Lecture 5-1

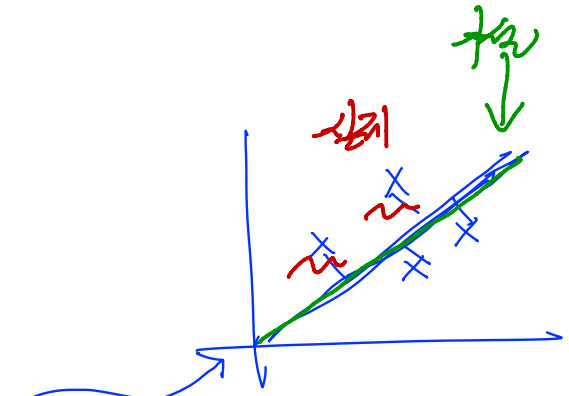
Logistic (regression) classification

Sung Kim <hunkim+mr@gmail.com>

Acknowledgement

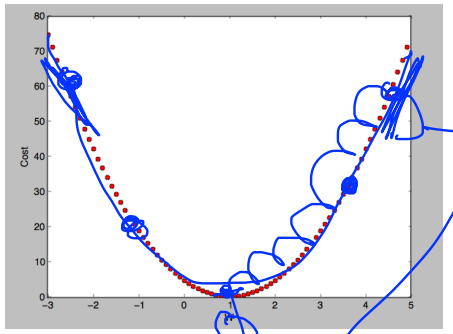
- Andrew Ng's ML class
  - https://class.coursera.org/ml-003/lecture
  - http://www.holehouse.org/mlclass/ (note)
- Convolutional Neural Networks for Visual Recognition
  - http://cs231n.github.io/
  - http://cs231n.stanford.edu/
- TensorFlow
  - https://www.tensorflow.org
  - https://github.com/aymericdamien/TensorFlow-Examples

Regression



x1 (hours)	x2 (attendance)	y (score)
10	5	90
9	5	80
3	2	50
2	4	60
11	1	40

- Hypothesis: $H(X) = WX$
- Cost: $cost(W) = \frac{1}{m}\sum(WX - y)^2$

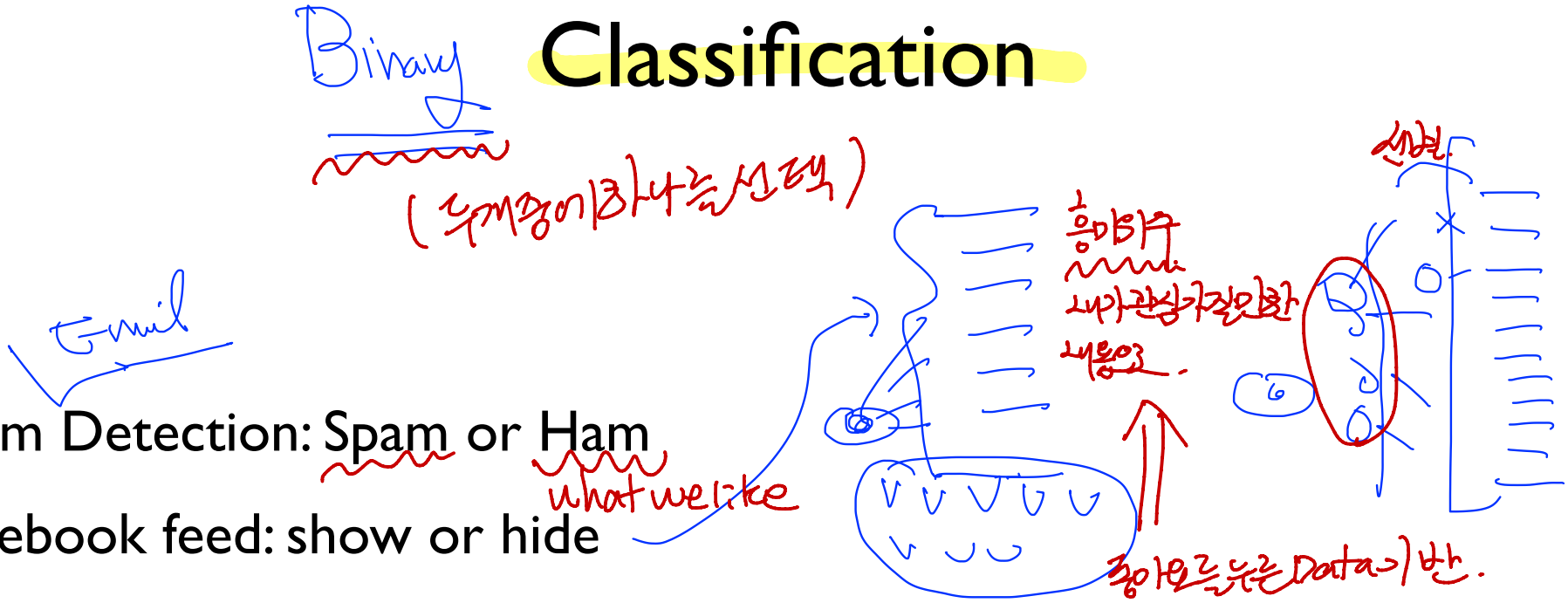


learning rate.

