

즐거운
프로그래밍
경험



with 파이썬

길벗

- 파이썬을 처음 시작하는 사람들의 눈높이에서 원리를 설명한다! ● 코랩, 케라스를 활용하여 간편하게 실습한다!
- 숫자 인식, 코로나 확진자 수 예측, 숫자 생성 인공지능, 모델을 직접 만들어 본다!

UNIT07

텐서플로 플레이그라운드로 딥러닝 체험하기

모두의
인공지능

이영호 지음

1. 텐서플로 플레이그라운드 접속하기

1. 텐서플로 플레이그라운드 접속하기 <http://playground.tensorflow.org/>

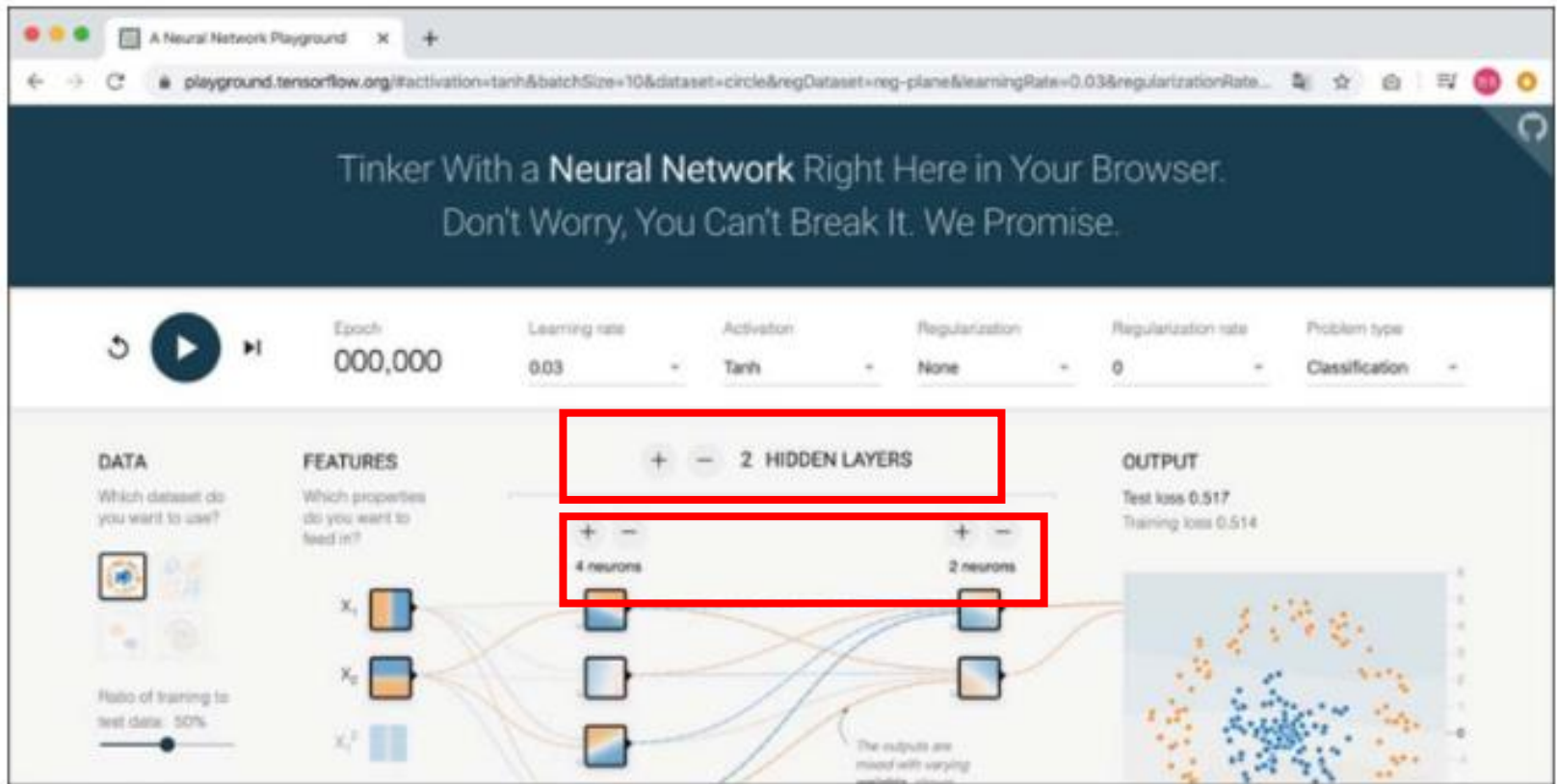


그림 7-1 | 텐서플로 플레이그라운드 접속



2. 시작 버튼 누르기

- 시작 버튼을 누르면 애니메이션 효과가 나오며, 이런 저런 값들이 변하는 모습을 볼수있다.
- 에포크(Epoch)라고 적힌 숫자의 값이 늘어남
 - 에포크- 전체 데이터를 사용하여 인공 신경망이 학습한 횟수를 의미

