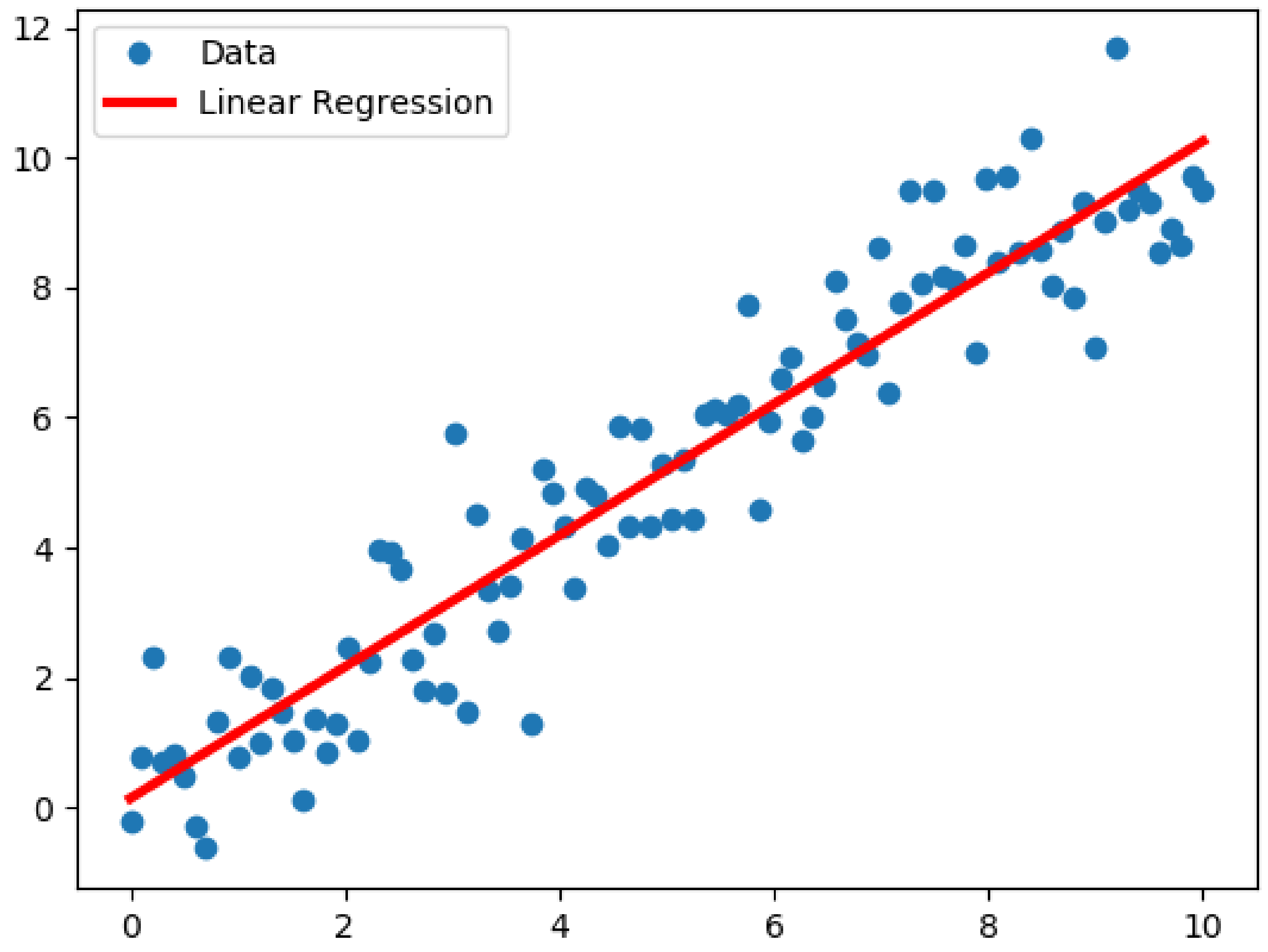
CS231n Base Lecture

- BOAZ 10기 박성현
- BOAZ 11기 김태희
- BOAZ 11기 홍지민
- BOAZ 10기 김용규

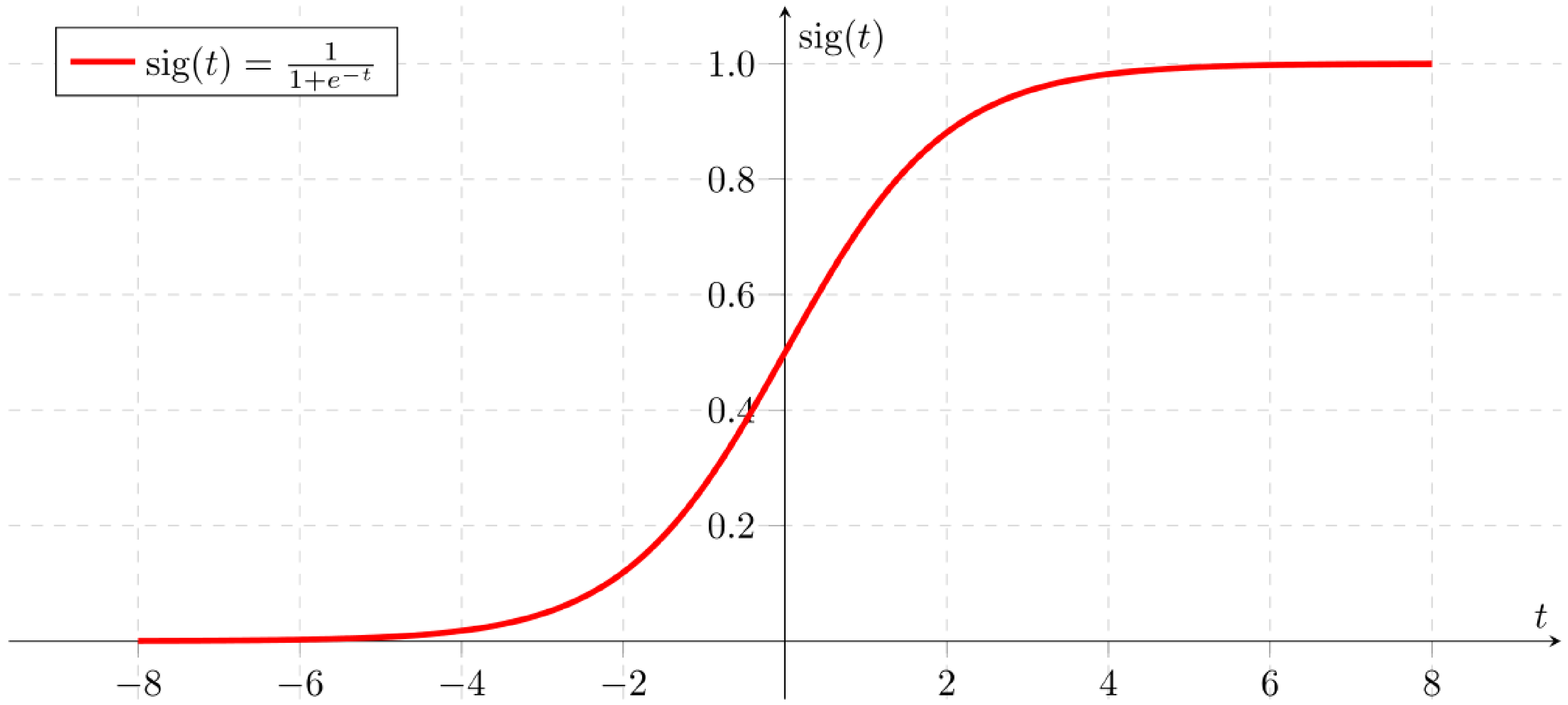
01

Linear Regression

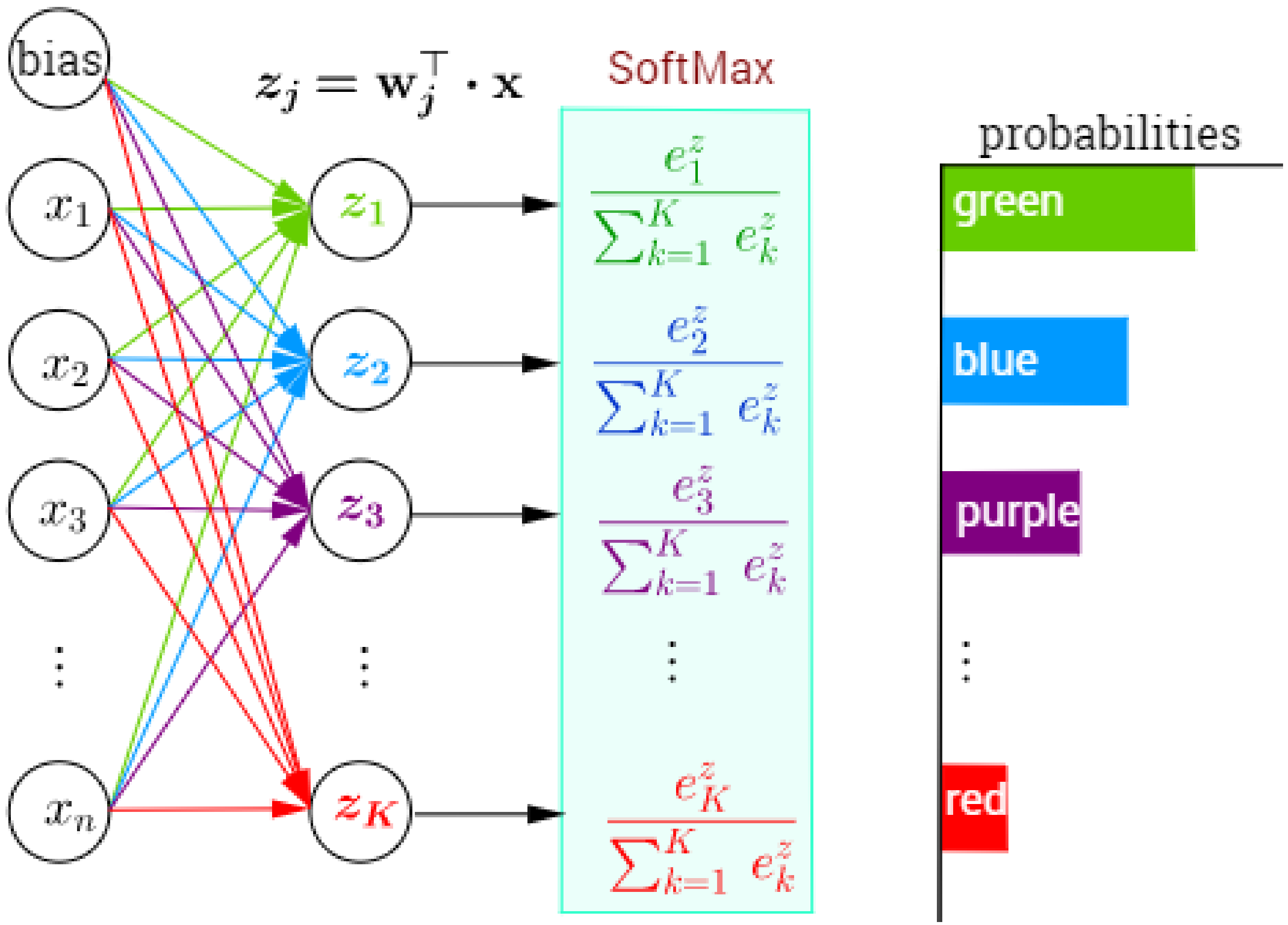


BOAZ

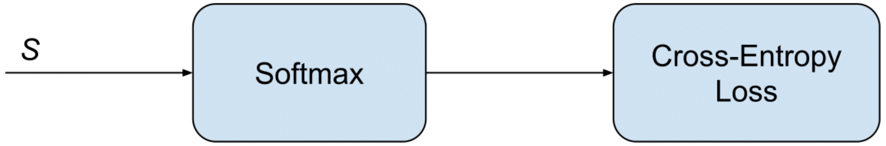