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

http://www.cse.iitk.ac.in/users/se367/10/presentation_local/Binary%20Classification.html

Classification



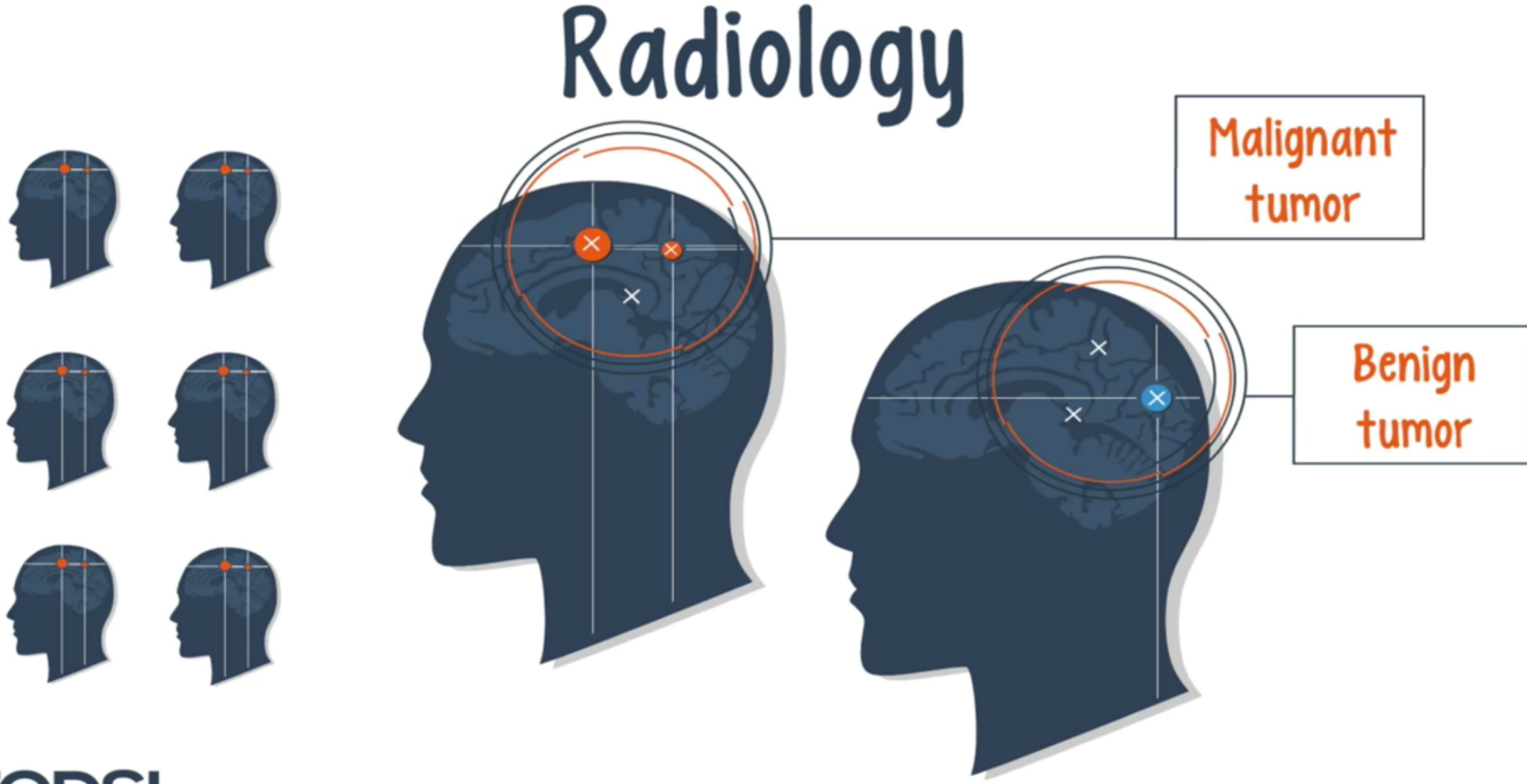
- Spam Detection: Spam or Ham
- Facebook feed: show or hide
- Credit Card Fraudulent Transaction detection: legitimate/fraud

http://www.cse.iitk.ac.in/users/se367/10/presentation_local/Binary%20Classification.html

