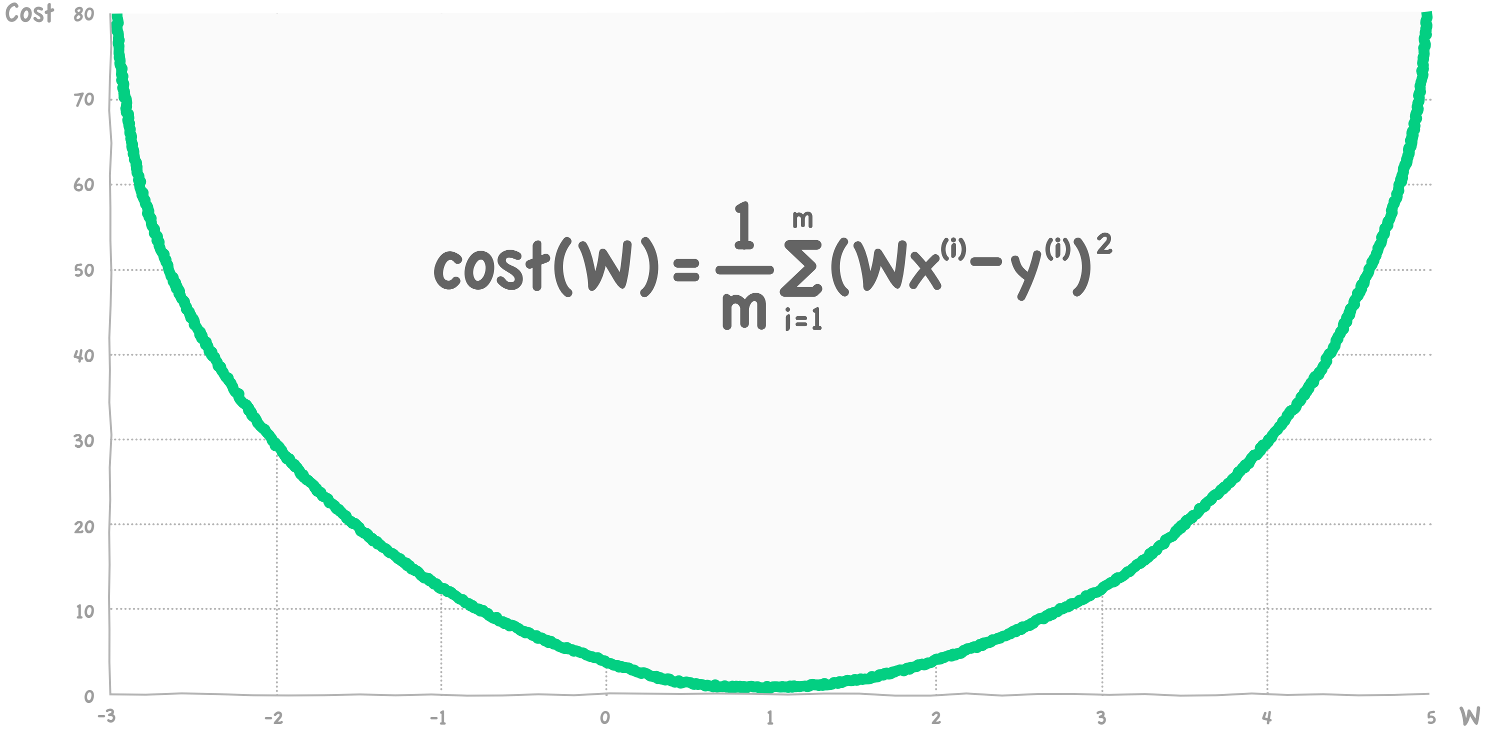
$$\frac{1}{3}((0\times1-1)^2 + (0\times2-2)^2 + (0\times3-3)^2)$$

