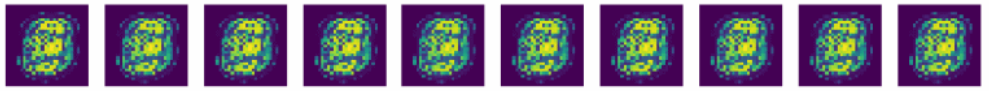
한림대학교 Summer Deep Learning Workshop

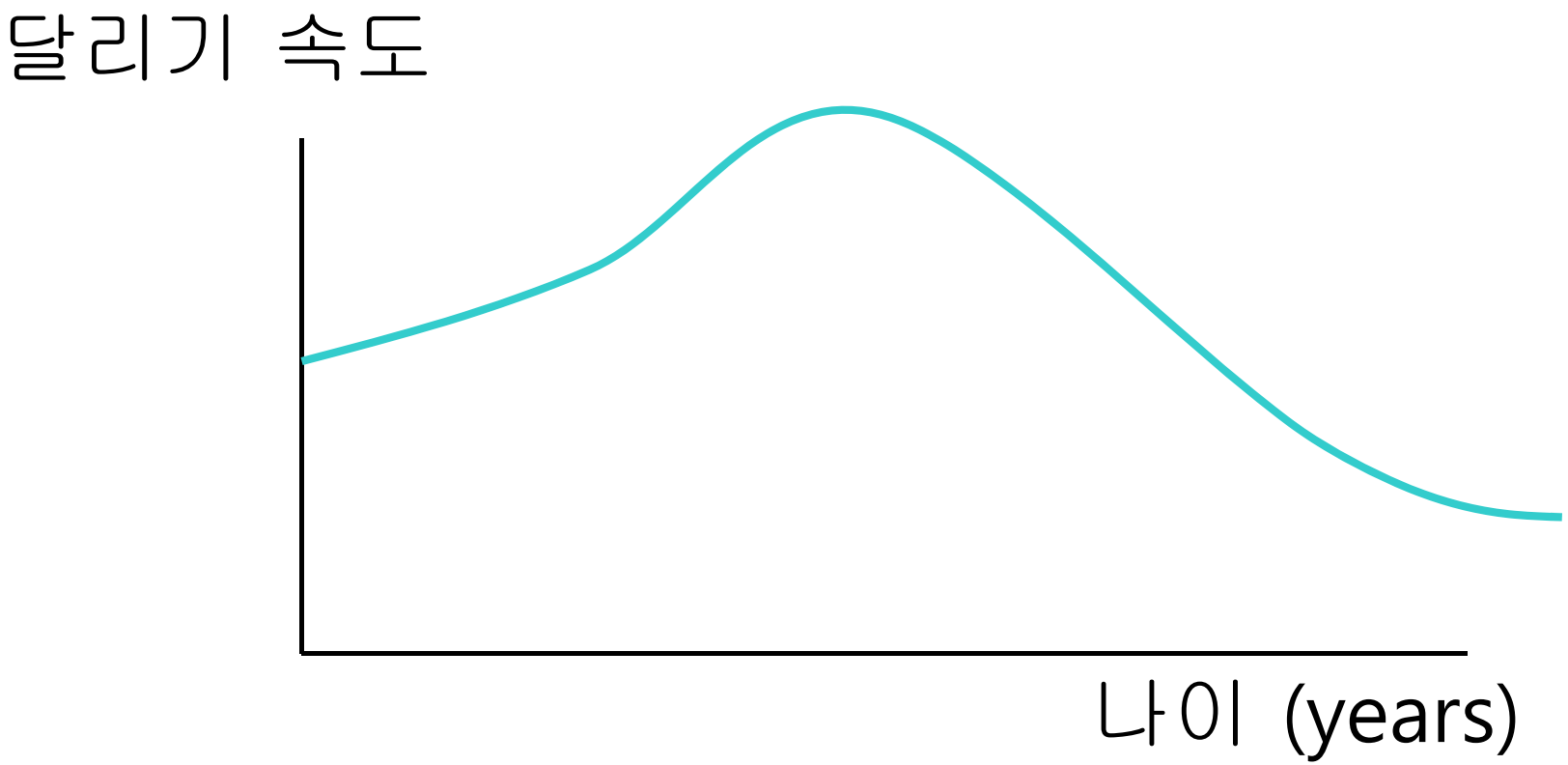
Logistic Regression

July 11, 2017

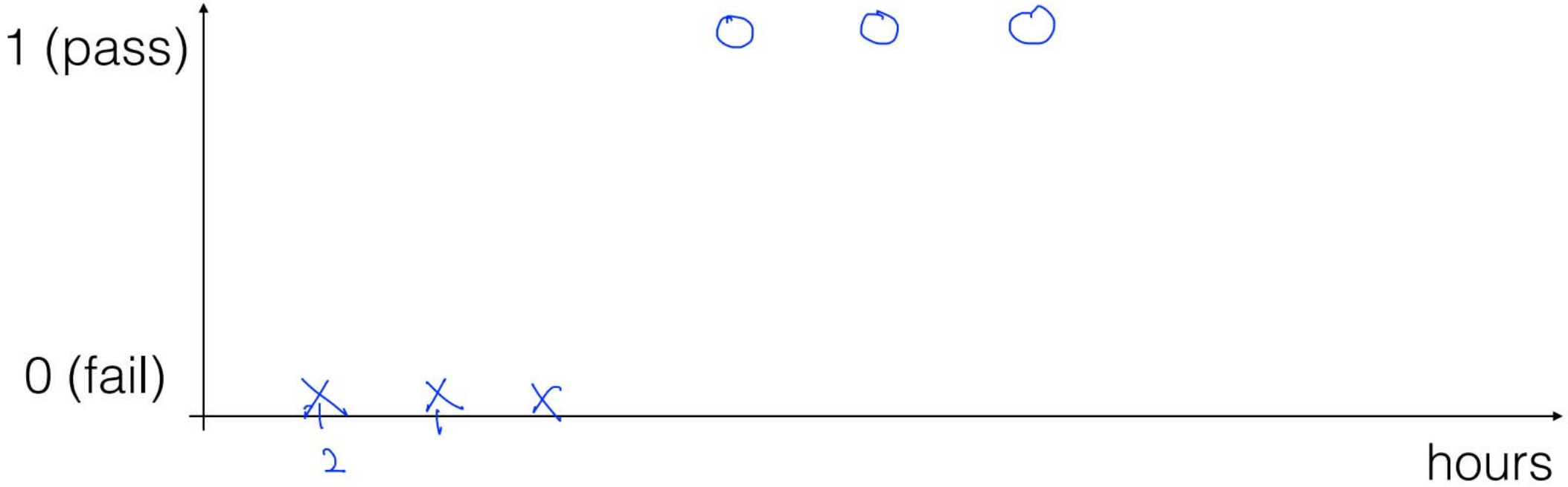
Seung-Chan Kim, Ph. D



