

딥러닝 입문

기본 용어

목 차

01 Epoch과 Batch Size

02 Gradient Descent

03 Stochastic Gradient Descent

04 Batch Size

05 과적합(Overfitting)

06 과적합 해결하기

01 Epoch과 Batch size

▶ 둘 다 정수의 값을 갖는다.

▶ 배치 크기(Batch size)는 모델의 내부 매개 변수가 업데이트 되기 전의 처리할 샘플 수를 정의하는 변수(파라미터)이다.

▶ epoch의 수는 학습 알고리즘이 전체 학습 데이터 세트에 대해 동작하는 횟수를 정의하는 파라미터

▶ 하나의 에포크(epoch)는 하나 이상의 배치로 구성.

02 Gradient Descent

- ▶ GD(Gradient Descent)는 반복적인 알고리즘이다.
- ▶ 가장 좋은 결과를 찾기 위한 머신러닝을 사용하는 알고리즘
- ▶ Gradient는 경사의 기울기 또는 경사의 비율을 의미
- ▶ Gradient descent는 Learning rate(학습률)의 하나의 파라미터를 갖는다.
- ▶ Cost 함수는 감소된다. 때때로 Cost를 Loss라고 이야기를 하기도 한다.

02 Gradient Descent

▶ 업데이트는 모델에 따라 다르지만 인공 신경망의 경우,
역전파 업데이트 알고리즘이 사용된다..

▶ 배치 사이즈에 따라 다음과 같이 분류된다.

Batch Gradient Descent **Batch Size** = **훈련 데이터의 크기**

Stochastic Gradient Descent **Batch Size** = 1

Mini-Batch Gradient Descent $1 < \text{Batch Size} < \text{훈련 데이터 크기}$

03 Stochastic Gradient Descent

- ▶ SGD(Stochastic Gradient Descent)는 최적화 알고리즘이다.
- ▶ 알고리즘의 역할은 로그 손실 또는 평균 제곱 오차와 같은 성능 측정에 대해 잘 수행되는 내부 모델 매개 변수를 찾는 것.
- ▶ 알고리즘의 반복적이다. 여러 개의 개별 단계를 걸쳐 모델 매개변수가 조금씩 향상된다.
- ▶ 매단계에서는 모델을 이용하여 예측하고 실제 결과와 비교하고, 오차를 계산하고 오차를 이용하여 내부 모델의 매개 변수를 업데이트 한다.