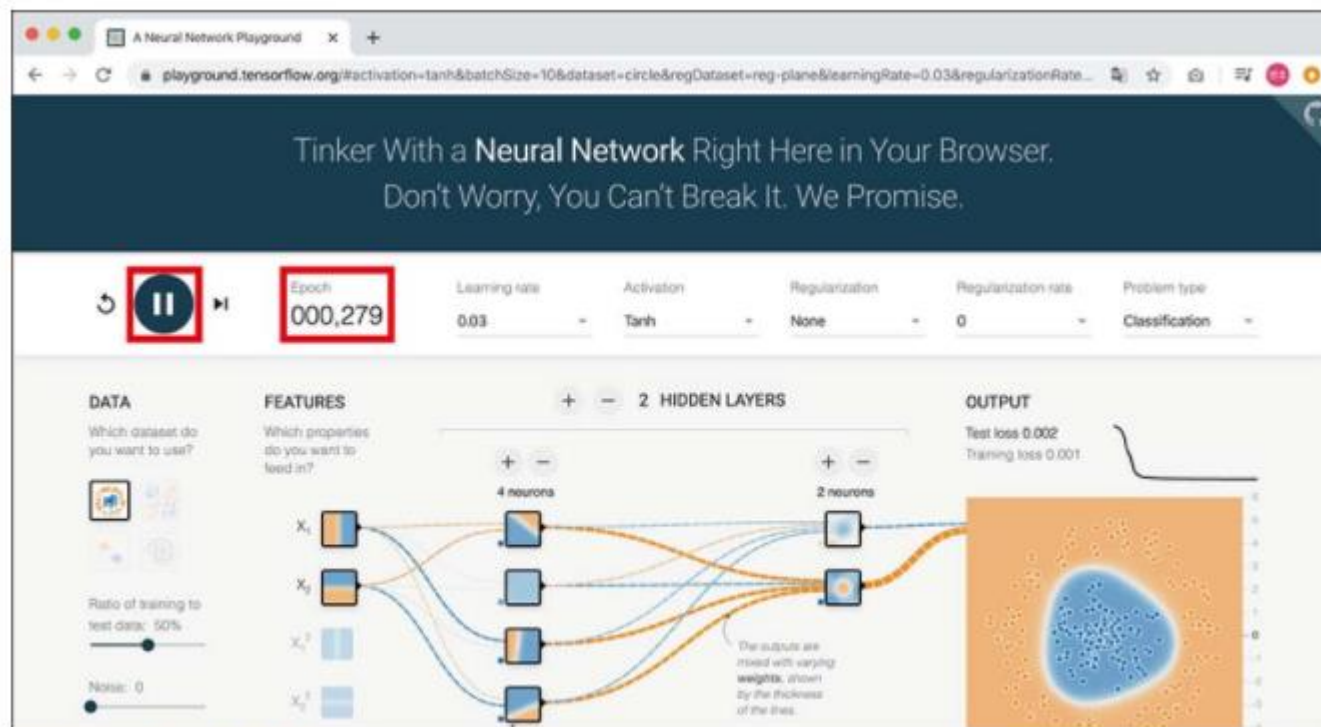
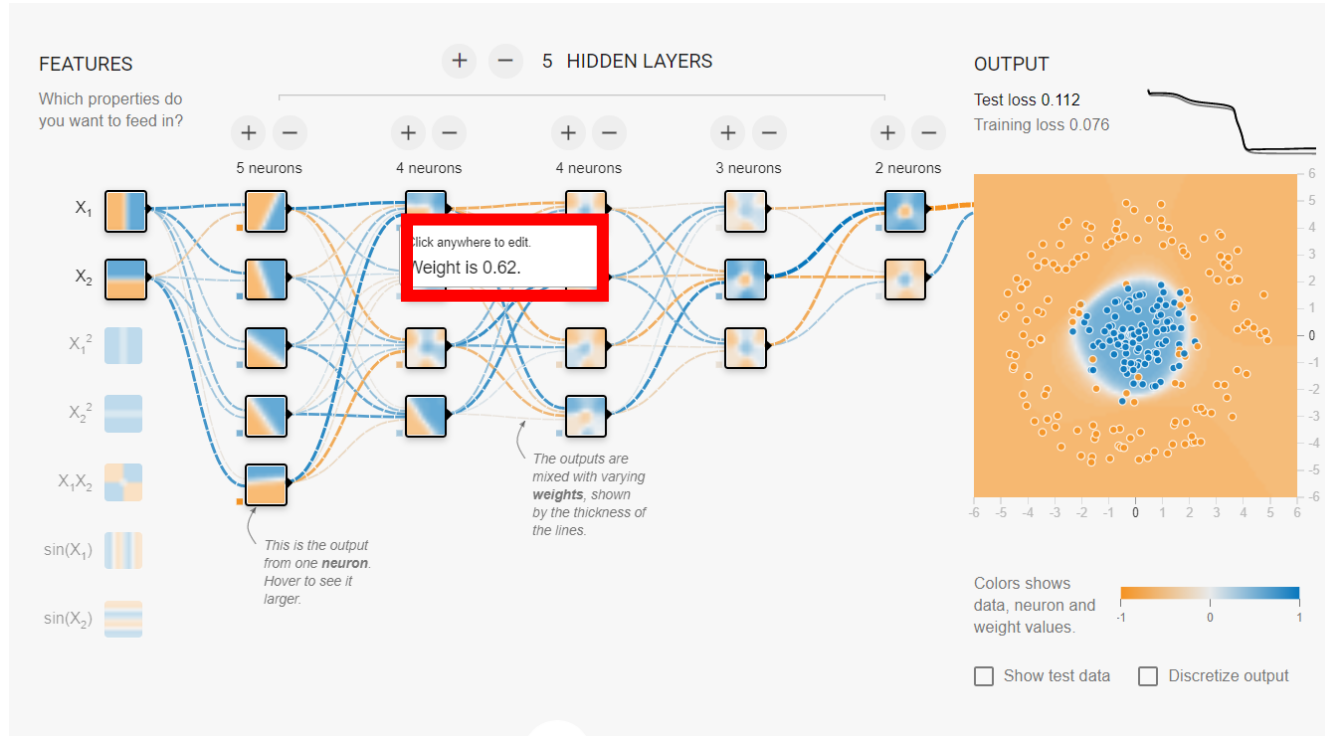


그림 7-2 | 시작 버튼을 클릭



선들의 가중치



잠깐만요

뉴런과 뉴런이 연결된 선은 무엇인가요?

이 선들이 바로 가중치입니다. 그 선에 마우스 포인터를 가져가면 특정한 숫자가 나오는 것을 볼 수 있습니다. 그 숫자들이 가중치 값이며, 가중치 값이 클수록 선의 두께 또한 두꺼워집니다.



3. 출력 부분 살펴보기



오차가 0에 가까움!

파란색 점과 주황색 점이
잘 구분됨!!

그림 7-3 | 출력 부분

TIP

계속 반복 학습을 하다가 주황색과 파란색의 구분이 명확해지기 시작한다면 중단 버튼을 눌러도 됩니다. 주황색과 파란색을 잘 구분하도록 학습이 잘 되었다는 의미이기 때문이죠.