선형성의 한계



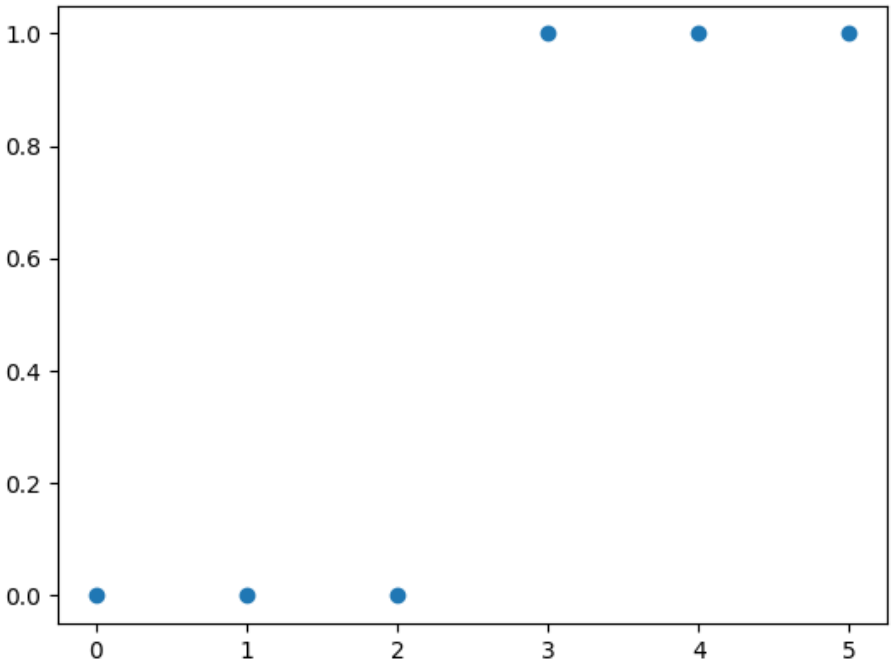
Pass(1)/Fail(0) based on study hours



https://hunkim.github.io/ml/lec5.pdf

Binary Classification - 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)