Logistic Regression



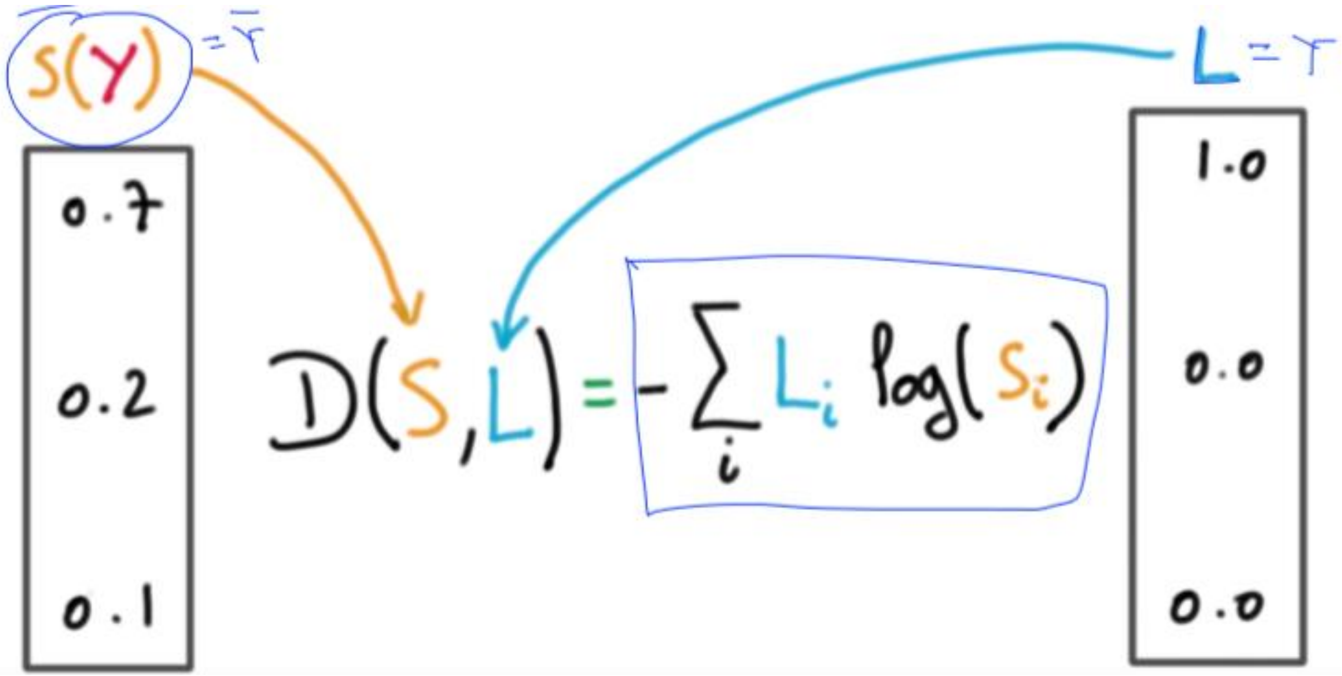
03 Softmax & Cross Entropy Loss



03 Softmax & Cross Entropy Loss

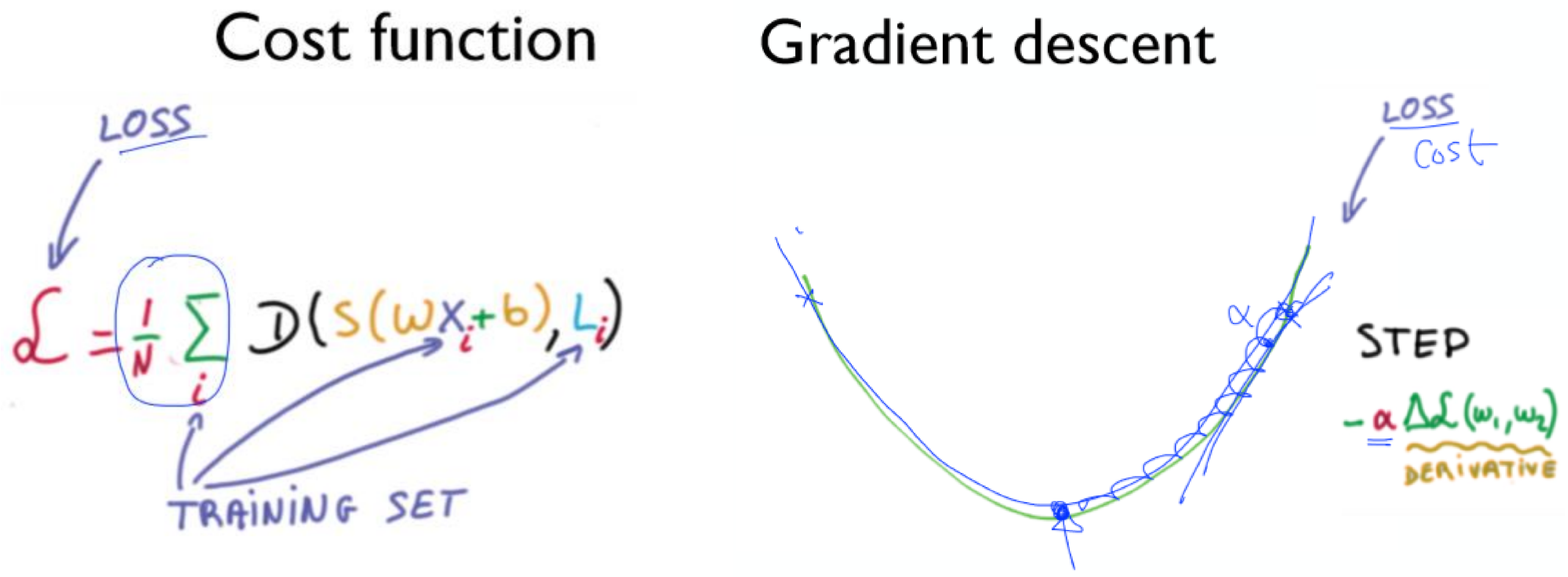


$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}} \qquad CE = -\sum_i^C t_i log(f(s)_i)$$



