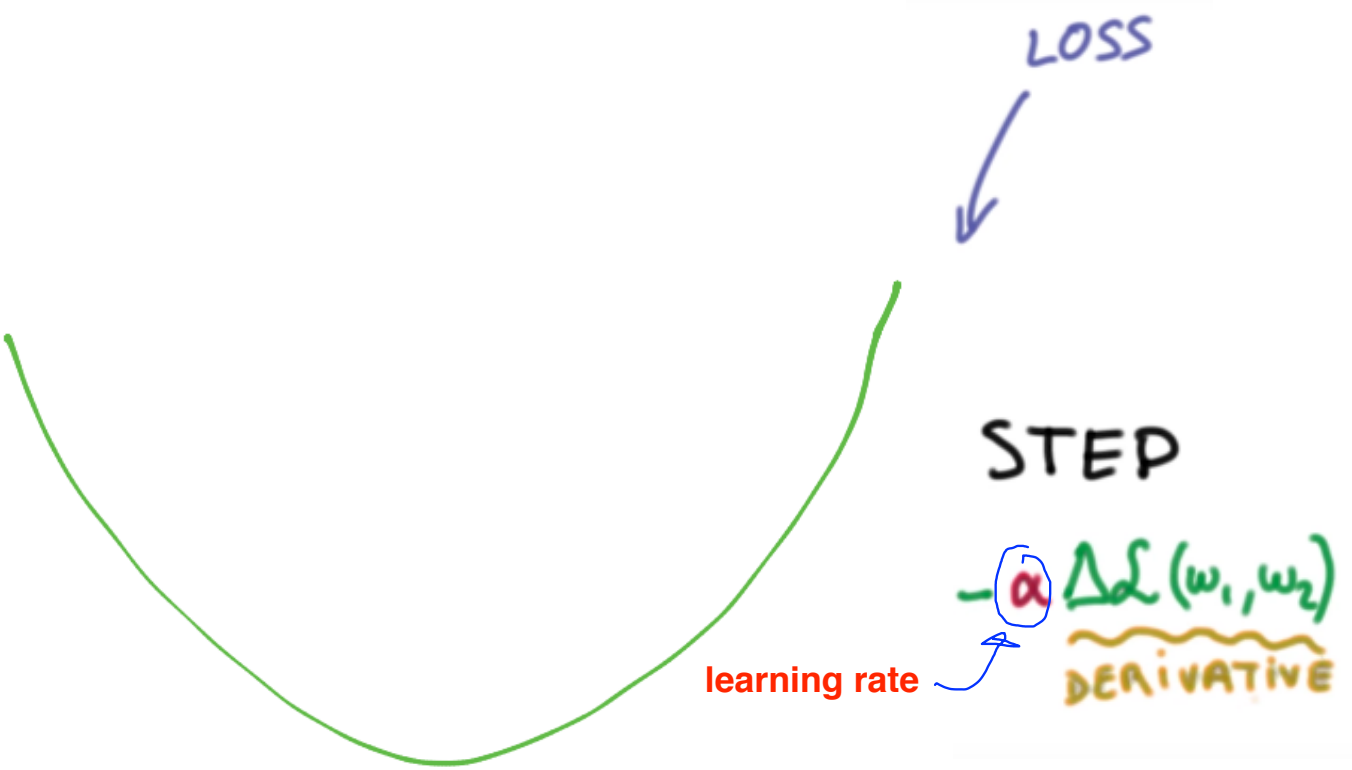
Lecture 7-1

Application & Tips:
Learning rate, data preprocessing, overfitting

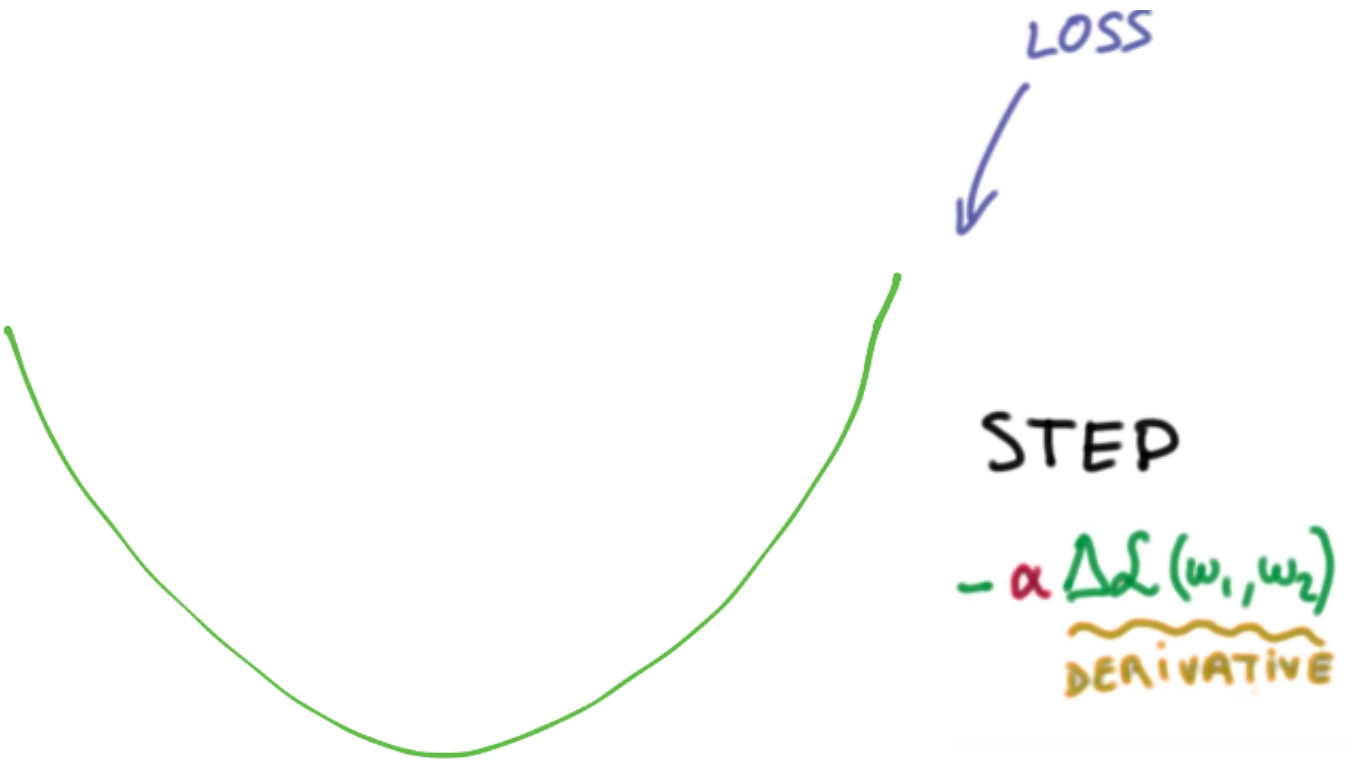
Sung Kim <hunkim+mr@gmail.com>

Gradient descent



https://www.udacity.com/course/viewer#!/c-ud730/l-6370362152/m-6379811827