0, 1 encoding

- Spam Detection: Spam (1) or Ham (0)
- Facebook feed: show(1) or hide(0)
- Credit Card Fraudulent Transaction detection: legitimate(0) or fraud (1)

http://www.cse.iitk.ac.in/users/se367/10/presentation_local/Binary%20Classification.html

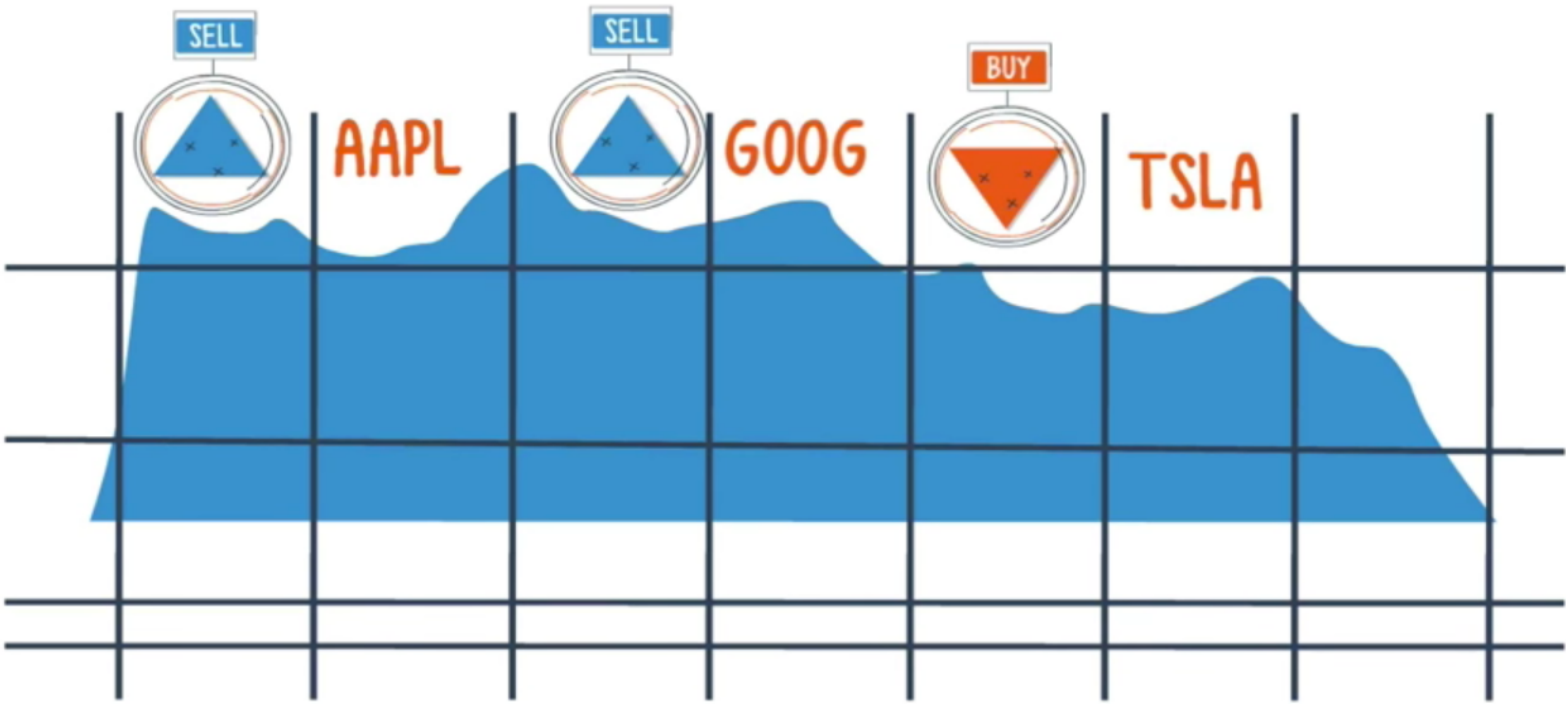


TODSI
Become A Better Data Scientist

https://www.youtube.com/watch?v=BmkA1ZsG2P4

Finance

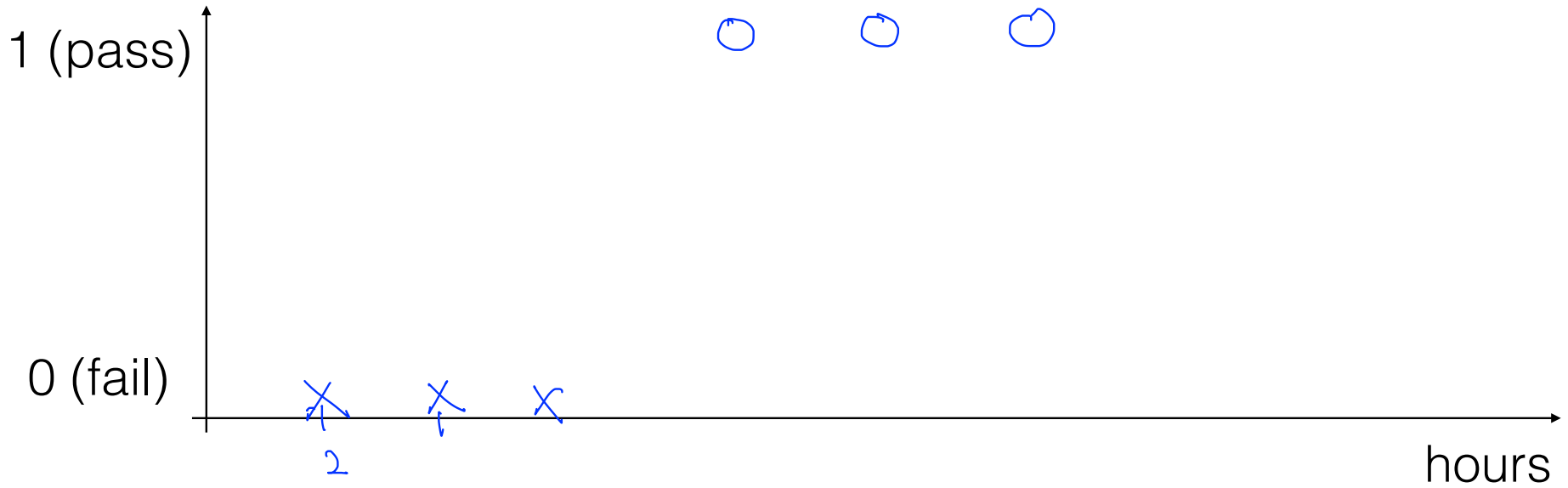
DWJI	17,499.10	▼
SP500	2,025.51	▼
NASDAQ	4,976.9	▲
AAPL	107.71	▲
GOOG	750.06	▲
TSLA	234.24	▼