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

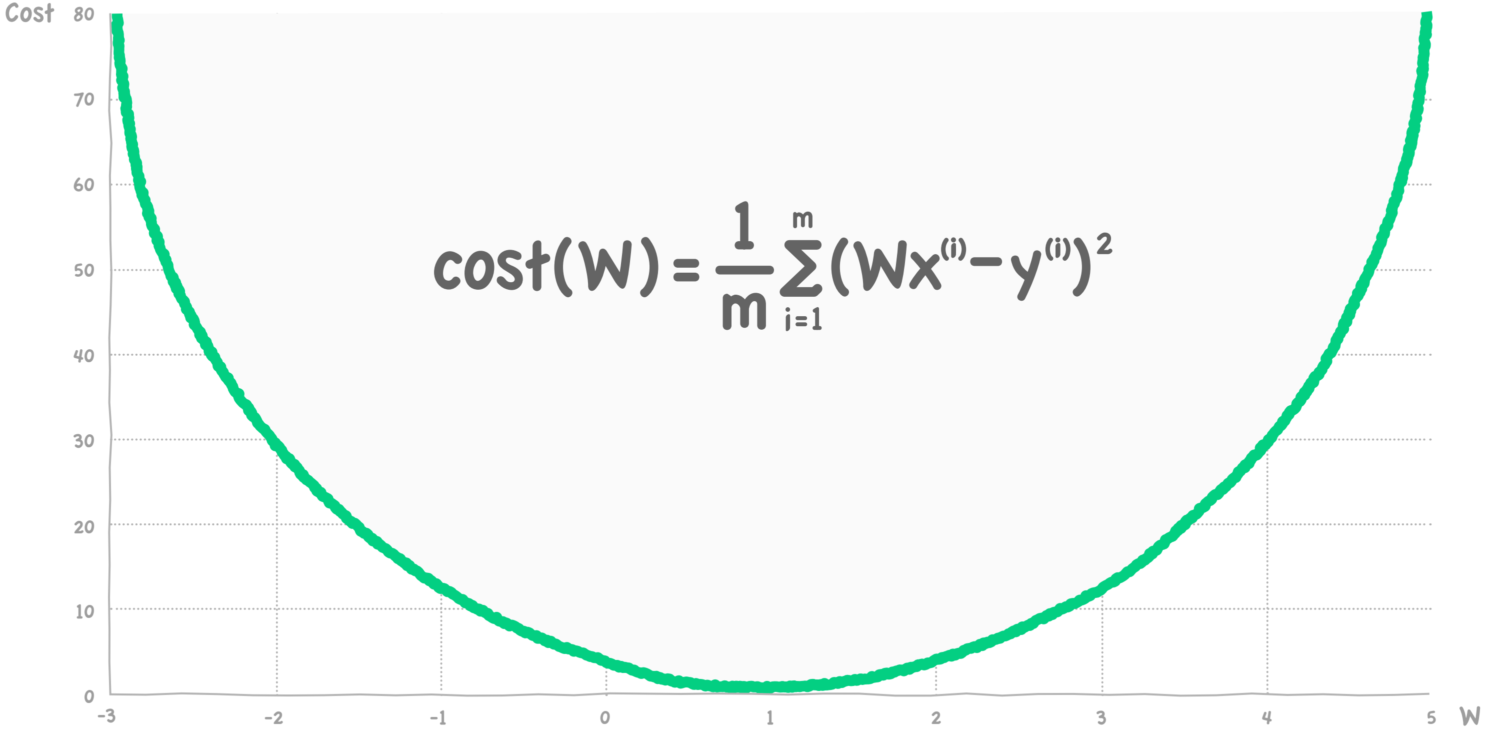
What Cost(W) Looks Like?

- W=1, cost(W)=0
- W=0, cost(W)=4.67
- W=2, cost(W)=4.67

What Cost(W) Looks Like?