04

Gradient Descent



Gradient Descent

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

Parameters: θ_0, θ_1

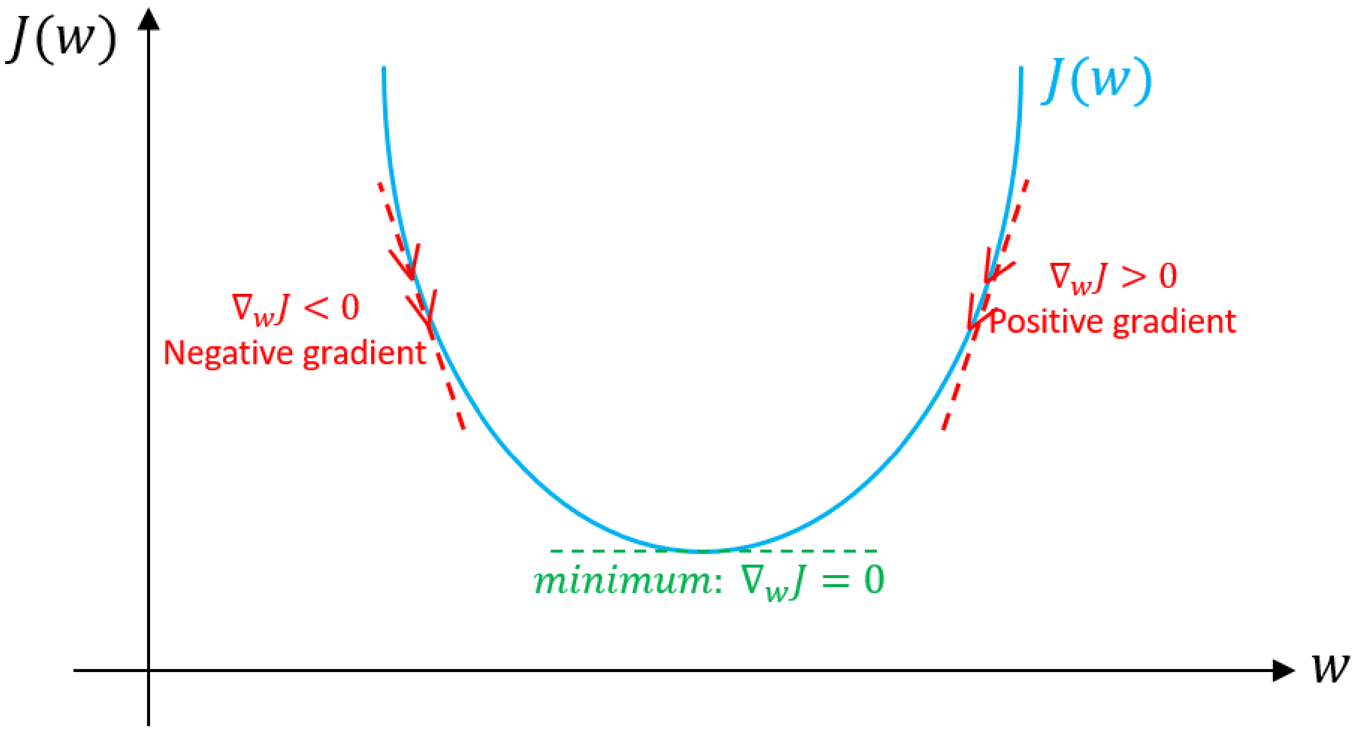
Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$

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

Andrew Ng

04

Gradient Descent



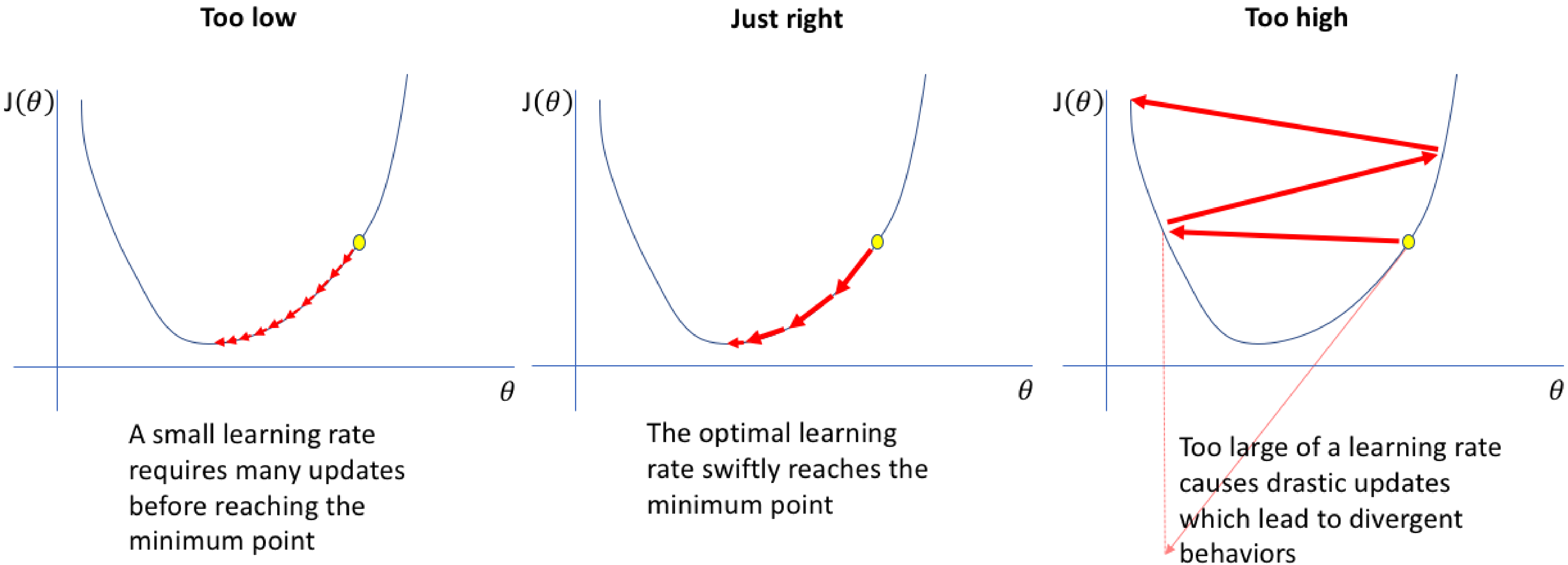
Repeat until convergence {

$$\theta_j \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

}

Gradient Descent

[Learning rate에 따른 변화]



A small learning rate requires many updates before reaching the minimum point

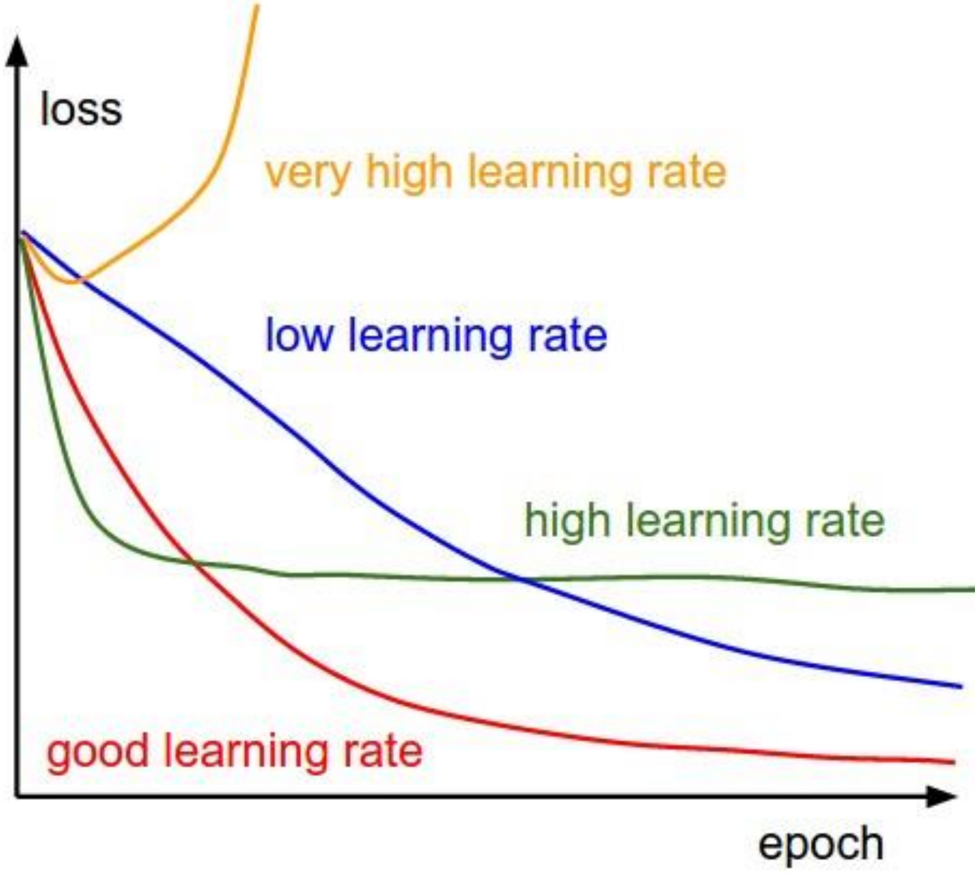
The optimal learning rate swiftly reaches the minimum point