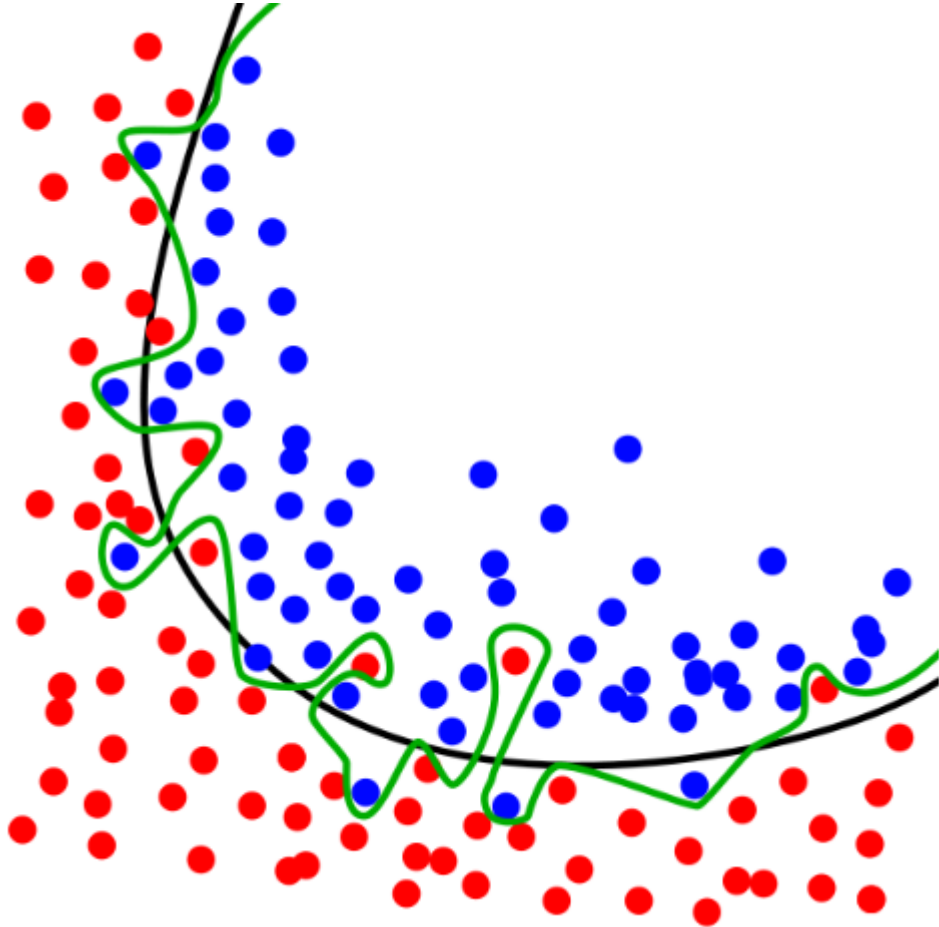
04 Batch Size

▶ mini-batch gradient descent의 경우, 보통 많이 쓰이는 배치는 32, 64, 128이 사용.

▶ 배치 사이즈에 학습 프로세스에 미치는 영향

<https://machinelearningmastery.com/how-to-control-the-speed-and-stability-of-training-neural-networks-with-gradient-descent-batch-size/>

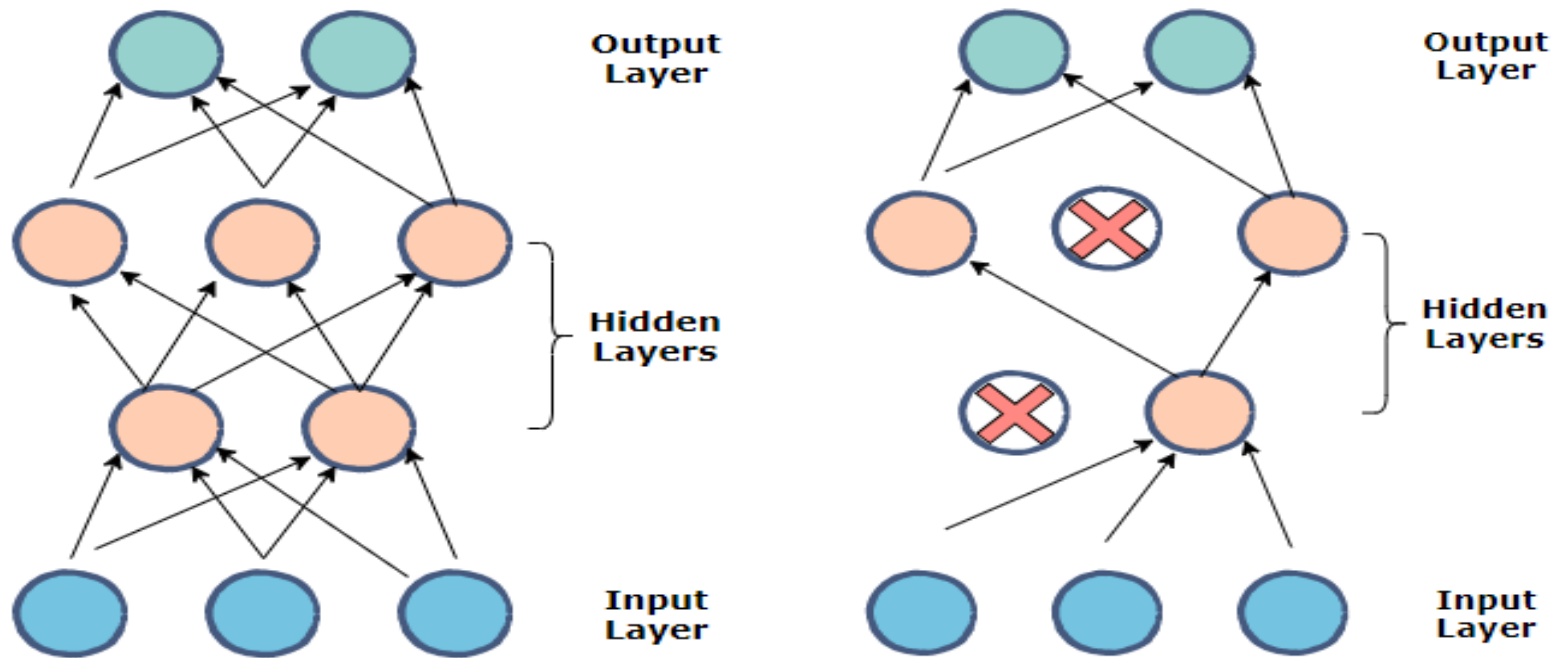
05 과적합(Overfitting)