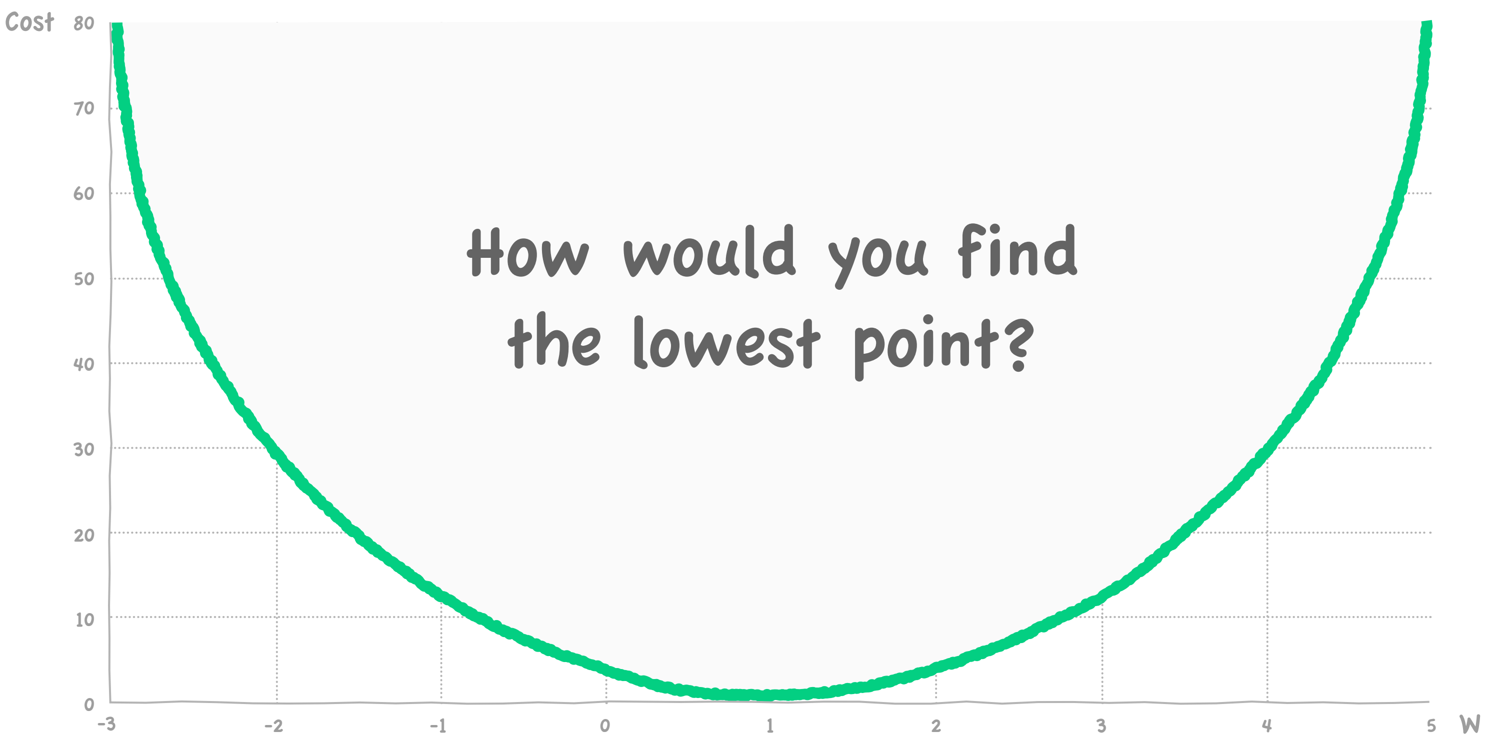
How to Minimize Cost?



Gradient Descent Algorithm

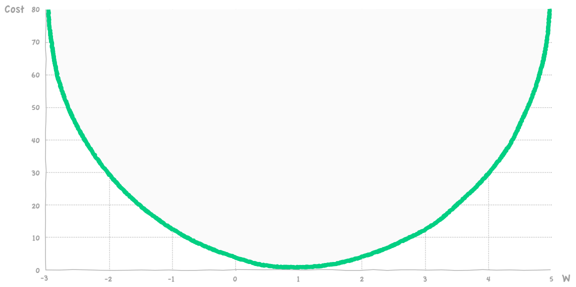
- Minimize cost function
- Gradient descent is used many minimization problems
- For a given cost function, cost(W, b), it will find W,b to minimize cost
- It can be applied to more general function: cost(w1, w2, …)

How It Works?