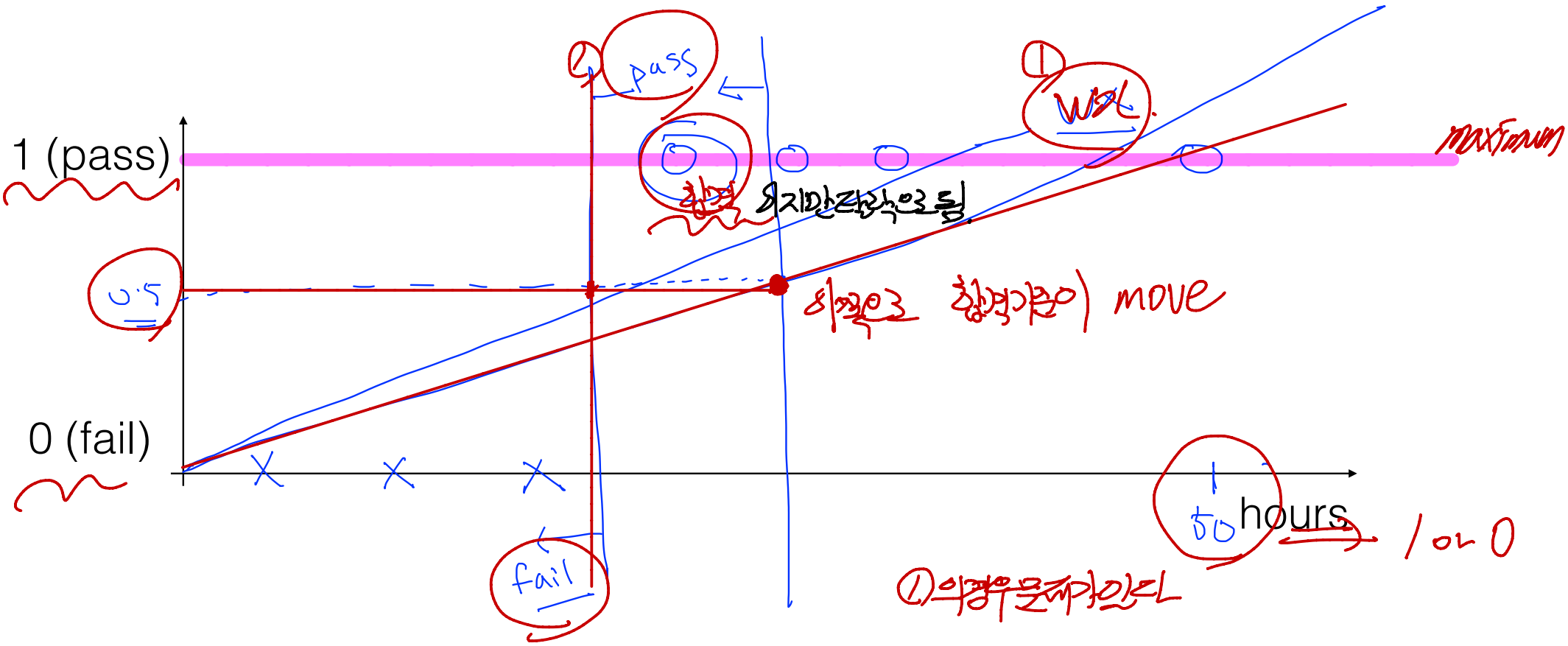
TODSI
Become A Better Data Scientist

https://www.youtube.com/watch?v=BmkA1ZsG2P4

Pass(1)/Fail(0) based on study hours



Linear Regression?

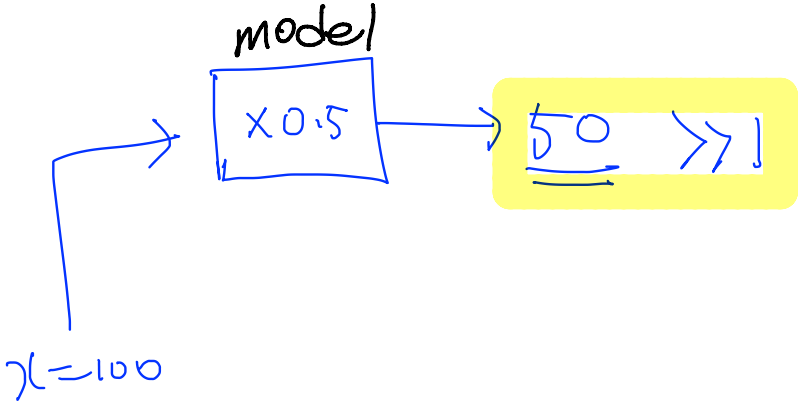


Linear regression

- We know Y is 0 or 1

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

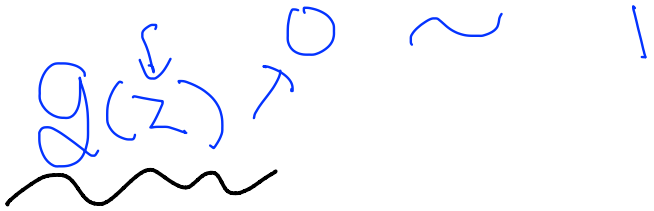
- Hypothesis can give values large than 1 or less than 0