Too large of a learning rate causes drastic updates which lead to divergent behaviors

Gradient Descent

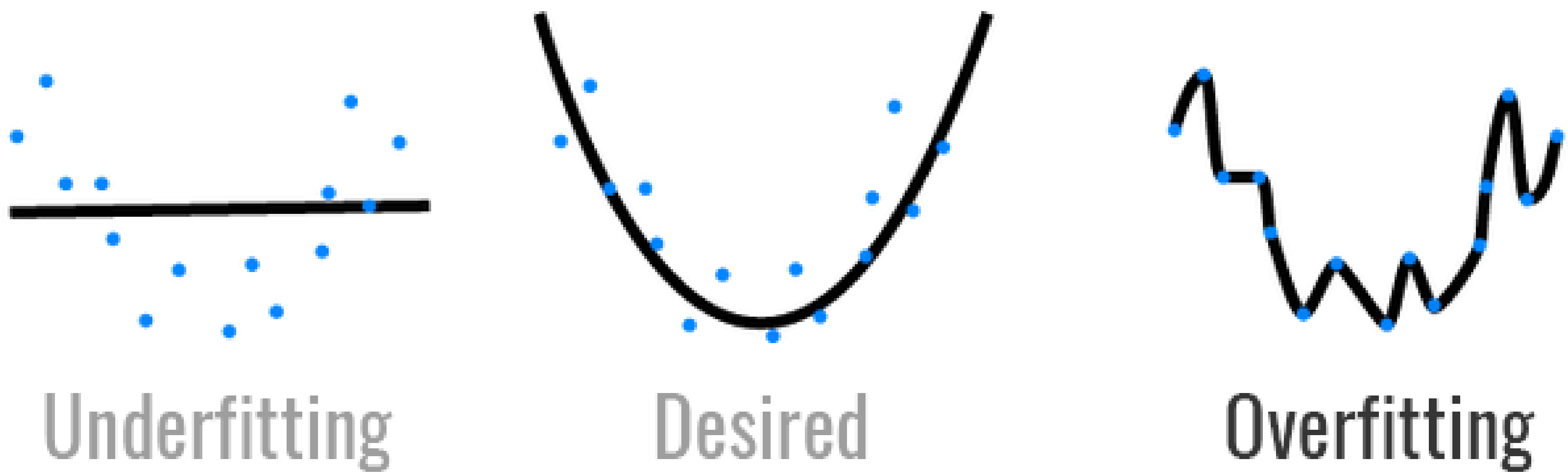
[Learning rate에 따른 변화]

Learning rate와 같은 변수를 **Hyperparameter**라고 한다.



BOAZ

Overfitting & Underfitting & Regularization



05

Overfitting & Underfitting & Regularization

Linear regression with regularization

Model: $h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

Regularization

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2m}\sum_{j=1}^{m}\theta_j^2$$

Large λ
High bias (underfit)
$\lambda = 10000.\ \theta_1 \approx 0, \theta_2 \approx 0, \ldots$
$h_\theta(x) \approx \theta_0$

Intermediate λ
"Just right"

Small λ
High variance (overfit)
$\lambda \approx 0$

BOAZ

Evaluation

[Train / Val / Test로 나누는 이유]

Idea #1: Choose hyperparameters that work best on the data

BAD: K = 1 always works perfectly on training data

Your Dataset

Idea #2: Split data into **train** and **test**, choose hyperparameters that work best on test data

BAD: No idea how algorithm will perform on new data

train	test

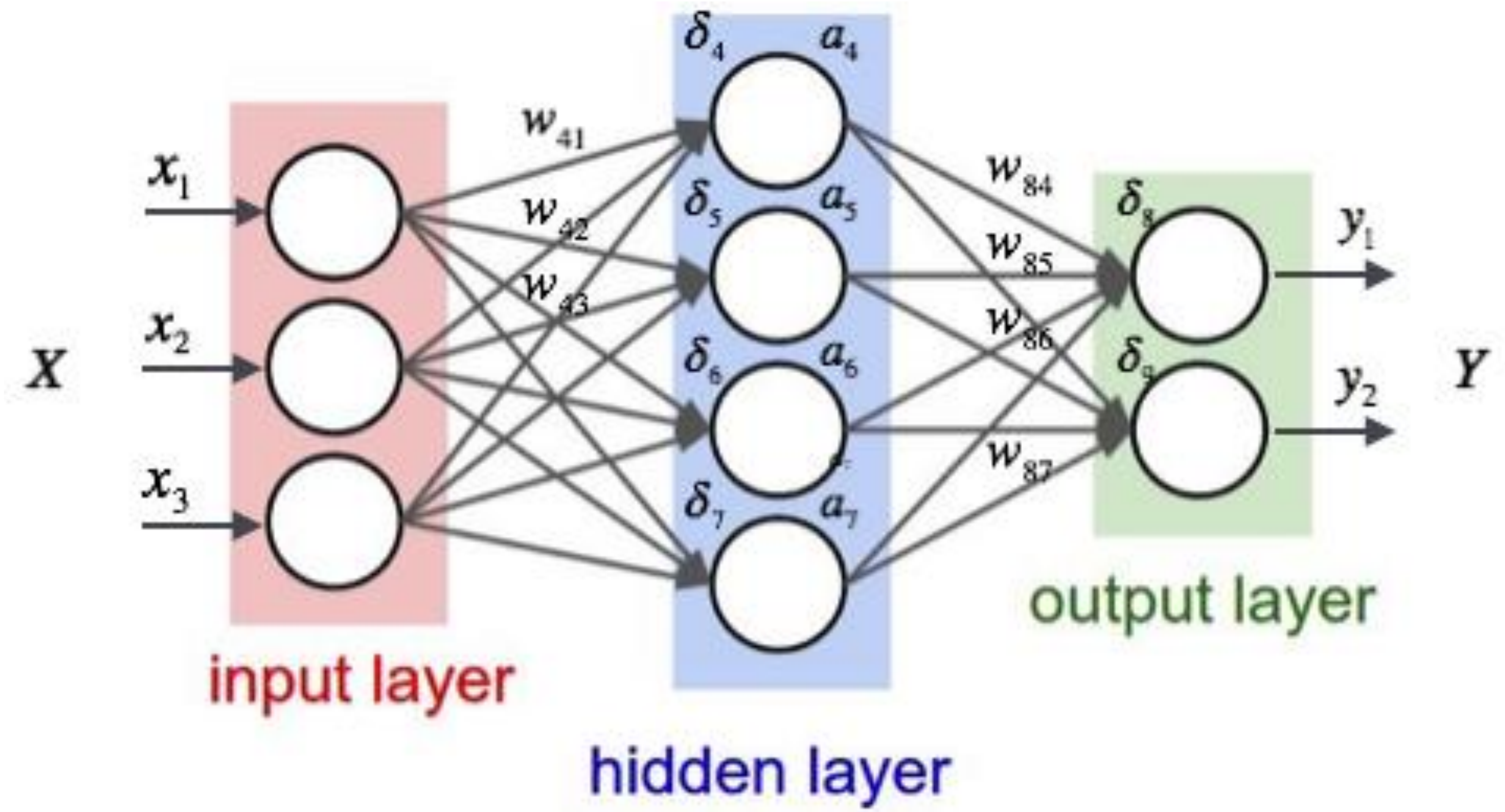
Idea #3: Split data into **train**, **val**, and **test**; choose hyperparameters on val and evaluate on test

Better!