Cost
80
70
60
50
40
30
20
10
0
-3
-2
-1
0
1
2
3
4
5
W
How would you find
the lowest point?

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 cost(tW,b)
02. Each time you change the parameters, you select the gradient which reduces 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



http://www.holehouse.org/mlclass/01_02_Introduction_regression_analysis_and_gr.html

Formal Definition

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

$$\downarrow$$

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

Formal Definition

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

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

Formal Definition

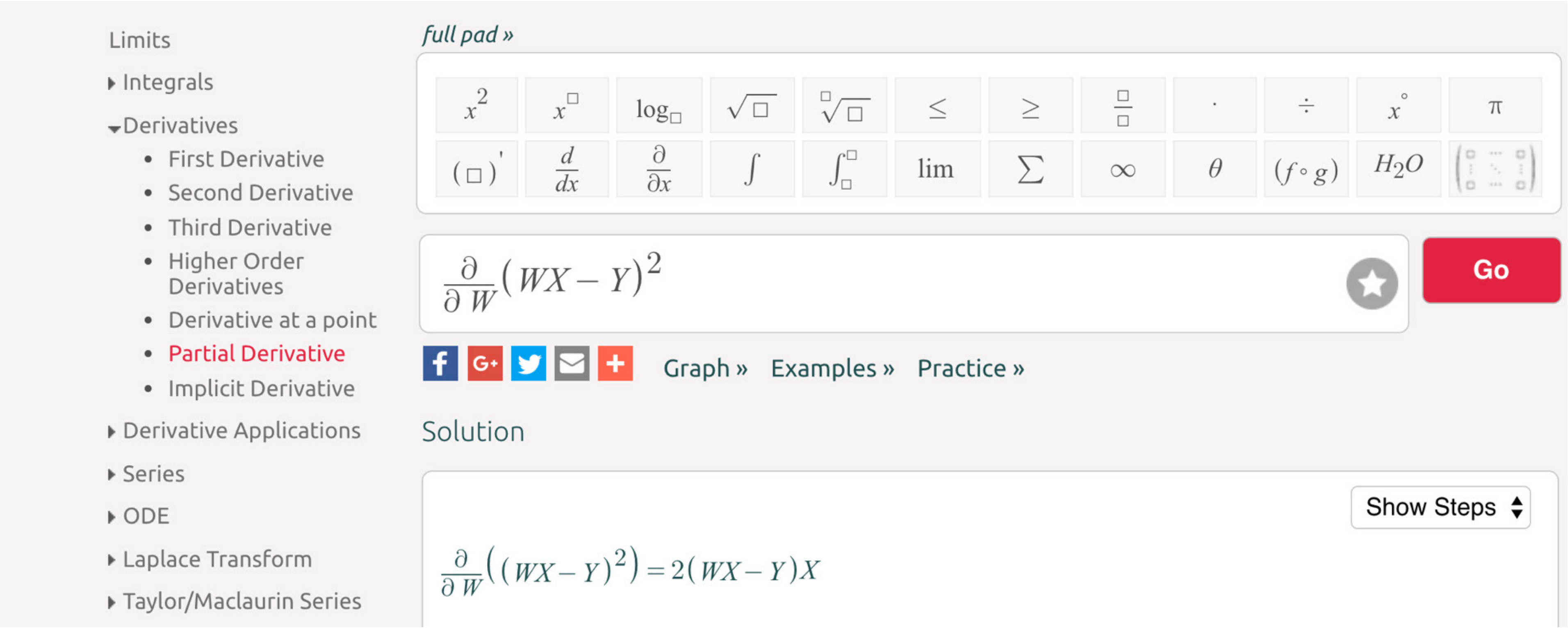
$$W := W - \alpha \frac{\partial}{\partial W} \frac{1}{2m} \sum_{i=1}^{m} (Wx^{(i)} - y^{(i)})^2$$