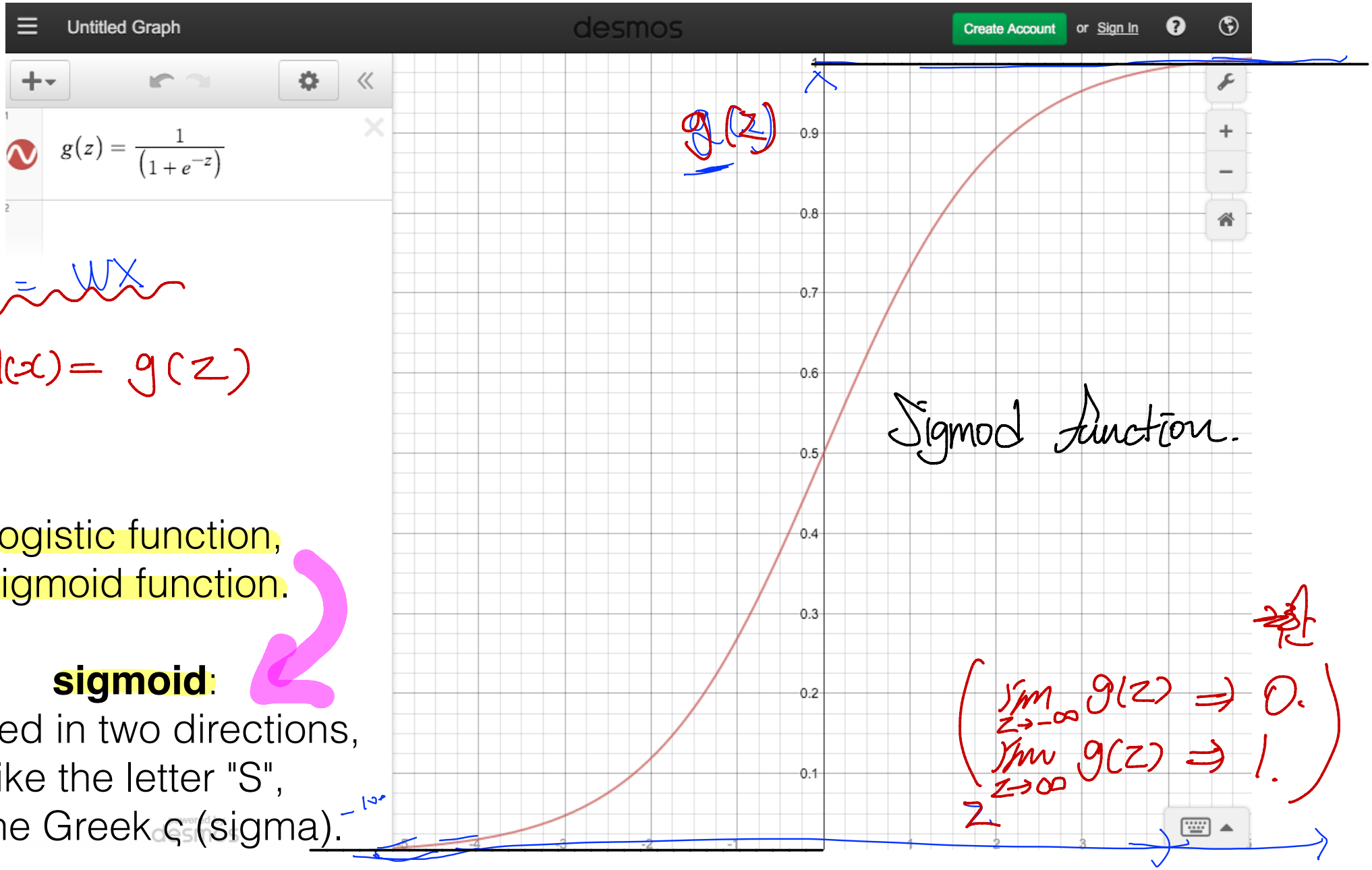


http://www.holehouse.org/mlclass/06_Logistic_Regression.html

Logistic Hypothesis

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





logistic function,
sigmoid function.

sigmoid:
Curved in two directions,
like the letter "S",
or the Greek ς (sigma).

Logistic Hypothesis

$$H(X) = \frac{1}{1 + e^{-W^T X}}$$

WX

$W^T X = Z$.