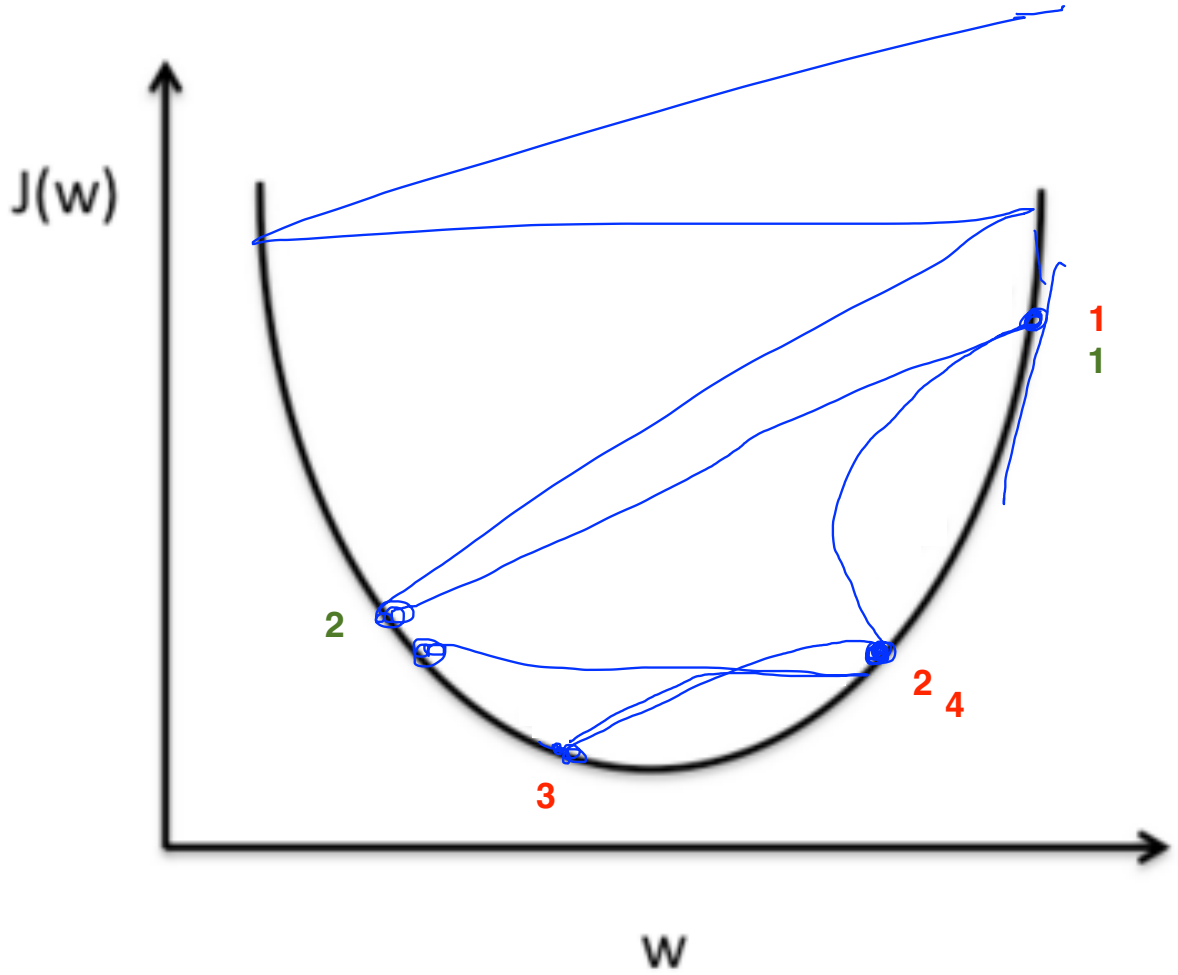
```python
# Minimize error using cross entropy
learning_rate = 0.001
cost = tf.reduce_mean(-tf.reduce_sum(Y*tf.log(hypothesis), reduction_indices=1)) # Cross entropy
optimizer = tf.train.GradientDescentOptimizer(learning_rate).minimize(cost) # Gradient Descent
```