train	validation	test

Neural Network

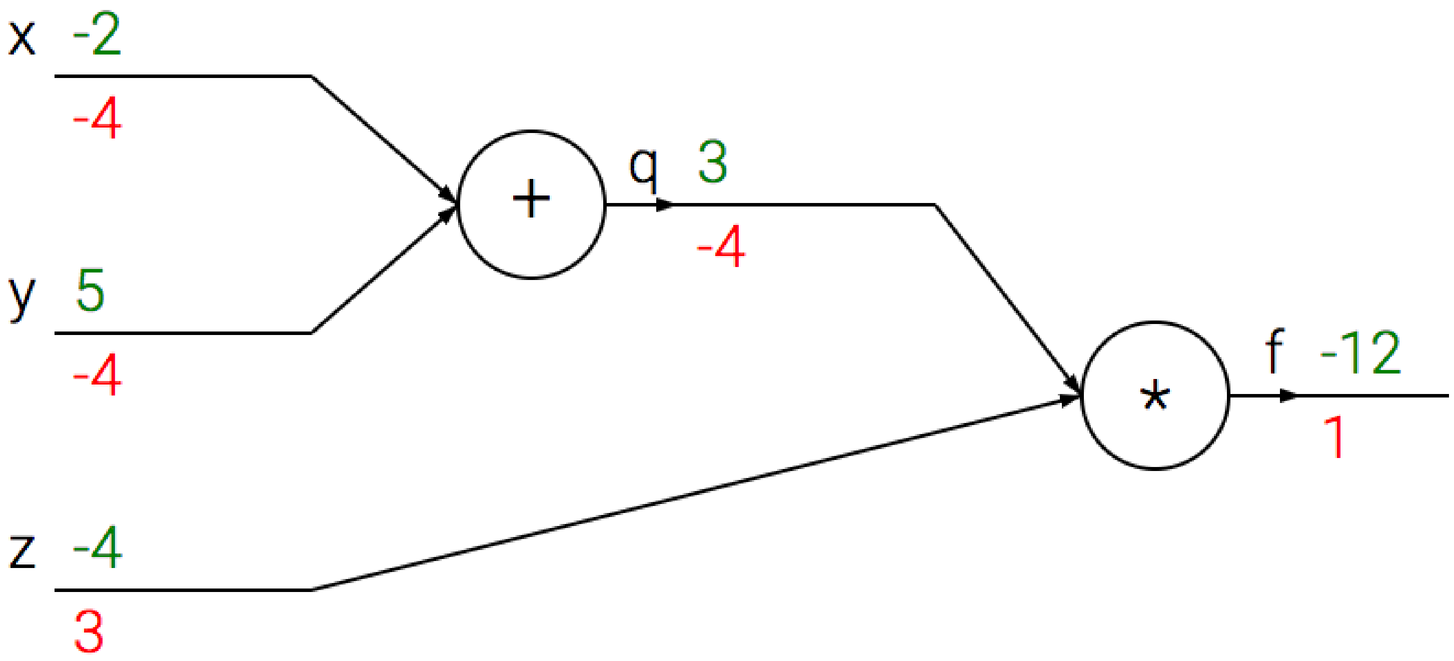
[Neural Network 구조]



Backpropagation

[Backpropagation 예제]

Backpropagation은 CS231n Lecture4에서 자세히



09

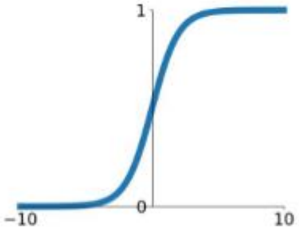
Activation Function

[Activation Function]

Activation Function은 CS231n Lecture6에서 자세히

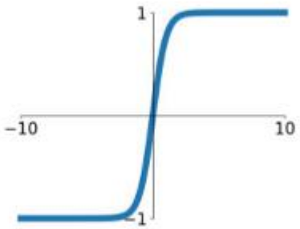
Sigmoid

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



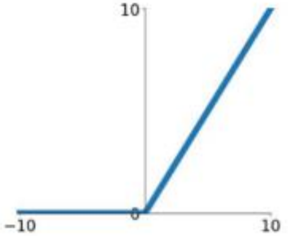
tanh

$\tanh(x)$



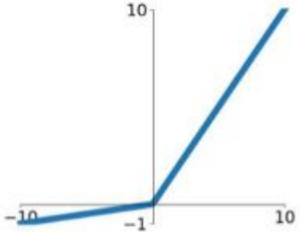
ReLU

$\max(0, x)$



Leaky ReLU

$\max(0.1x, x)$

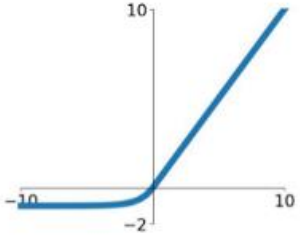


Maxout

$\max(w_1^T x + b_1, w_2^T x + b_2)$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Weight Initialization

[Weight Initialization]
Weight Initialization은 CS231n Lecture6에서 자세히

Xavier/He initialization

- Makes sure the weights are 'just right', not too small, not too big
- Using number of input (fan_in) and output (fan_out)

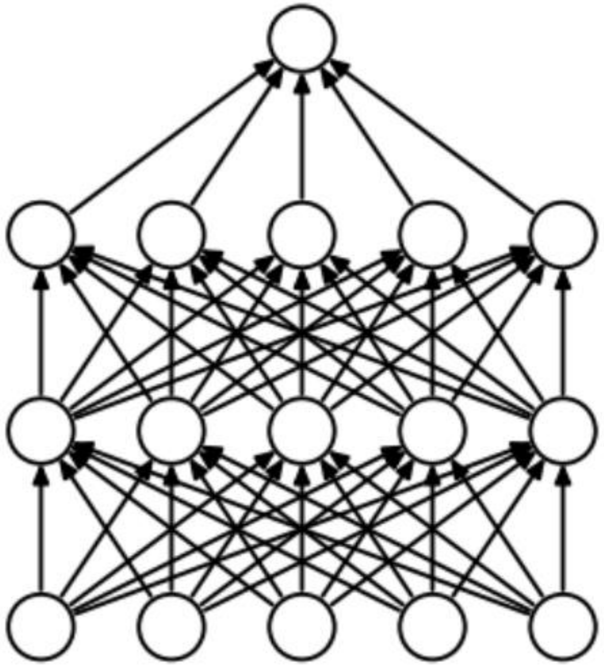
```
# Xavier initialization
# Glorot et al. 2010
W = np.random.randn(fan_in, fan_out)/np.sqrt(fan_in)

# He et al. 2015
W = np.random.randn(fan_in, fan_out)/np.sqrt(fan_in/2)
```

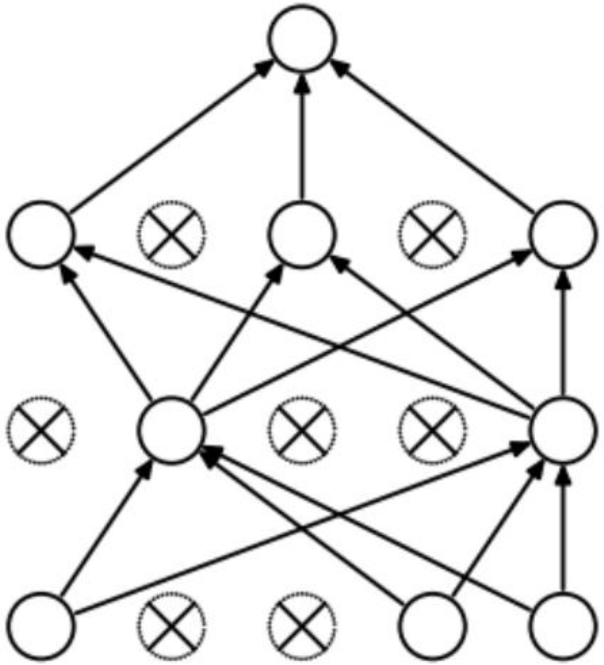
BOAZ

Dropout & Model Ensemble

[Dropout]
Deep Learning에서 Overfitting을 줄이는 1가지 방법



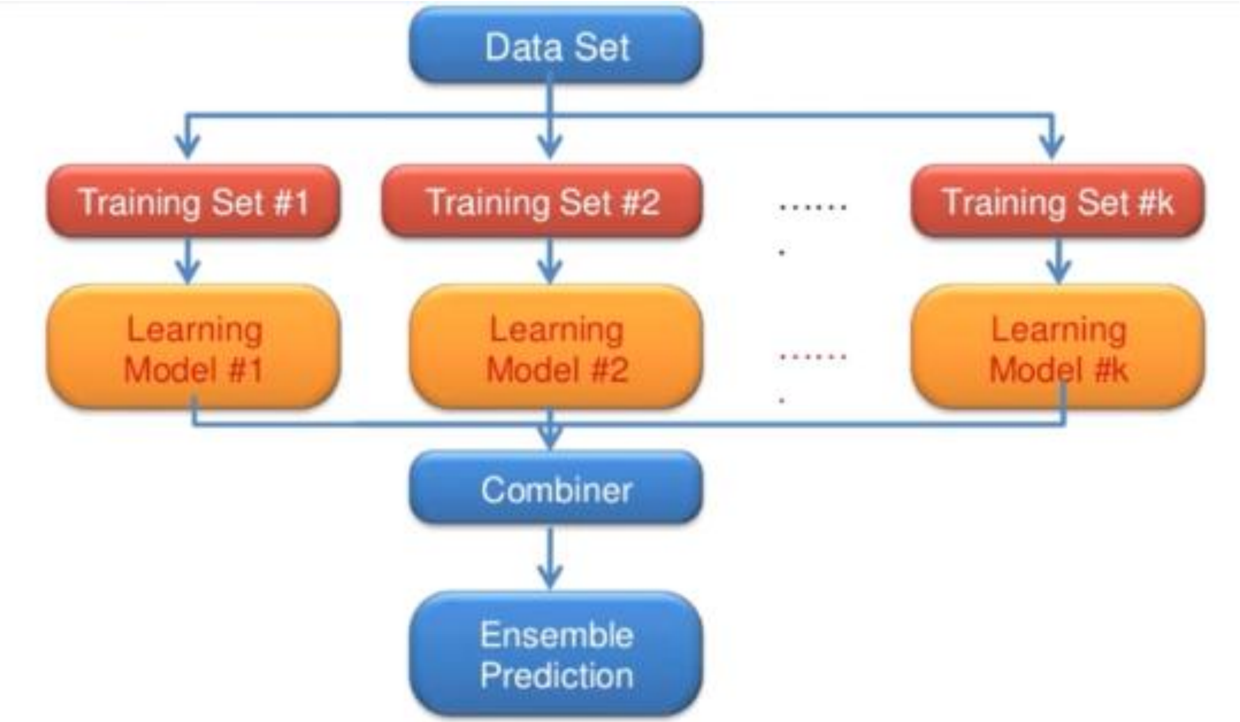
(a) Standard Neural Net



(b) After applying dropout.