Next : Cost & minimize

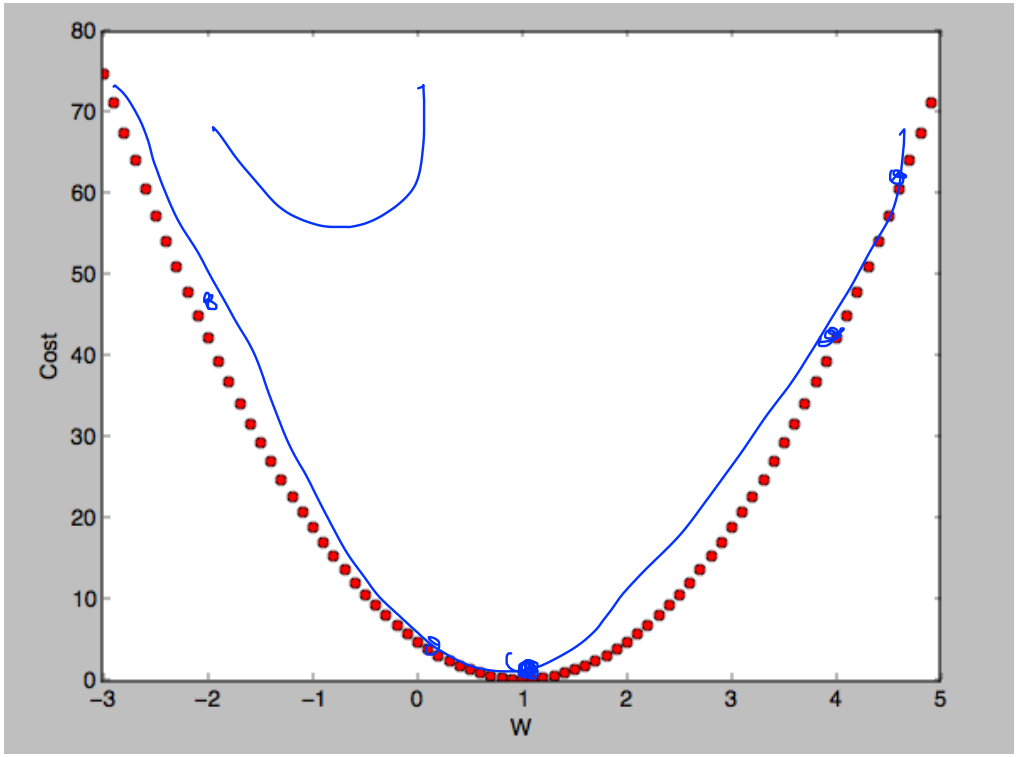
Lecture 5-2

Logistic (regression) classification: cost function & gradient decent

Sung Kim <hunkim+mr@gmail.com>

Cost

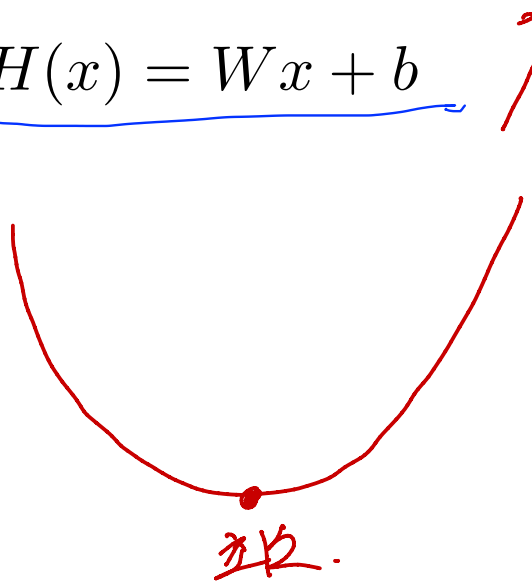
$$cost(W,b) = \frac{1}{m}\sum_{i=1}^{m}(H(x^{(i)}) - y^{(i)})^2 \quad \text{when} \quad H(x) = Wx + b$$



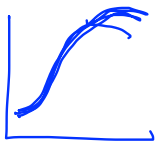
Cost function

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

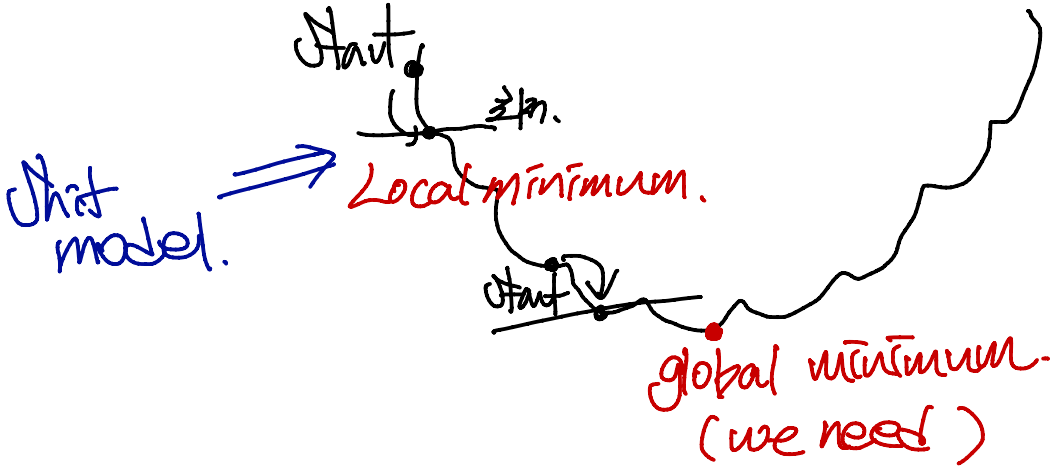
$H(x) = Wx + b$



$$H(X) = \frac{1}{1+e^{-W^T X}}$$



Not Linear.