https://www.udacity.com/course/viewer#!/c-ud730/l-6370362152/m-6379811827

경사면을 내려가는 step

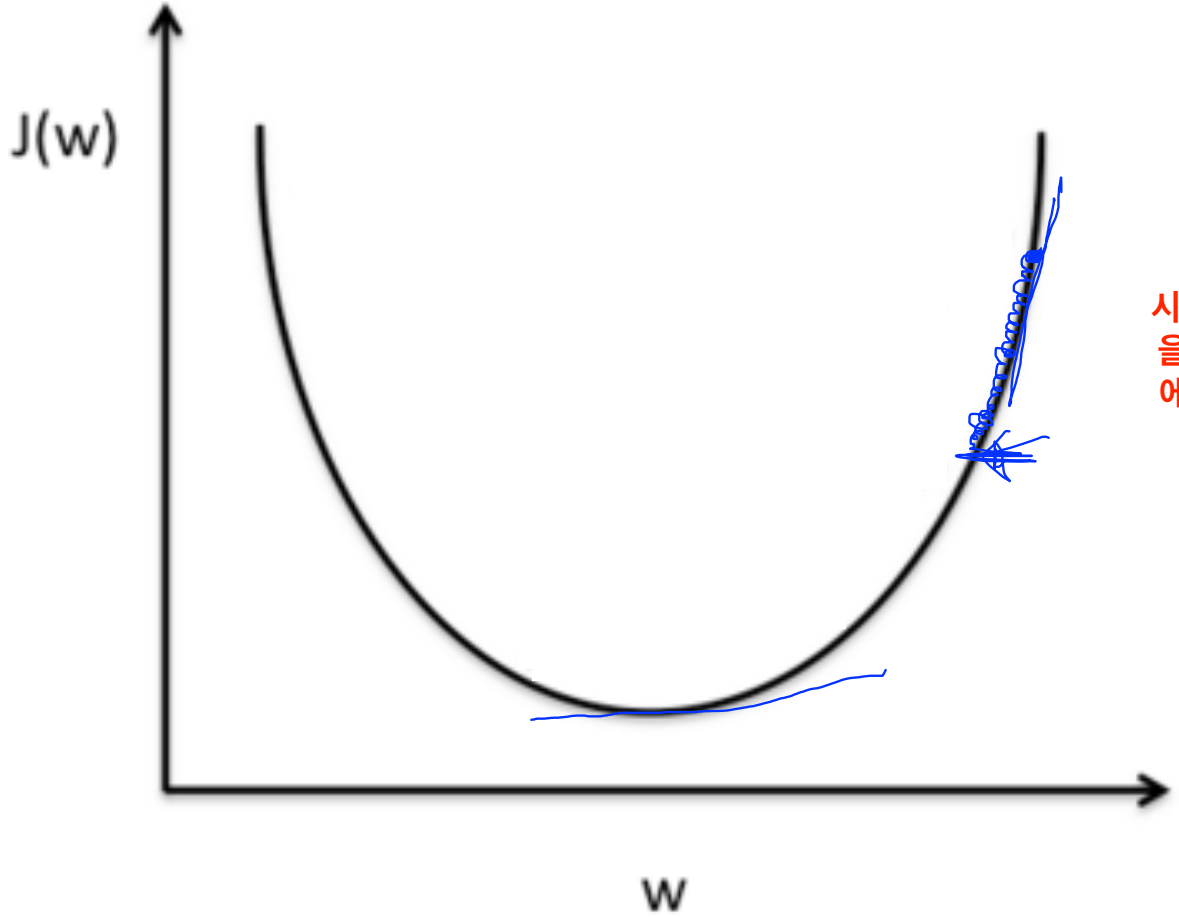
Large learning rate: overshooting



learning rate가 너무 크면 1번에서 시작했다가 2번으로 갔는데, 최저점을 넘어 3번으로 갔다가 다시 4번으로 가서 global minimum을 못 찾을 수도 있다. 또한 녹색과 같이 1번에서 시작했다가 2번으로 간 후 중간을 왔다 갔다 거리다가 튕겨져 나갈 수도 있다.