훈련 데이터와 검증 데이터 구분

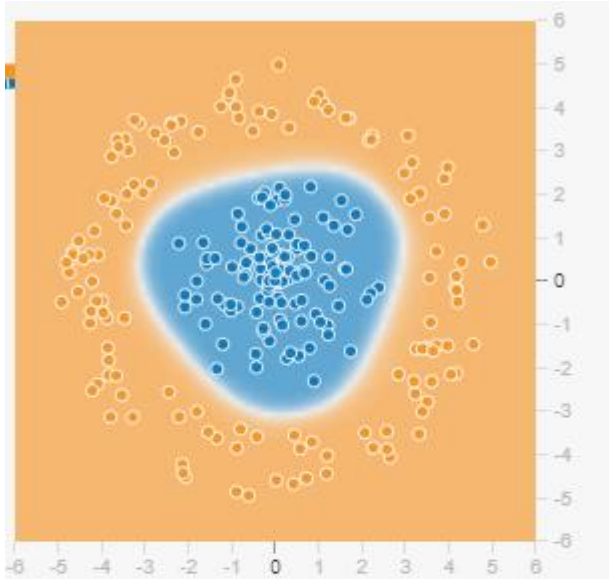
OUTPUT

Test loss 0.002

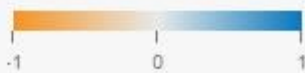
Training loss 0.000

검증데이터의 오차값

훈련데이터의 오차값

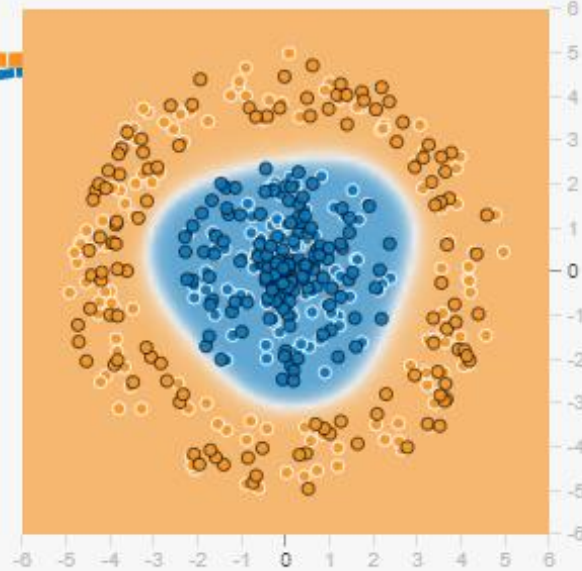


Colors shows
data, neuron and
weight values.

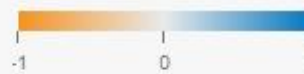


☐ Show test data

☐ Discretize output



Colors shows
data, neuron and
weight values.

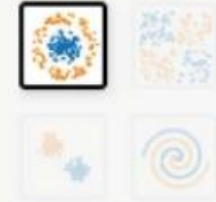


☒ Show test data

☐ Discretize output

DATA

Which dataset do
you want to use?



Ratio of training to
test data: 50%



Noise: 0



Batch size: 10



REGENERATE

검증 데이터가 훈련 데이터에 비해 너무 많다고요?

화면의 왼쪽 DATA 영역에서 그 비율을 수정할 수 있습니다. Ratio of training to test data 영역에서 비율을 조정하면 됩니다. 실제 인공지능 개발에서는 검증 데이터의 비율을 20~30% 정도로 한답니다.