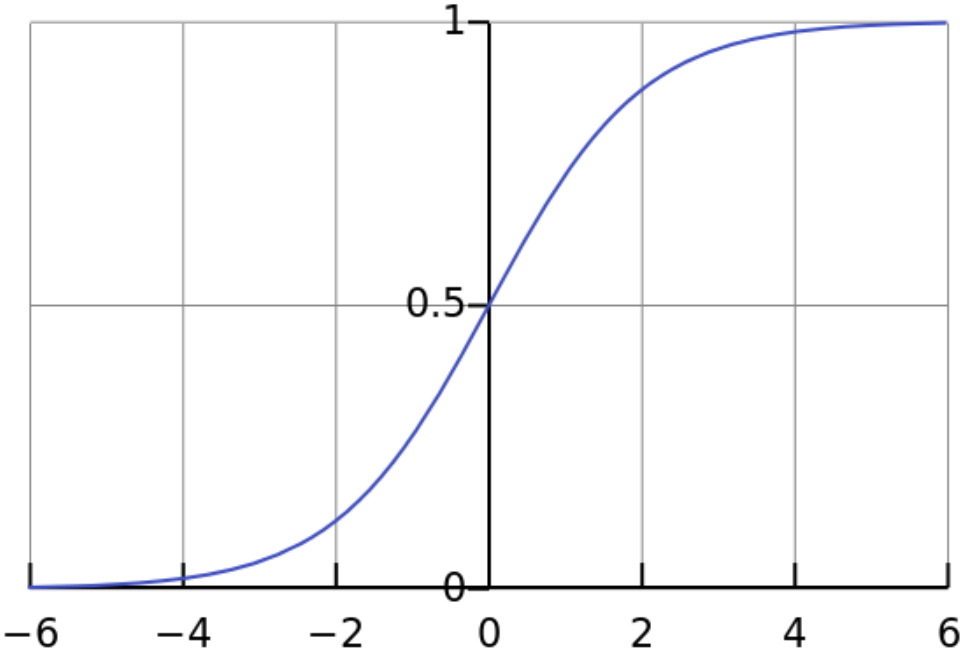

1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5

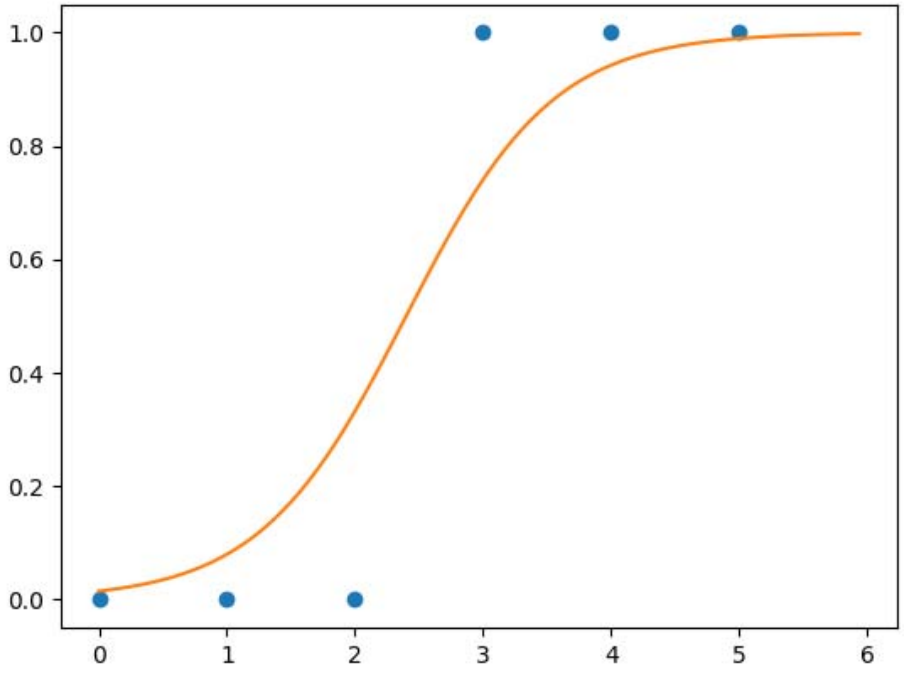
Sigmoid function