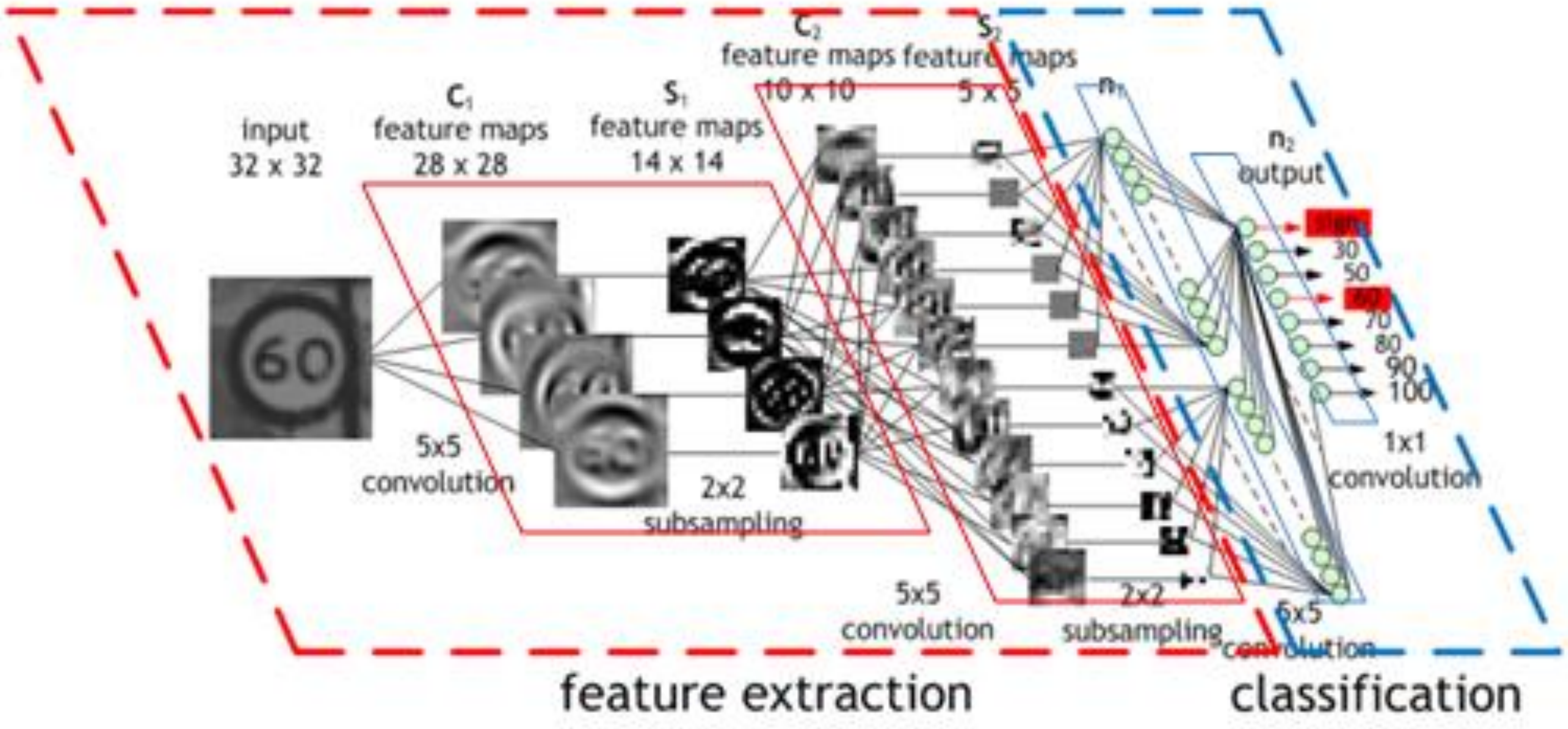
Dropout & Model Ensemble

[Model Ensemble]
마지막에 모델의 성능을 조금 더 향상시키기 위한 방법



CNN

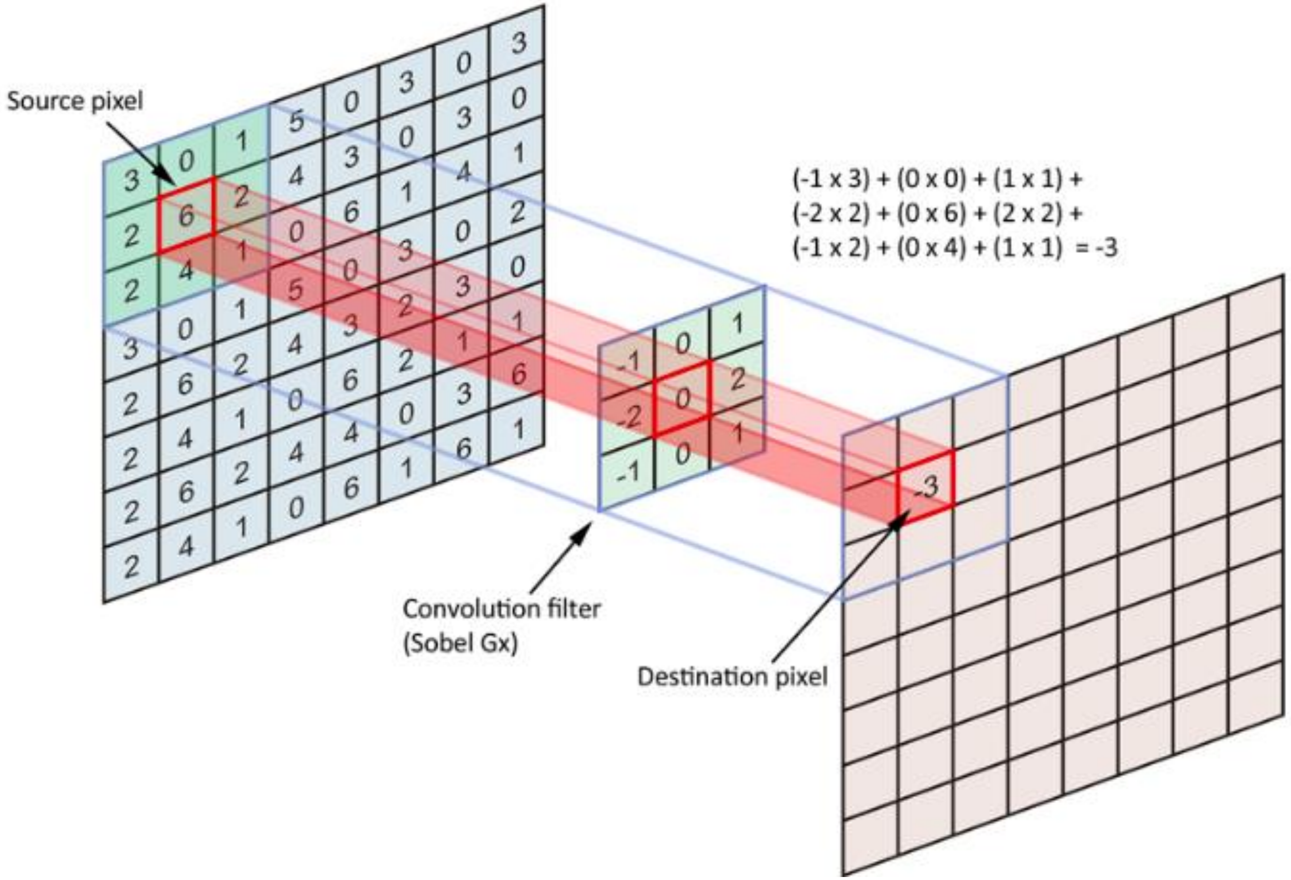
[CNN 모델 예시]


input
32 x 32
C1
feature maps
28 x 28
S1
feature maps
14 x 14
C2
feature maps
10 x 10
S2
feature maps
5 x 5
n1
n2
output
30
50
70
80
90
100
5x5
convolution
2x2
subsampling
5x5
convolution
2x2
subsampling
6x5
convolution
1x1
convolution
feature extraction
classification