학습용 데이터에만 넘 잘 맞는 것.

06 과적합 해결하기

▶ Dropout – 과적합 해결



왼쪽이 원래 신경망, 오른쪽에 Dropout된 신경망

06 과적합 해결하기

- ▶ Batch Normalization(배치 정규화)
- ▶ Data Augmentation

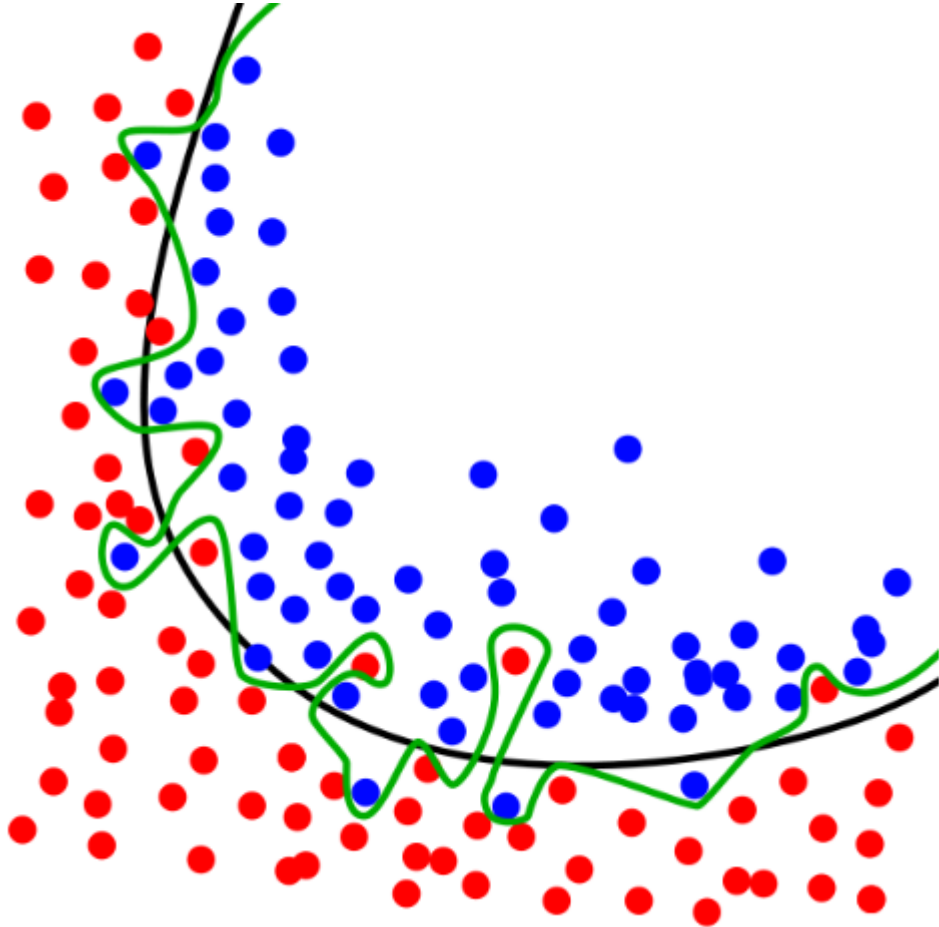
04 Batch Size

▶ mini-batch gradient descent의 경우, 보통 많이 쓰이는 배치는 32, 64, 128이 사용.

▶ 배치 사이즈에 학습 프로세스에 미치는 영향

<https://machinelearningmastery.com/how-to-control-the-speed-and-stability-of-training-neural-networks-with-gradient-descent-batch-size/>

Overfitting(과적합)



학습용 데이터에만 넘 잘 맞는 것.