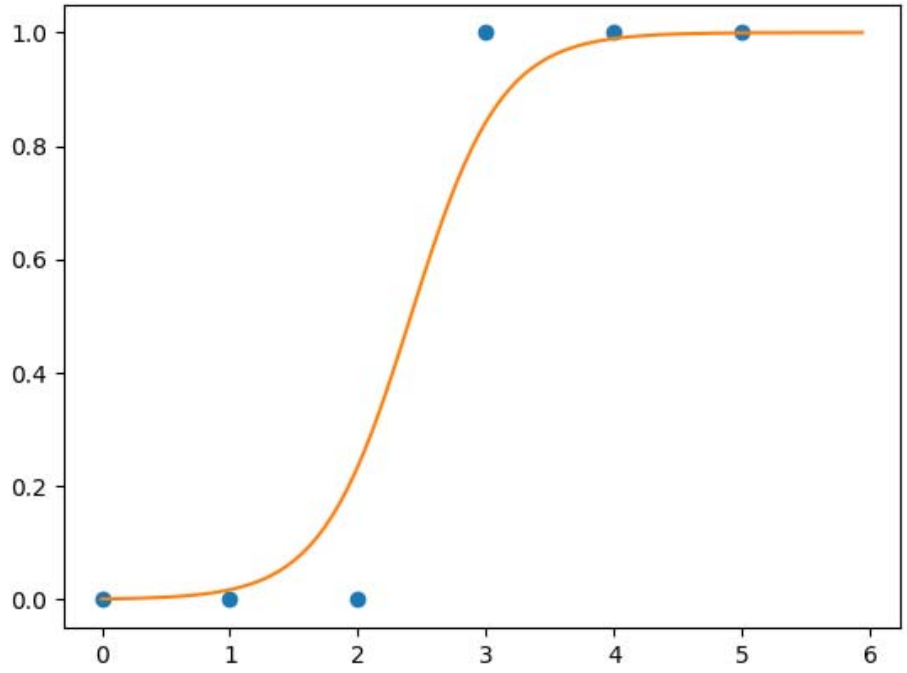
- A sigmoid function is a mathematical function having a characteristic "S"-shaped curve or sigmoid curve.