4. 신경망 구조 설계하기

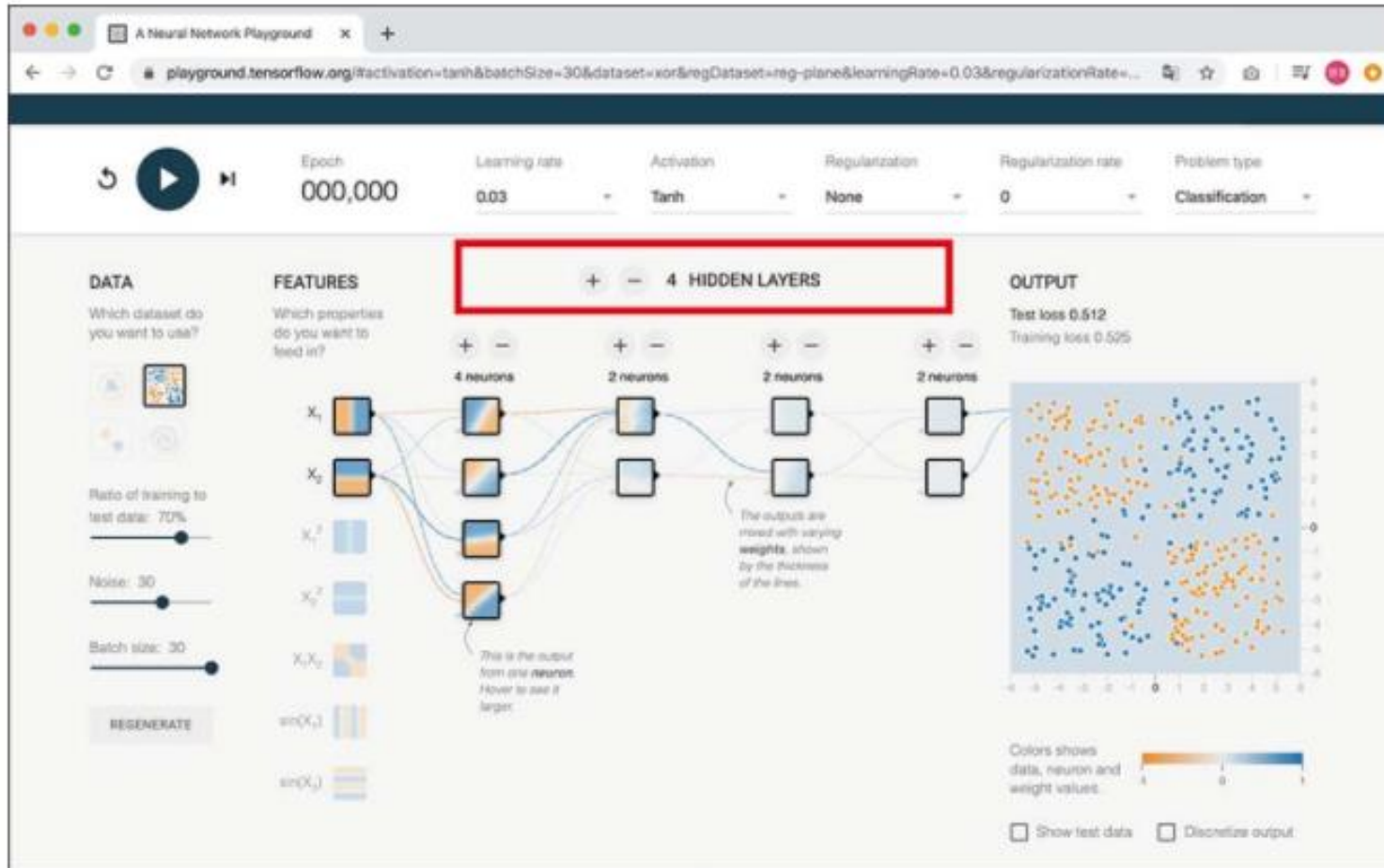


그림 7-5 | 신경망 층의 개수 조정



4. 신경망 구조 설계하기

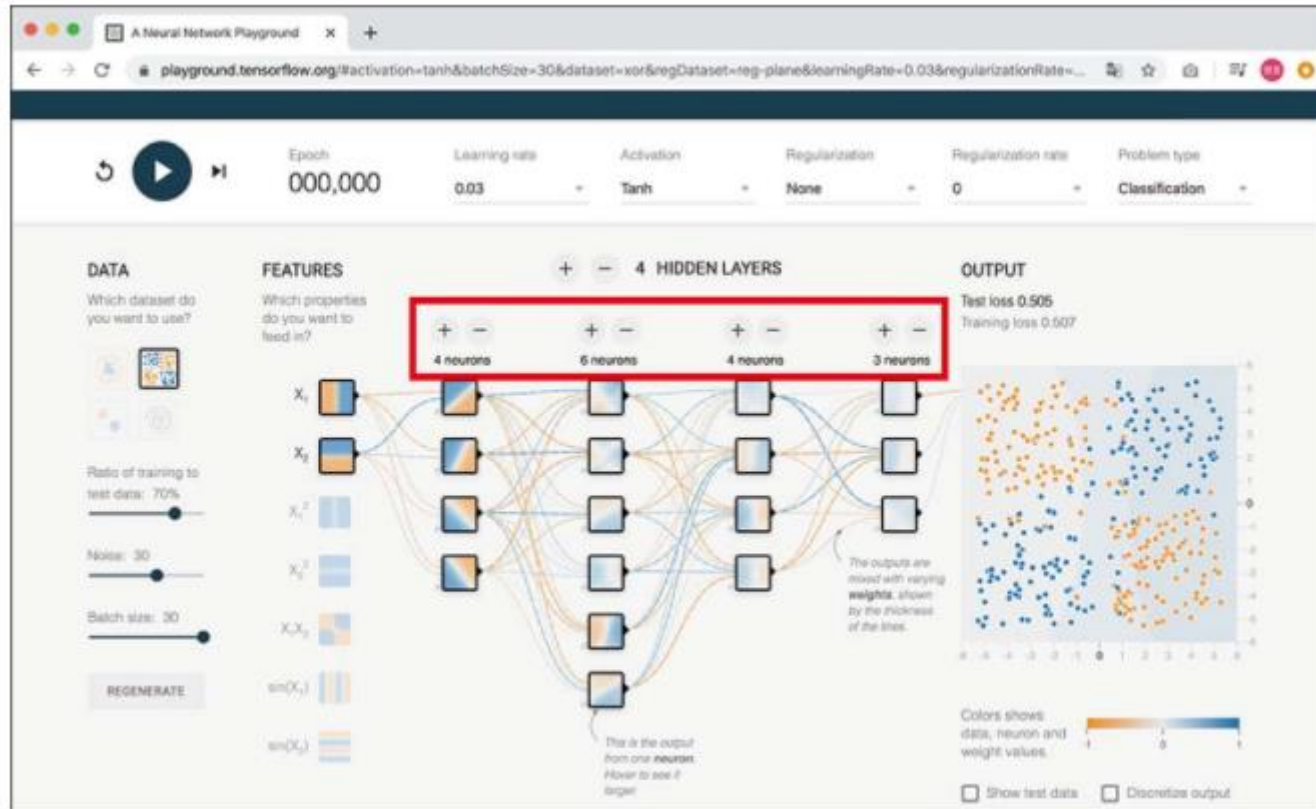


그림 7-6 | 뉴런의 수를 조정




5. 데이터 입력 형태 선택하기


- 화면 왼쪽의 Data에서 데이터의 입력 형태를 선택
- 기본적으로 x축과 y축의 값 입력
- 데이터의 형태, 검증 데이터와 훈련 데이터의 비율, 노이즈, 배치 사이즈, 에포크 등의 값을 수정한 후 regenerate(재생성)버튼 클릭

DATA


Which dataset do you want to use?




Ratio of training to test data: 50%



Noise: 0



Batch size: 10



REGENERATE

그림 7-7 | 데이터의 입력 형태를 선택 후 REGENERATE(재생성) 버튼 클릭



잠깐만요

관련 용어를 정리해요!