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

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

Partial Derivative Calculator

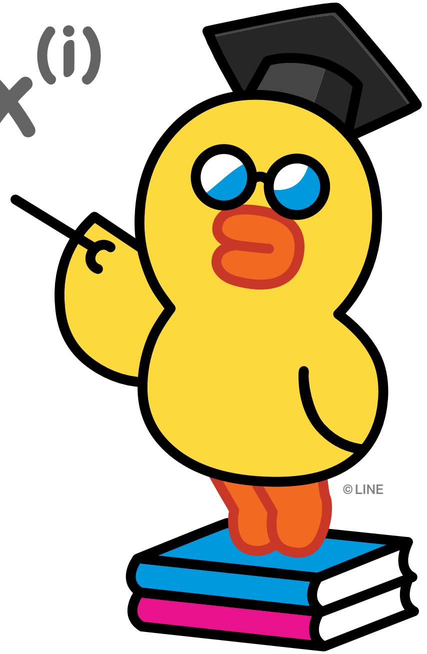
Partially differentiate functions step-by-step



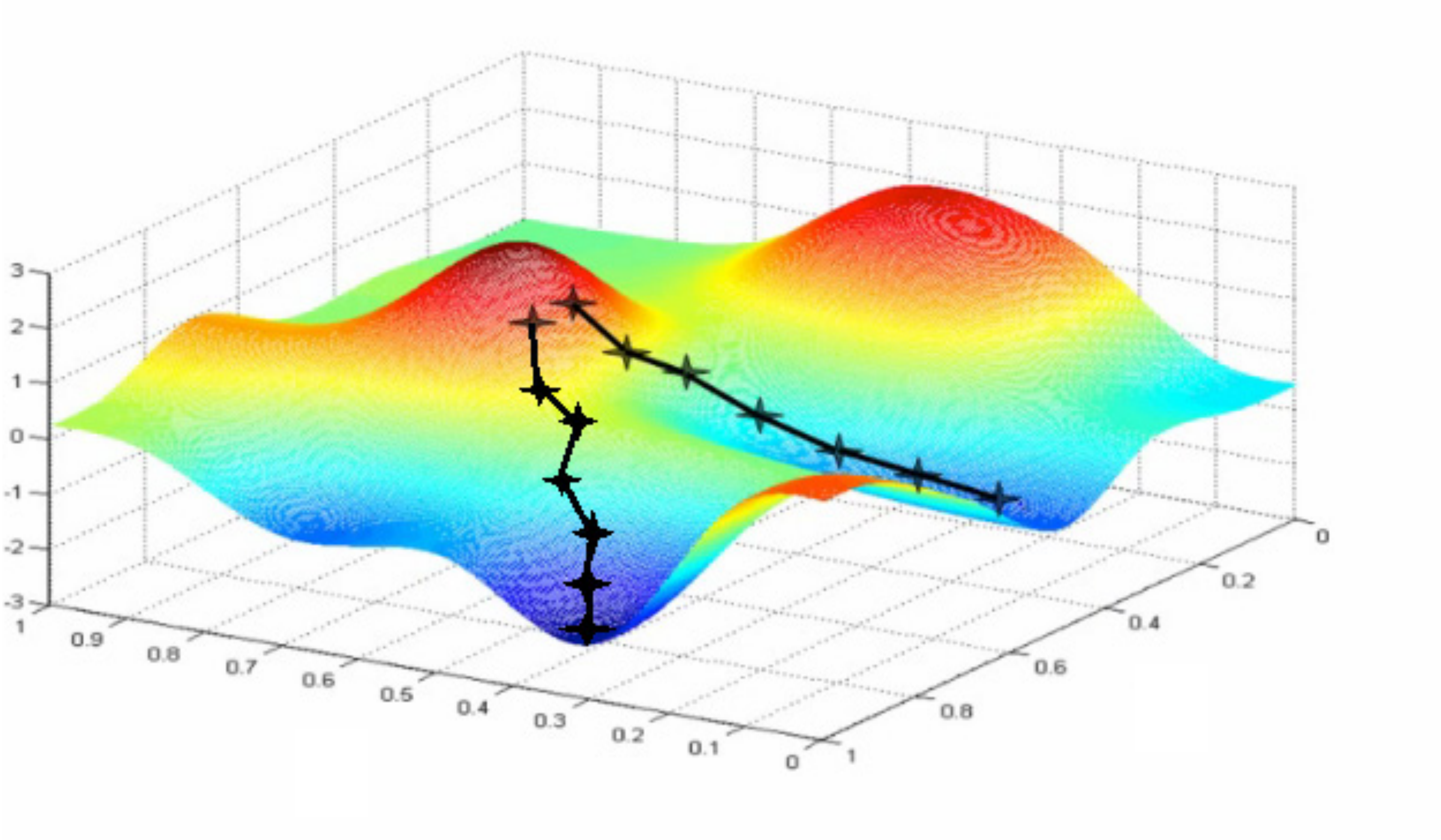
https://www.symbolab.com/solver/partial-derivative-calculator