New cost function for logistic

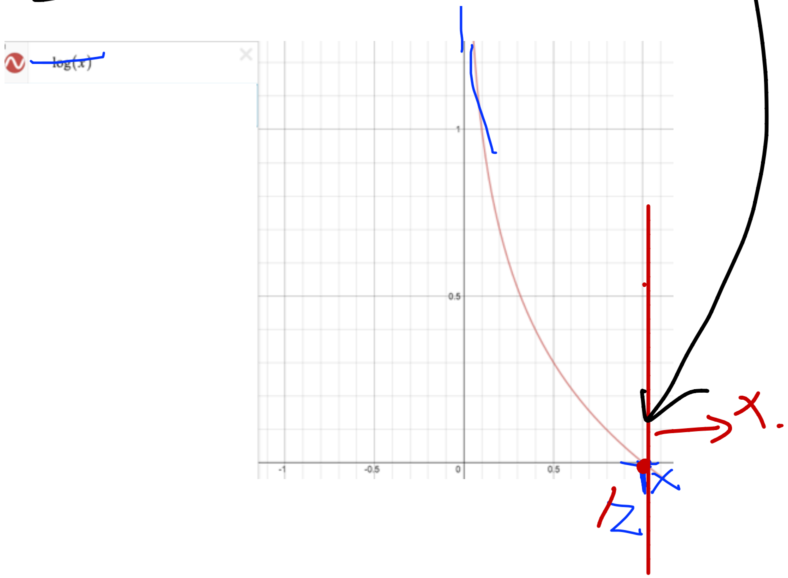
$$cost(W) = \frac{1}{m}\sum c(H(x), y)$$

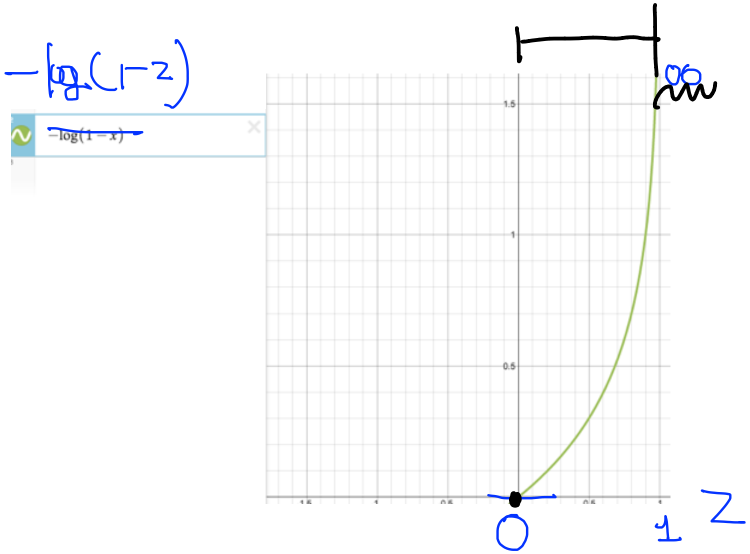
$$c(H(x), y) = \begin{cases} -log(H(x)) & : y = 1 \\ -log(1 - H(x)) & : y = 0 \end{cases}$$

Next

understanding cost function

$$C(H(x), y) = \begin{cases} -log(H(x)) & : y = 1 \\ -log(1 - H(x)) & : y = 0 \end{cases}$$





Cost function

$$cost(W) = \frac{1}{m}\sum \quad c(H(x), y)$$

$$c(H(x), y) = \begin{cases} -log(H(x)) & : y = 1 \\ -log(1 - H(x)) & : y = 0 \end{cases}$$

⇒ programming complexity.

$$c(H(x), y) = -ylog(H(x)) - (1 - y)log(1 - H(x))$$

① ②

y=1 ② $(1 - y)log(1 - H(x))$ ⇒

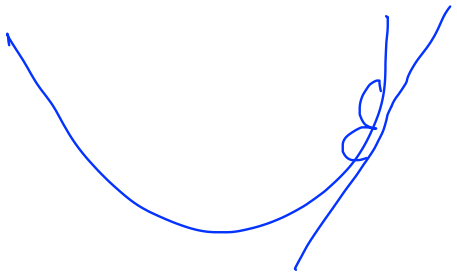
y=0 ① $-ylog(H(x))$ ⇒

y=1 $-ylog(H(x))$

y=0 $-1 * log(1 - H(x))$

Minimize cost - Gradient decent algorithm

$$cost(W) = -\frac{1}{m}\sum ylog(H(x)) + (1-y)log(1-H(x))$$



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

Gradient decent algorithm

$$cost(W) = -\frac{1}{m}\sum ylog(H(x)) + (1-y)log(1-H(x))$$

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

```
# cost function
cost = tf.reduce_mean(-tf.reduce_sum(Y*tf.log(hypothesis) + (1-Y)*tf.log(1-hypothesis)))

# Minimize
a = tf.Variable(0.1) # Learning rate, alpha
optimizer = tf.train.GradientDescentOptimizer(a)
train = optimizer.minimize(cost)
```

Next

Multinomial classification (Softmax)