괭장히 작은 learning rate 을 준 경우

Small learning rate:
takes too long, stops at local minimum



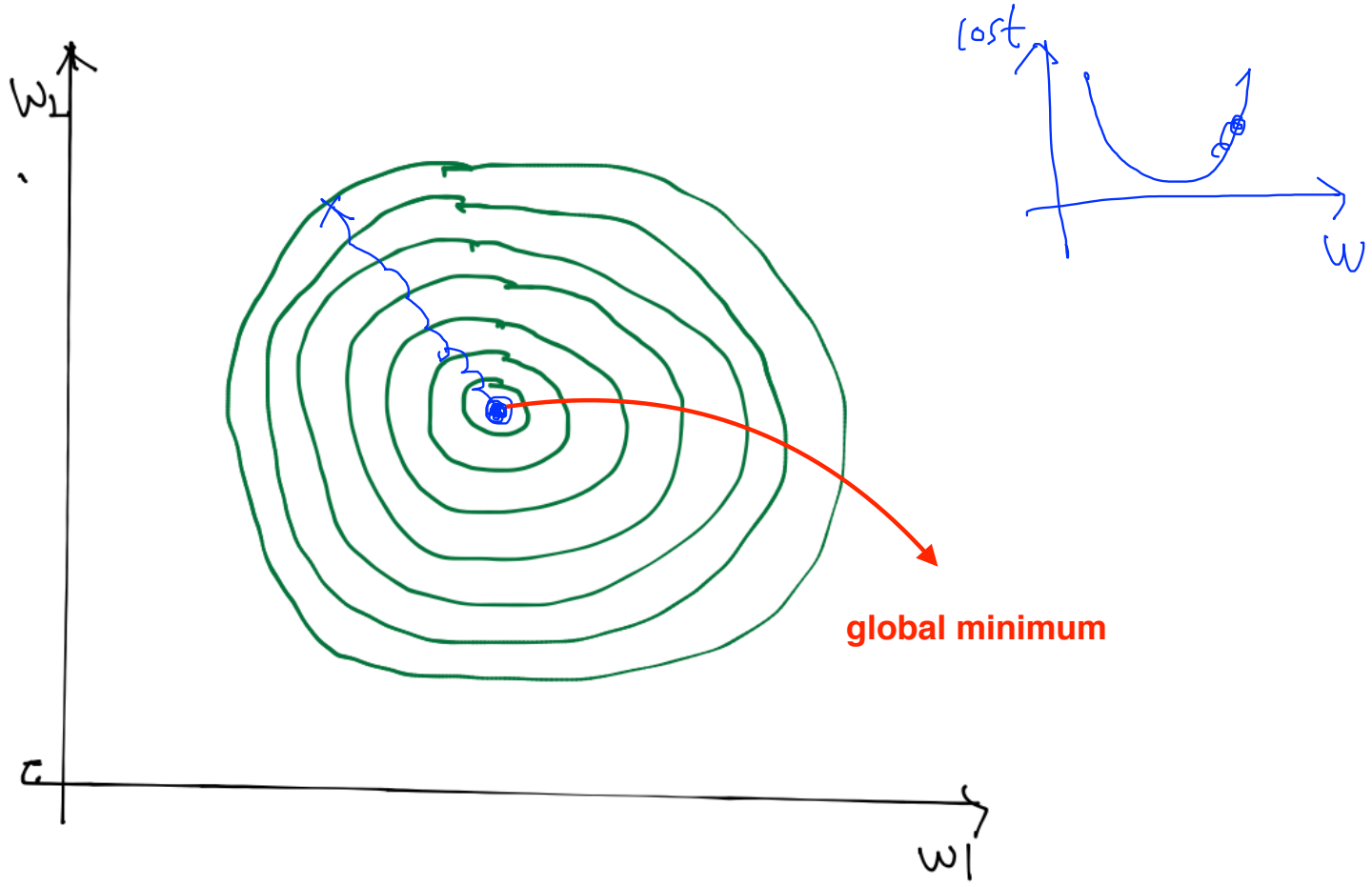
시간이 오래 걸려서 시간에 제한을 준 경우 global minimum 에 도달하지 못했지만 stop 할 수도 있다

Try several learning rates

- Observe the cost function
- Check it goes down in a reasonable rate

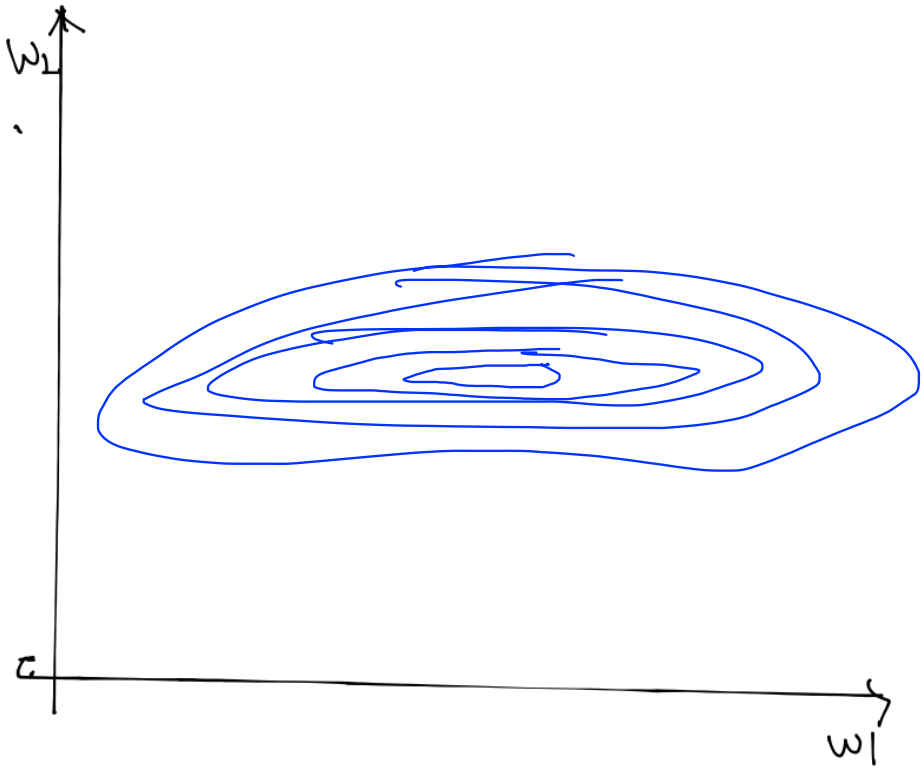
결론, learning rate 을 정하는 데 답이 있는 것은 아니다
작은 값, 큰 값 입력해보면서 확인