CNN

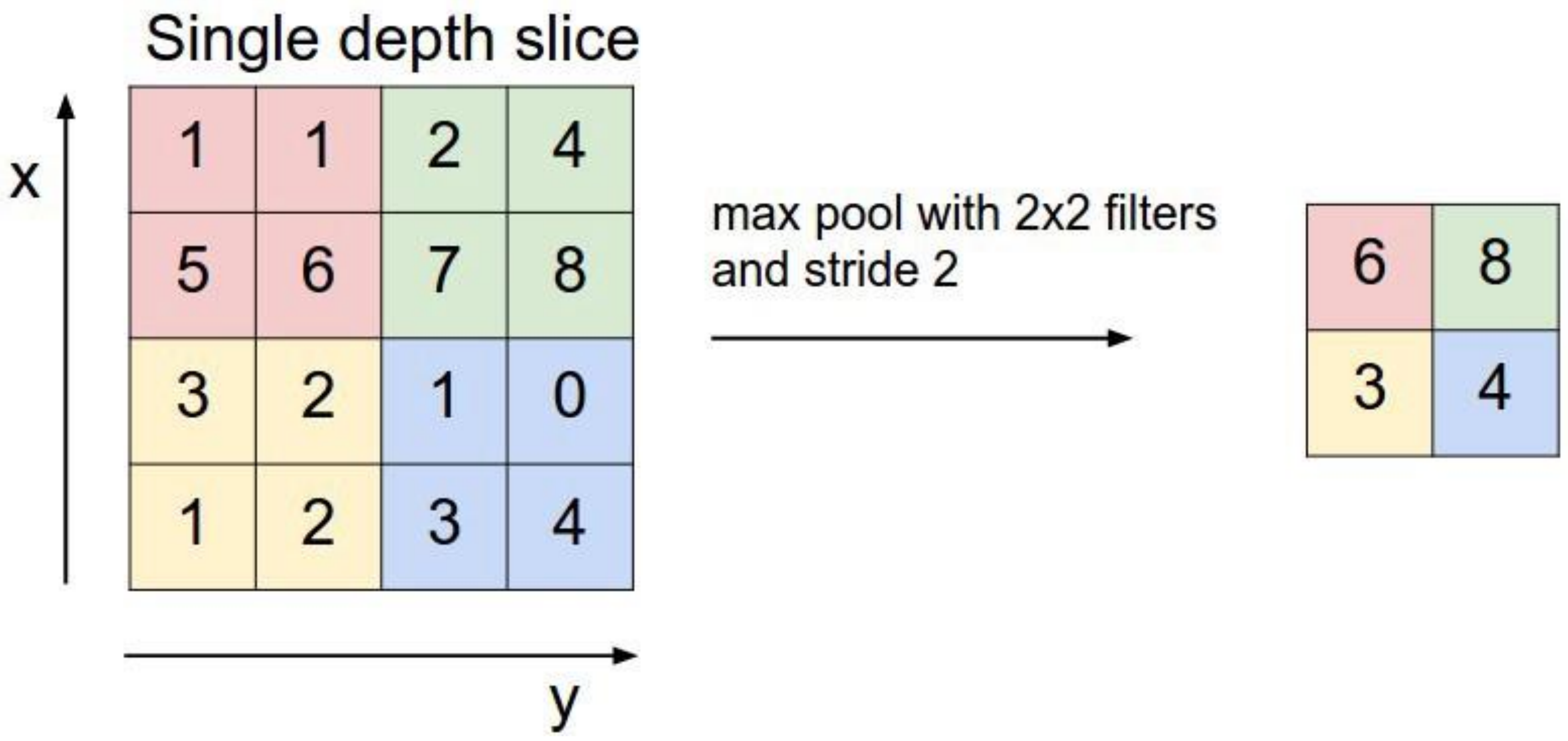
[CNN 계산 방법]

CNN 계산 방법은 숙지를 해놓을 것!



[Max Pooling 계산 방법]

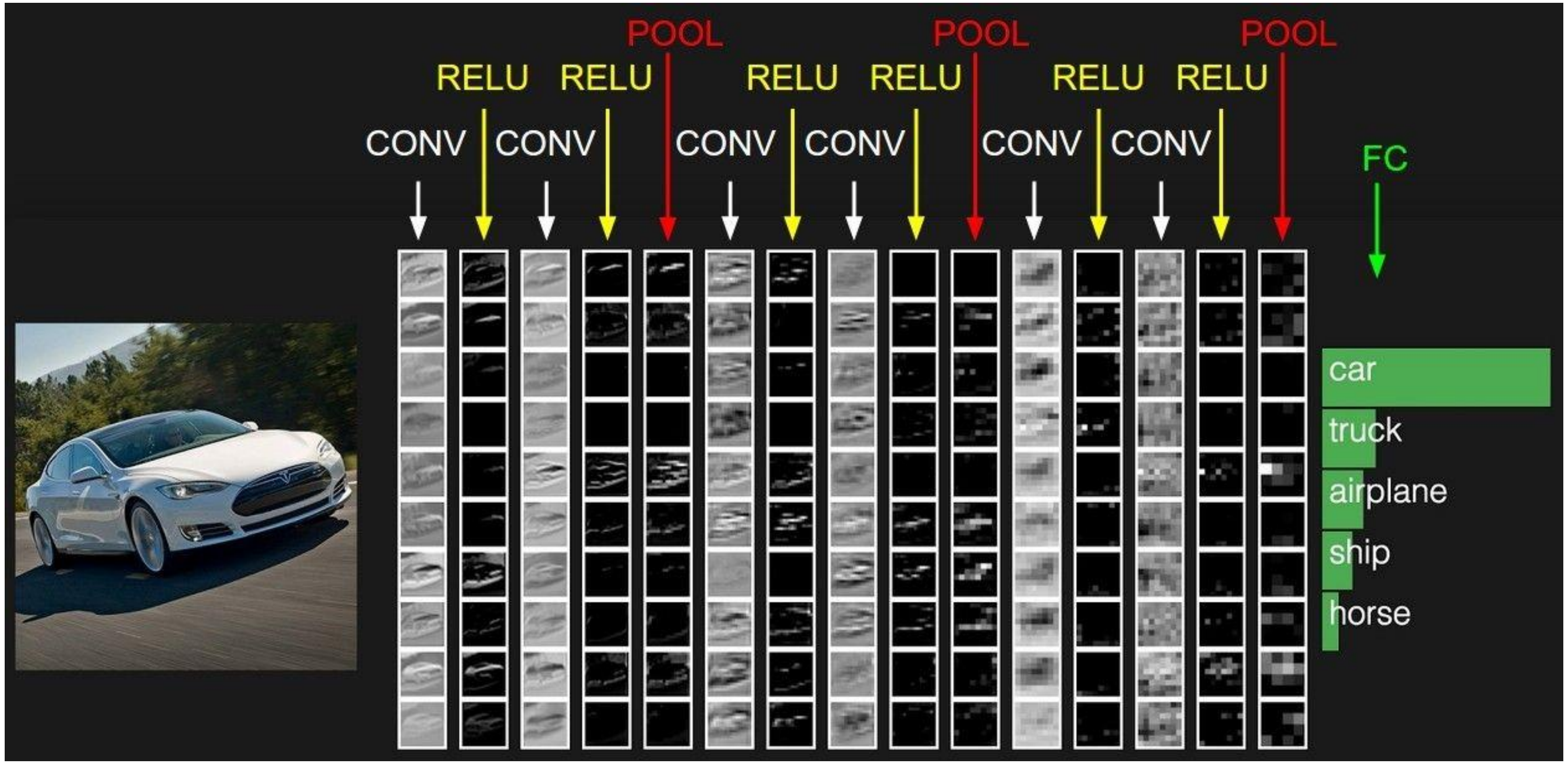
Convolution Layer와 같이 활용되기도 함.



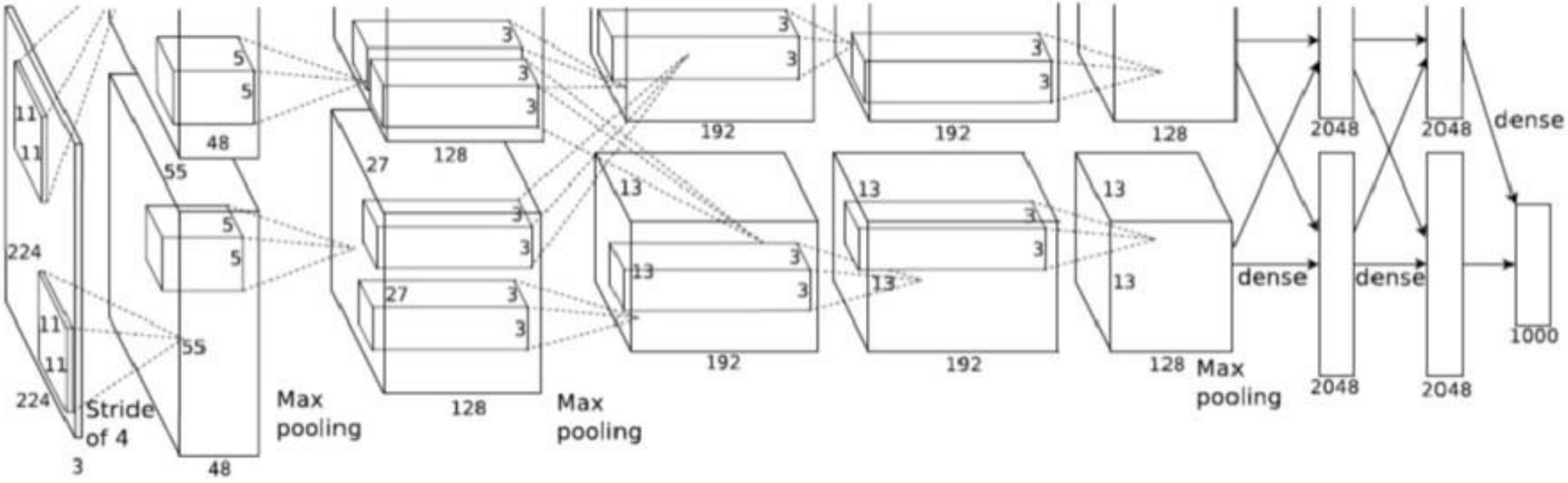
CNN

[CNN 모델 예시]

Conv - Batch Normalization - ReLU - Pooling 순서로 사용됨.