- ① Ratio of training to test data: 훈련 데이터와 검증 데이터를 나누는 비율을 의미합니다. 검증 데이터가 너무 많거나 너무 적어도 신경망이 학습할 때 문제가 될 수 있습니다. 보통 신경망을 훈련시킬 때 훈련 데이터와 검증 데이터의 비율을 7:3 혹은 8:2로 합니다. 하지만 비율에 정답이 있는 것은 아니니 여러 비율로 테스트해 보세요.
- ② Noise: 분류 문제에서 실제 데이터는 정확하게 두 부분으로 나뉘지 않습니다. 데이터가 서로 섞여 있기 마련이죠. Noise가 0일 때에는 섞여 있지 않을 때를 말하며, Noise가 높아질수록 데이터가 섞여 있는 정도가 심해집니다. 그러면 신경망을 학습시킬 때 모델 설계를 더 잘해야 합니다.
- ③ Batch Size: 신경망을 학습시킬 때 한 번에 학습하는 데이터의 양을 말합니다. 만약 데이터가 100개가 있을 때 Batch Size(배치 사이즈)가 100이라면 한 번 학습할 때 데이터 100개를 한꺼번에 학습시킨다는 의미입니다. 만약 Batch Size가 10이라면 한 번 학습할 때 데이터 10개를 사용하겠죠.
- ④ Epoch: Epoch(에포크)란 전체 데이터를 한 번 학습하는 것을 의미합니다. 만약 50Epoch라면 전체 데이터를 50번 학습한다는 의미입니다.

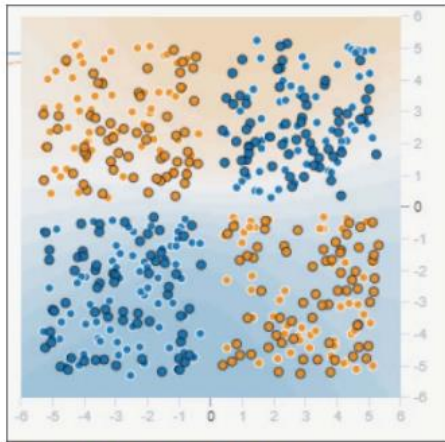


그림 7-8 | 사분면 모습의 두 번째 데이터

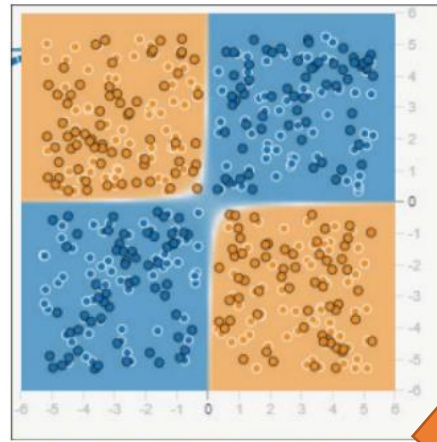
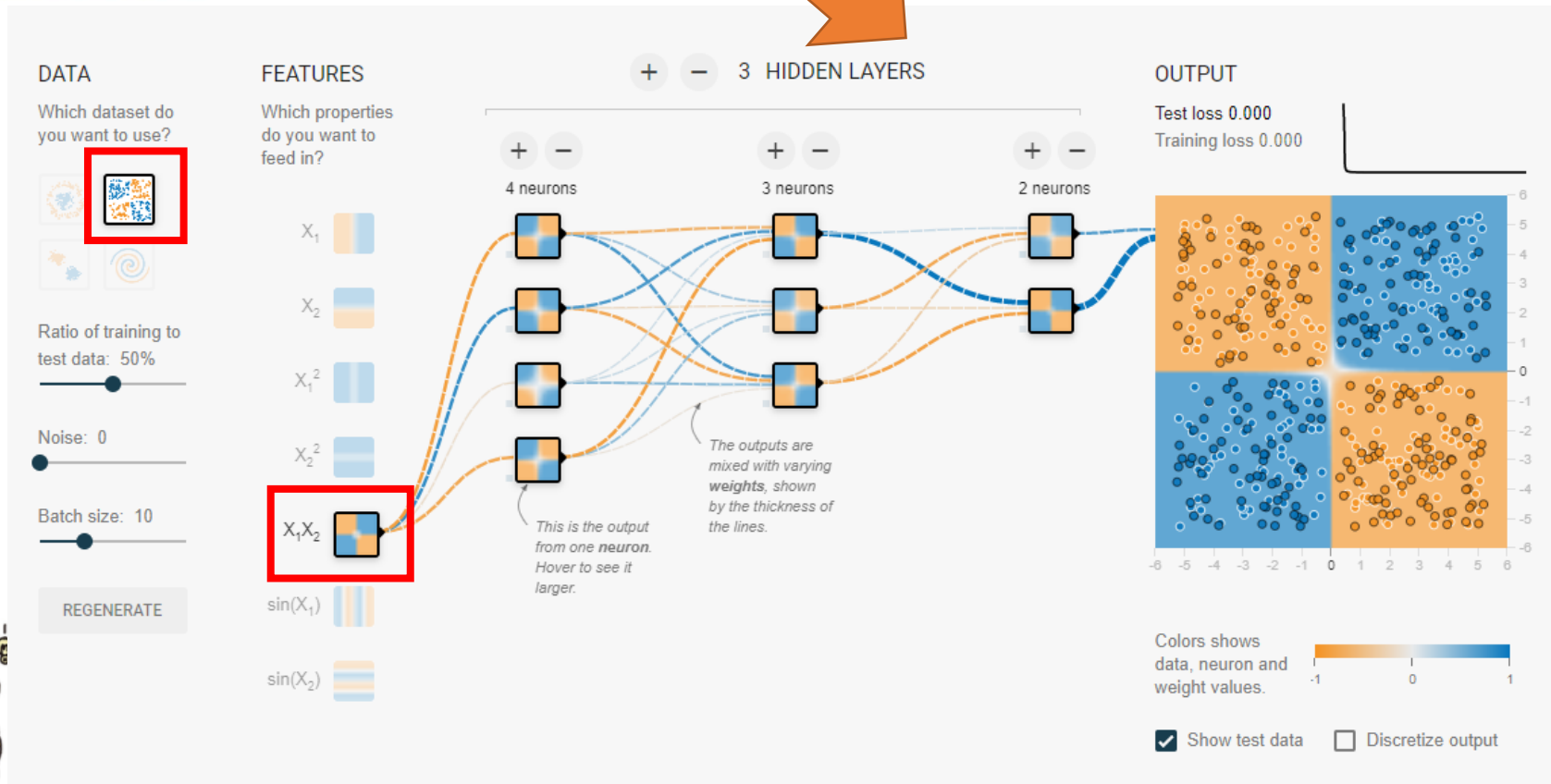