Data (X) preprocessing for gradient descent



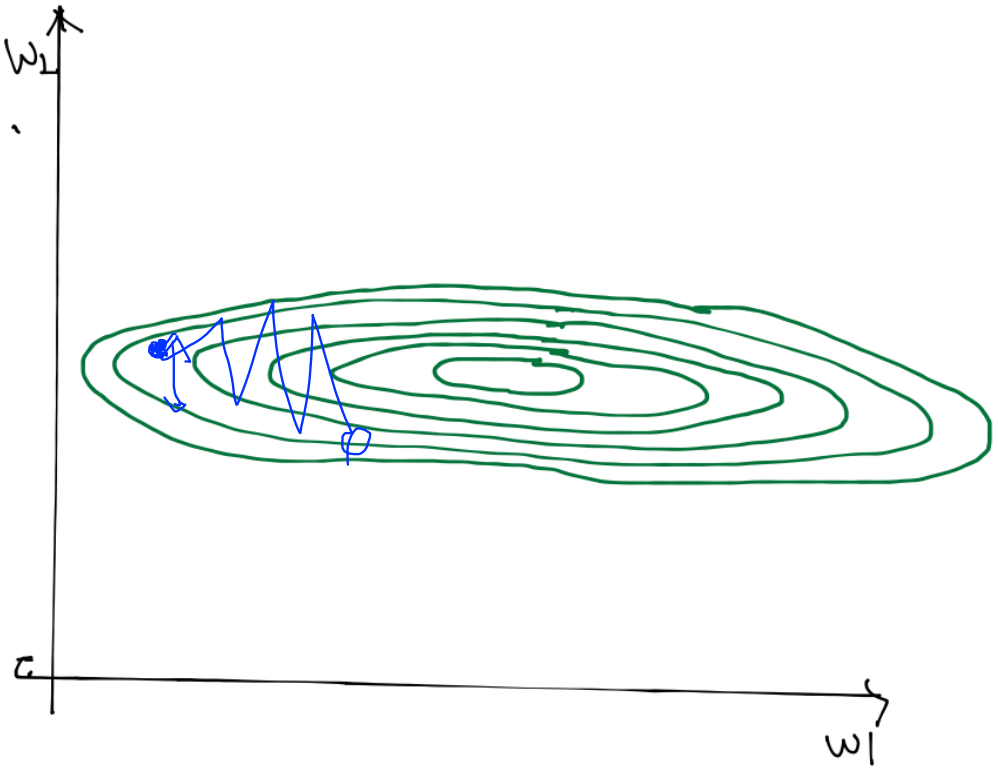
Data (X) preprocessing for gradient descent

x1	x2	y
1	9000	A
2	-5000	A
4	-2000	B
6	8000	B
9	9000	C



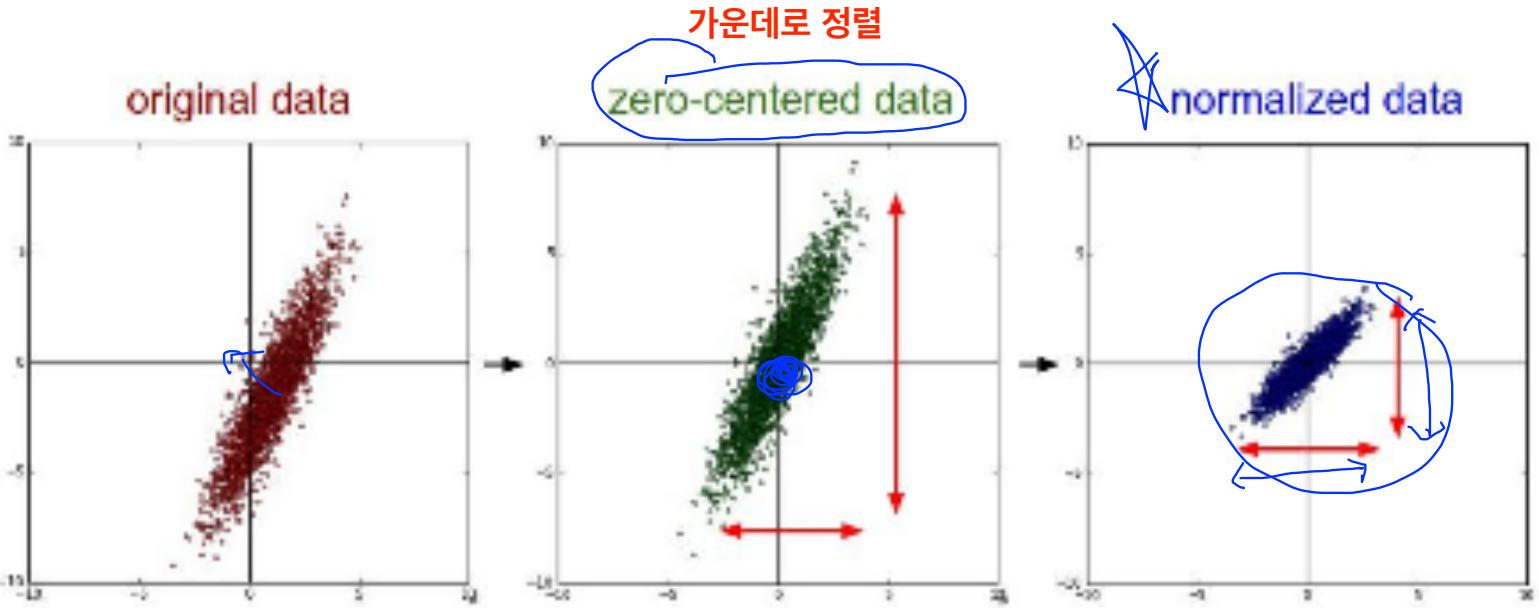
Data (X) preprocessing for gradient descent