- Examples

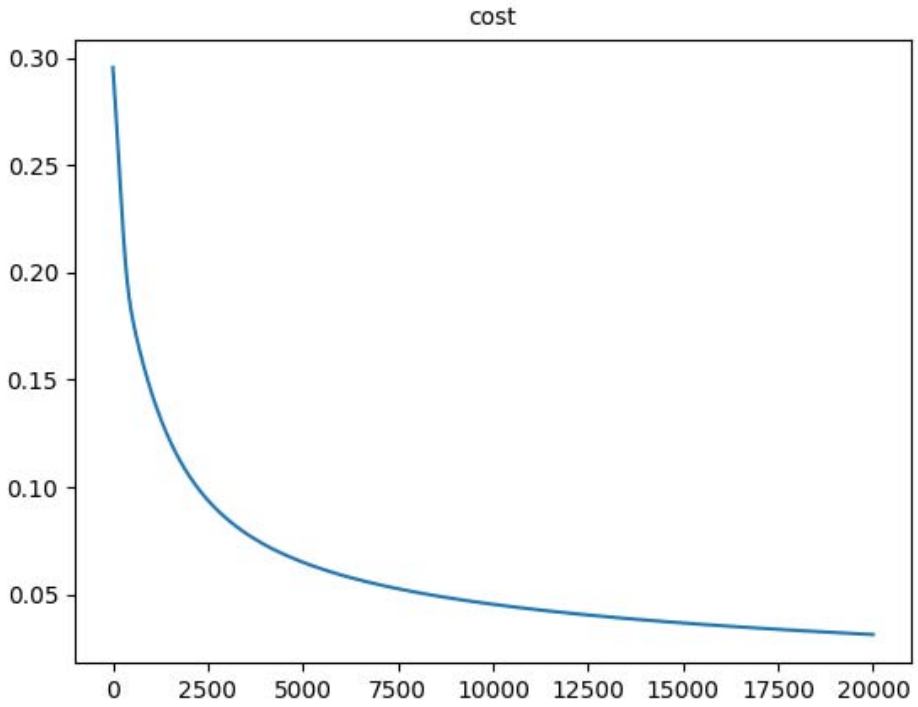
$$f(x) = \frac{1}{1 + e^{-x}}$$

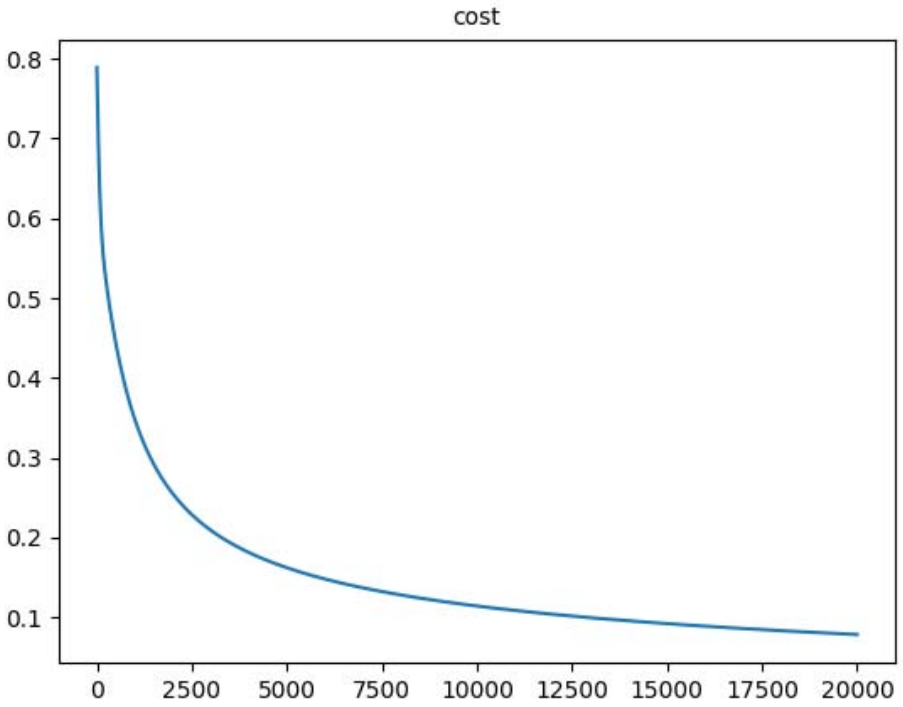






02-0-logistic-regression.py를 열어주세요


cost
0.30
0.25
0.20
0.15
0.10
0.05
0
2500
5000
7500
10000
12500
15000
17500
20000


cost
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
2500
5000
7500
10000
12500
15000
17500
20000