Gradient Descent Algorithm

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



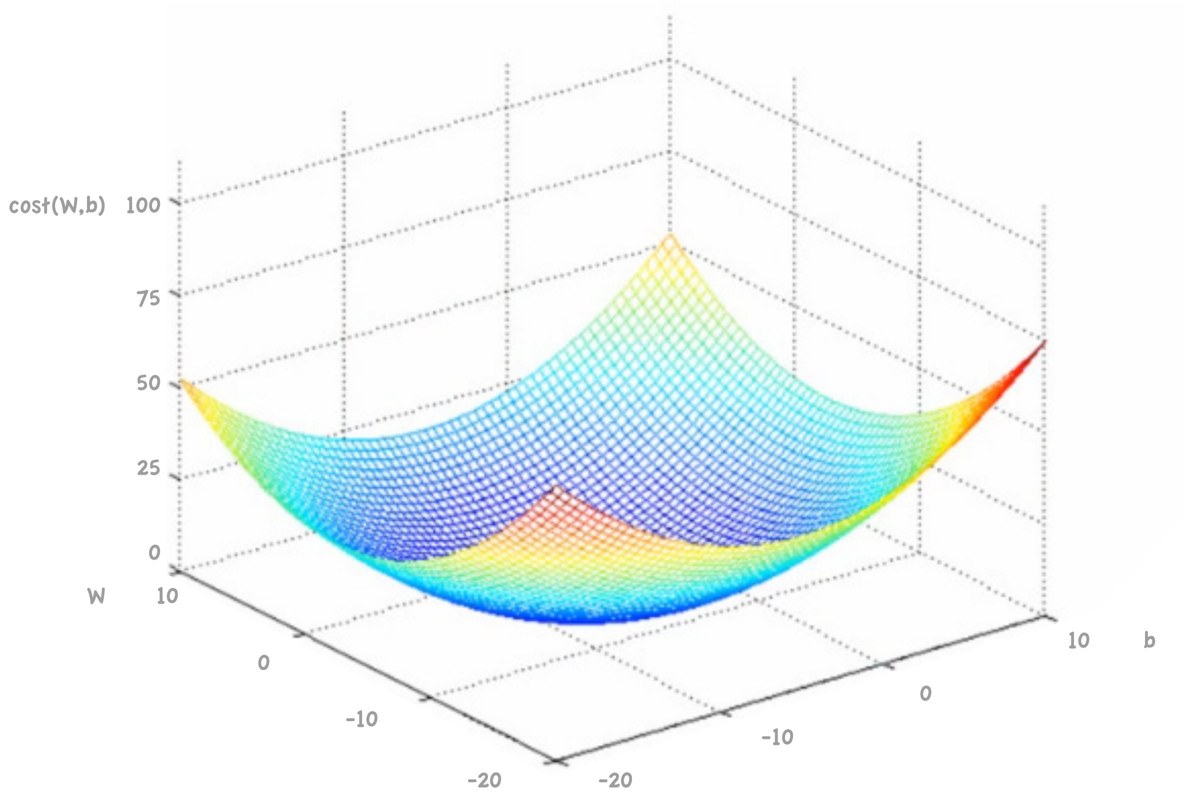
Convex Function



www.holehouse.org/mlclass/

Convex Function

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



www.holehouse.org/mlclass/

NEXT LECTURE

MULTIVARIABLE LOGISTIC REGRESSION