x1	x2	y
1	9000	A
2	-5000	A
4	-2000	B
6	8000	B
9	9000	C



이렇게 데이터 값에 큰 차이가 있을 경우, learning rate 을 잘 설정했음에도 불구하고 튀어나갈 수 있다

Data (X) preprocessing for gradient descent



Standardization

$$x'_j = \frac{x_j - \mu_j}{\sigma_j}$$

```
X_std[:,0] = (X[:,0] - X[:,0].mean()) / X[:,0].std()
```

http://sebastianraschka.com/Articles/2015_singlelayer_neurons.html

머신 러닝의 가장 큰 문제

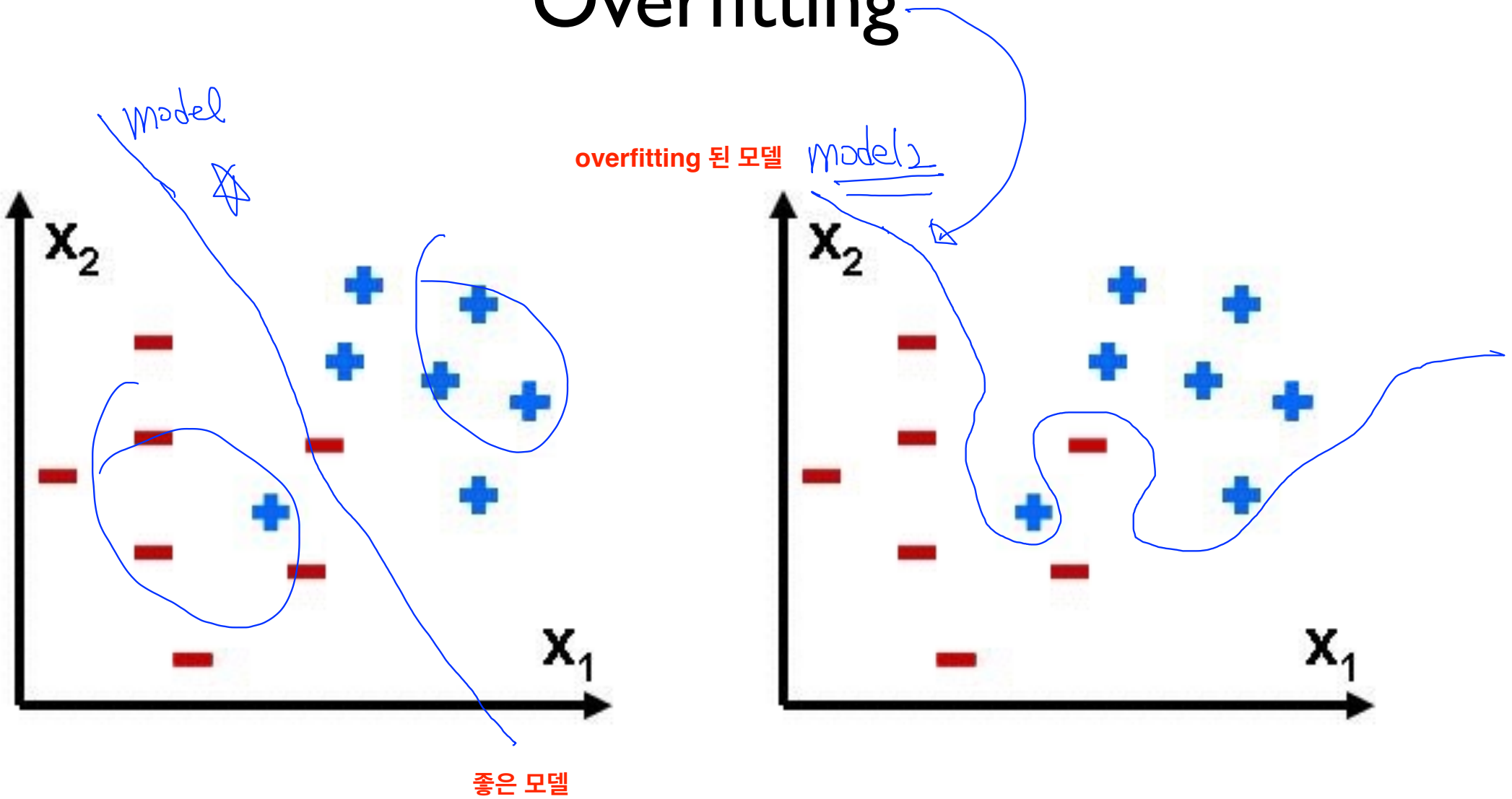
Overfitting

- Our model is very good with training data set (with memorization)
- Not good at test dataset or in real use