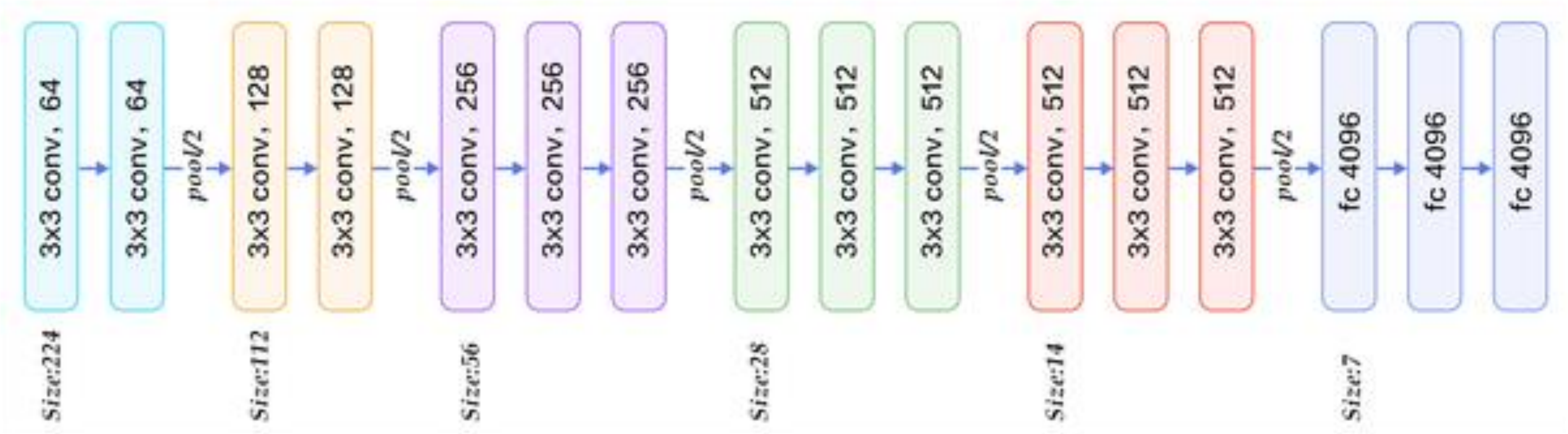
[AlexNet]
CS231n Lecture9에서 자세히



CNN

[VGG16]

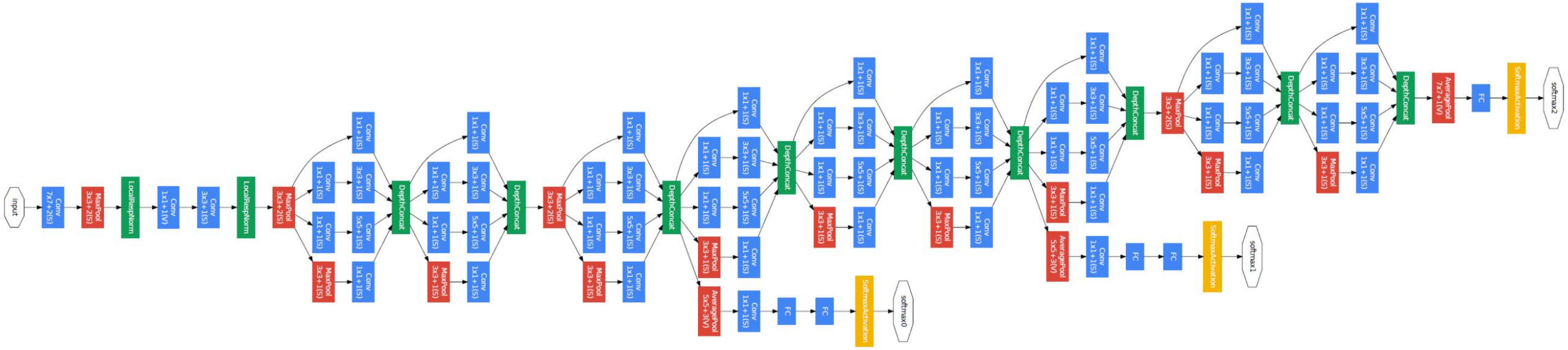
CS231n Lecture9에서 자세히



CNN

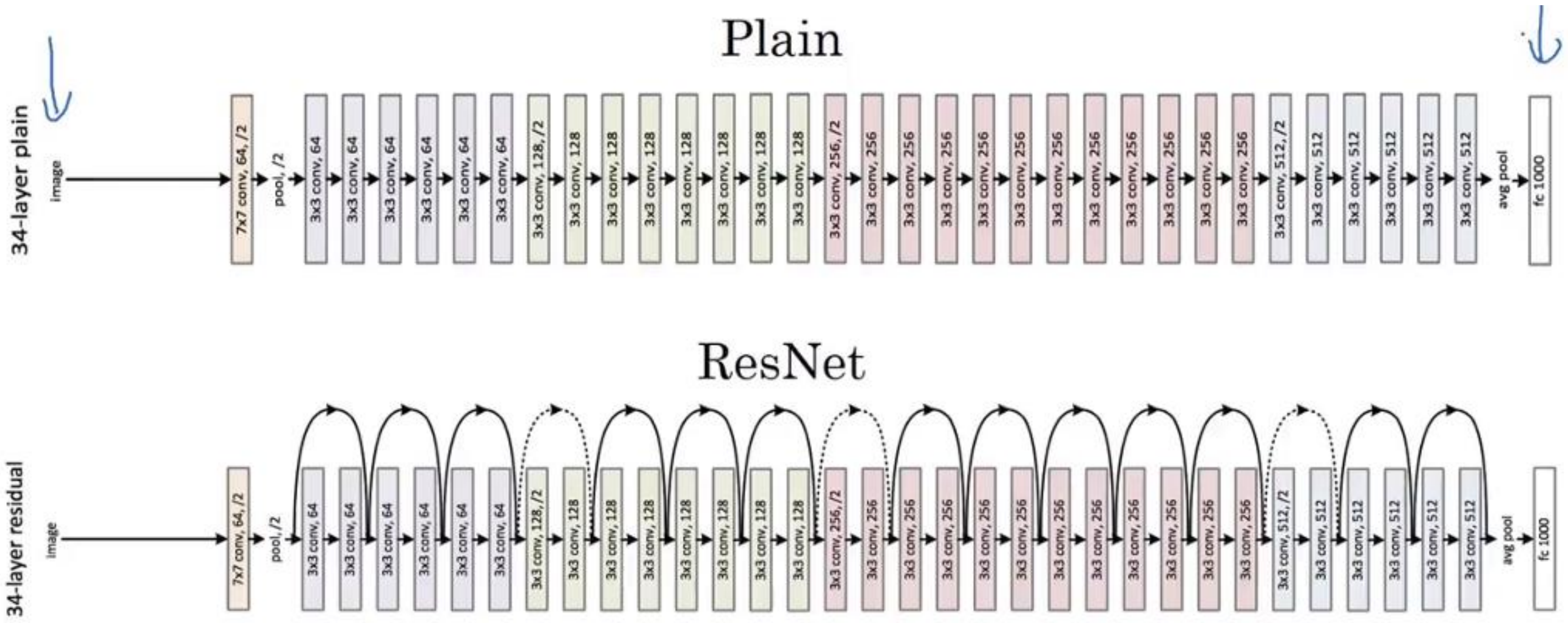
[Inception 또는 GoogleNet]

CS231n Lecture9에서 자세히



[ResNet]

CS231n Lecture9에서 자세히



End

참고 자료

모두의 딥러닝 : https://hunkim.github.io/ml/

Andrew Ng의 Machine Learning : https://ko.coursera.org/learn/machine-learning

CS231n : http://cs231n.stanford.edu/syllabus.html

BOAZ