Understanding cost function

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

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

https://hunkim.github.io/ml/lec5.pdf

Understanding cost function

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

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

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

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

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

https://hunkim.github.io/ml/lec5.pdf

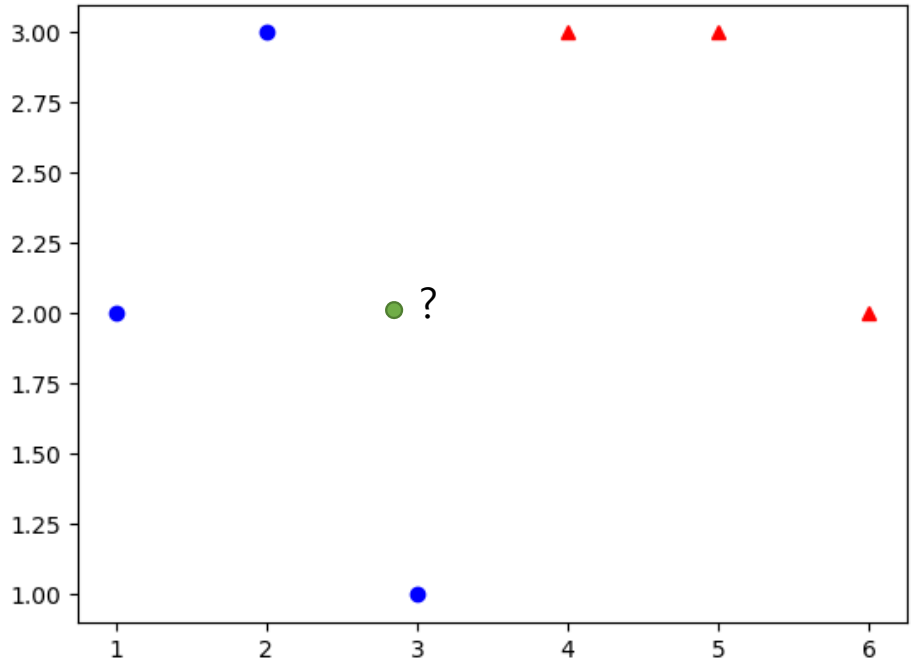
Understanding cost function

$$C(H(x), y) = -y\log(H(x)) - (1-y)\log(1-H(x))$$

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

Logistic (regression) classification

실제 문제로의 적용 가능



02-1-logistic-regression-classifier.py 를 열어주세요

Pima Indians Diabetes Data Set

Pima Indians Diabetes Data Set

Download: Data Folder, Data Set Description

Abstract: From National Institute of Diabetes and Digestive and Kidney Diseases; Includes cost data (donated by Peter Turney)

Data Set Characteristics:	Multivariate	**Number of Instances:**	768	**Area:**	Life
Attribute Characteristics:	Integer, Real	**Number of Attributes:**	8	**Date Donated**	1990-05-09
Associated Tasks:	Classification	**Missing Values?**	Yes	**Number of Web Hits:**	292585

Source:

Original Owners:

National Institute of Diabetes and Digestive and Kidney Diseases

Donor of database:

Vincent Sigillito (vgs '@' aplcen.apl.jhu.edu)
Research Center, RMI Group Leader
Applied Physics Laboratory
The Johns Hopkins University
Johns Hopkins Road
Laurel, MD 20707
(301) 953-6231

https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes

Pima Indians Diabetes Data Set

1. Number of times pregnant
2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test
3. Diastolic blood pressure (mm Hg)
4. Triceps skin fold thickness (mm)
5. 2-Hour serum insulin (mu U/ml)
6. Body mass index (weight in kg/(height in m)^2)
7. Diabetes pedigree function
8. Age (years)

9. Class variable (0 or 1)

https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes

data-03-diabetes.csv

```
-0.294118,0.487437,0.180328,-0.292929,0,0.00149028,-0.53117,-0.0333333,0
-0.882353,-0.145729,0.0819672,-0.414141,0,-0.207153,-0.766866,-0.666667,1
-0.0588235,0.839196,0.0491803,0,0,-0.305514,-0.492741,-0.633333,0
-0.882353,-0.105528,0.0819672,-0.535354,-0.777778,-0.162444,-0.923997,0,1
0,0.376884,-0.344262,-0.292929,-0.602837,0.28465,0.887276,-0.6,0
-0.411765,0.165829,0.213115,0,0,-0.23696,-0.894962,-0.7,1
-0.647059,-0.21608,-0.180328,-0.353535,-0.791962,-0.0760059,-0.854825,-0.833333,0
0.176471,0.155779,0,0,0,0.052161,-0.952178,-0.733333,1
```

```
753  0.0588235,0.708543,0.213115,-0.373737,0,0.311475,-0.722459,-0.266667,0
754  0.0588235,-0.105528,0.0163934,0,0,-0.329359,-0.945346,-0.6,1
755  0.176471,0.0150754,0.245902,-0.030303,-0.574468,-0.019374,-0.920581,0.4,1
756  -0.764706,0.226131,0.147541,-0.454545,0,0.0968703,-0.77626,-0.8,1
757  -0.411765,0.21608,0.180328,-0.535354,-0.735225,-0.219076,-0.857387,-0.7,1
758  -0.882353,0.266332,-0.0163934,0,0,-0.102832,-0.768574,-0.133333,0
759  -0.882353,-0.0653266,0.147541,-0.373737,0,-0.0938897,-0.797609,-0.933333,1
```

Acknowledgement

모두를 위한 머신러닝/딥러닝 강의

Seung-han

Seung-Chan

Jeung-Chan

감사합니다.