학습 데이터

학습 데이터에 너무 잘 맞는 모델이라 실제. 데이터에는 잘 안되는 것

Overfitting

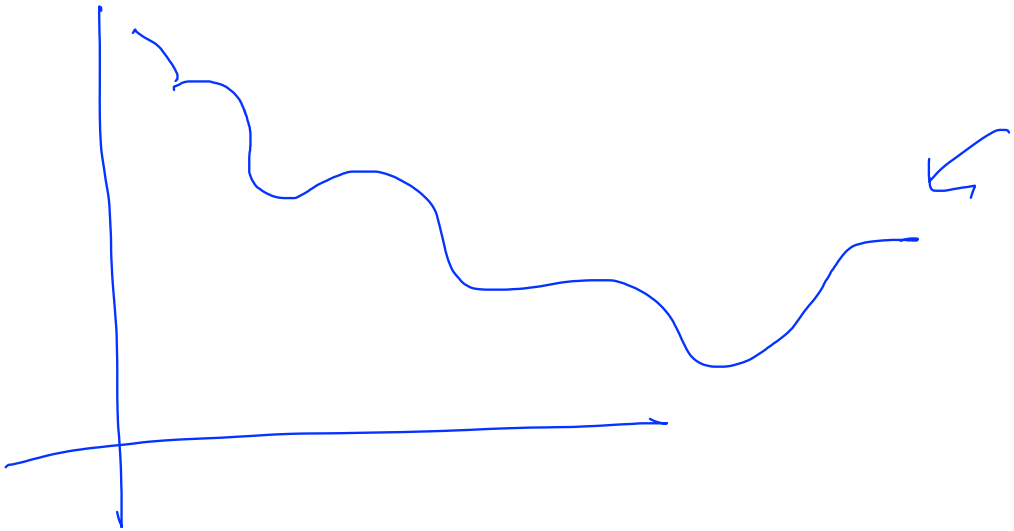


Solutions for overfitting

- More training data! training 데이터를 많이 가지면 가질수록 overfitting 을 줄일 수 있다
- Reduce the number of features
- **Regularization**

Regularization

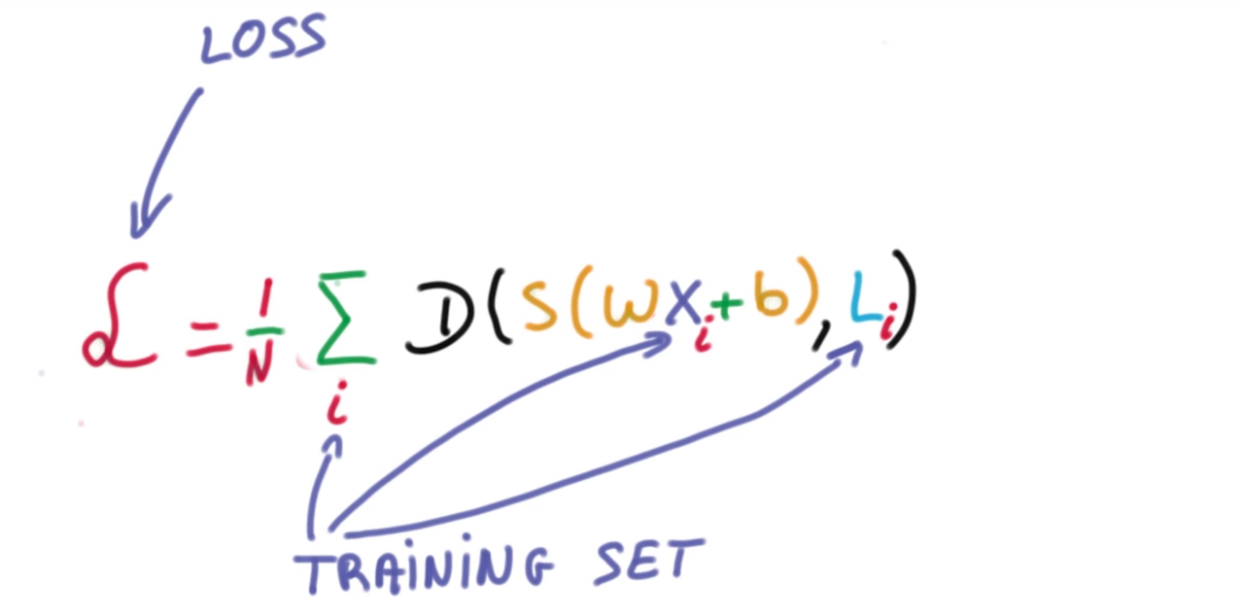
- Let's not have too big numbers in the weight



decision boundary를 데이터에 맞게 구부리는 것을 overfitting 이라고 한다
Weight이 작은 값을 가지면 펴지고, 큰 값을 가지면 구부러진다

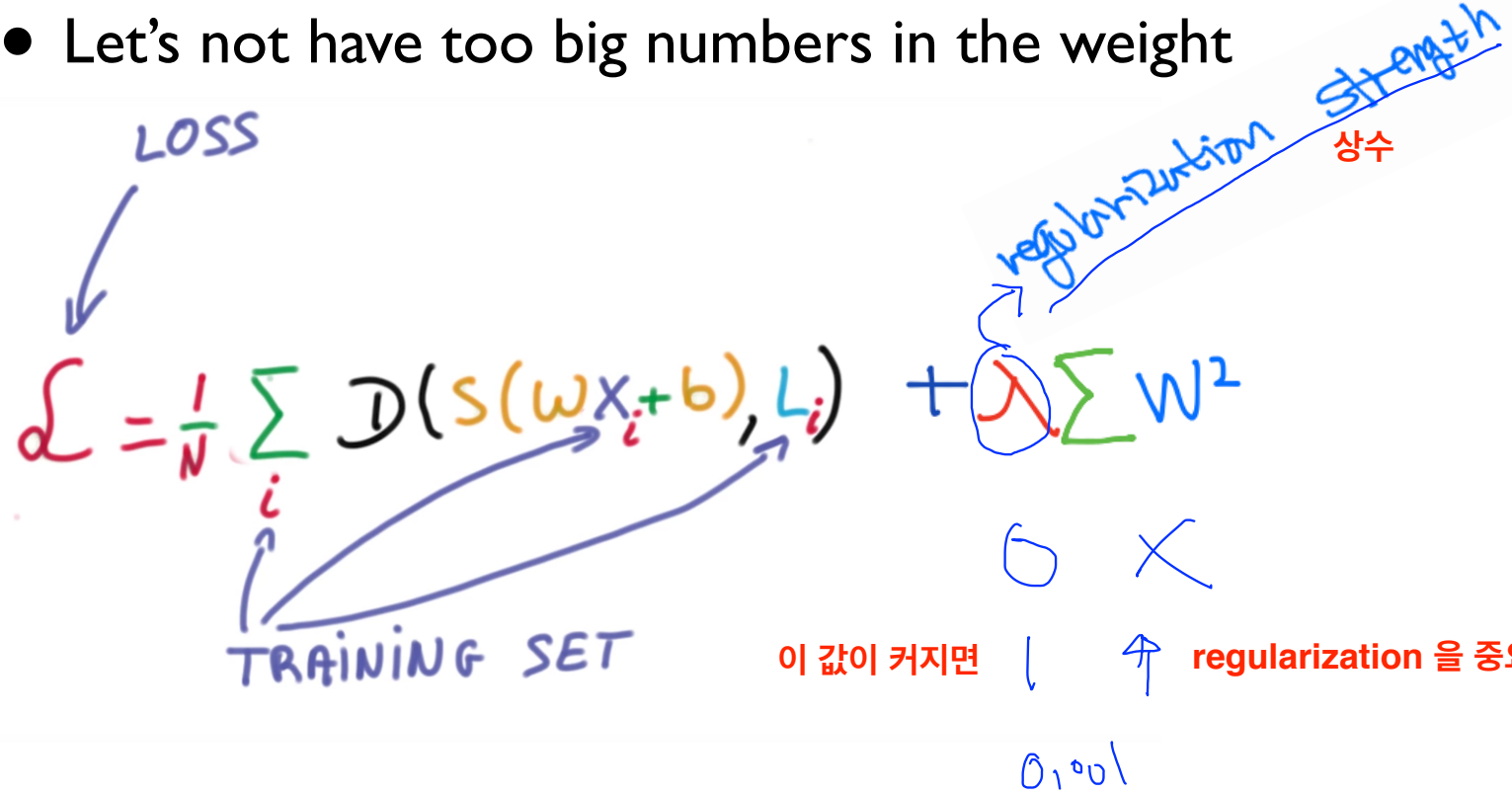
Regularization

- Let's not have too big numbers in the weight



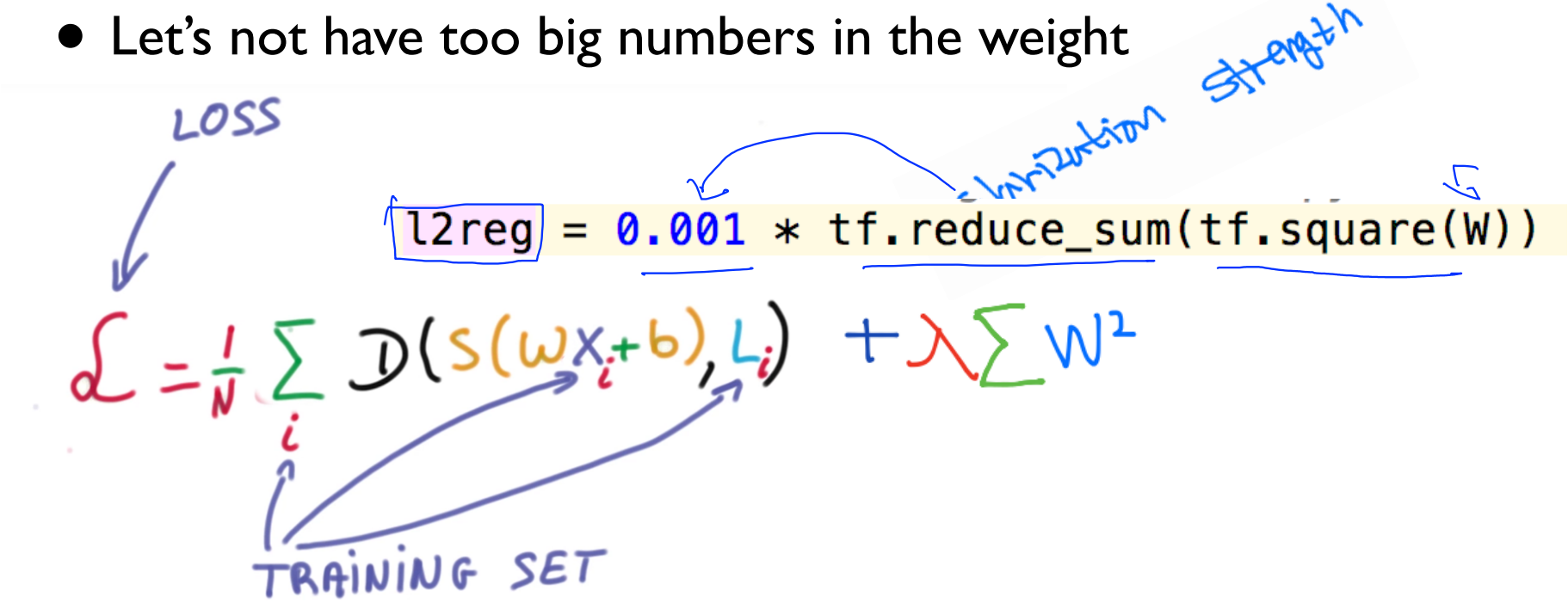
Regularization

- Let's not have too big numbers in the weight



Regularization

- Let's not have too big numbers in the weight



Summary

- Learning rate
- Data preprocessing
- Overfitting
  - More training data
  - Regularization