그림 7-9 | 학습이 잘 이루어진 데이터의 모습

이 때 입력값은
데이터의 모양과
비슷한 x_1x_2 를 선택



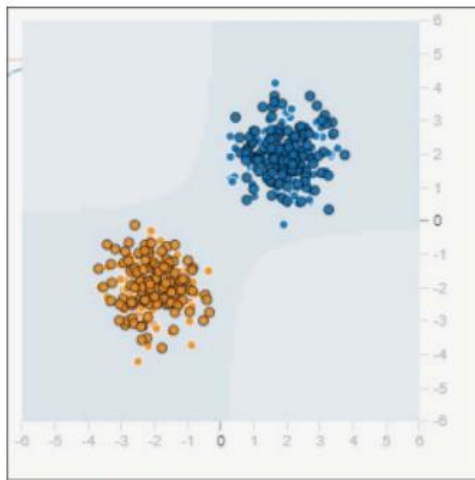


그림 7-11 | 좀 더 단순한 세 번째 데이터의 모습

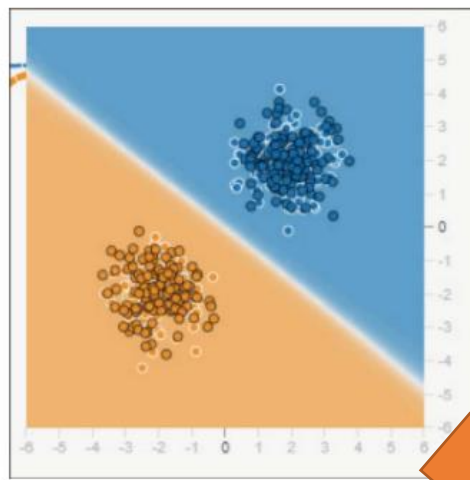
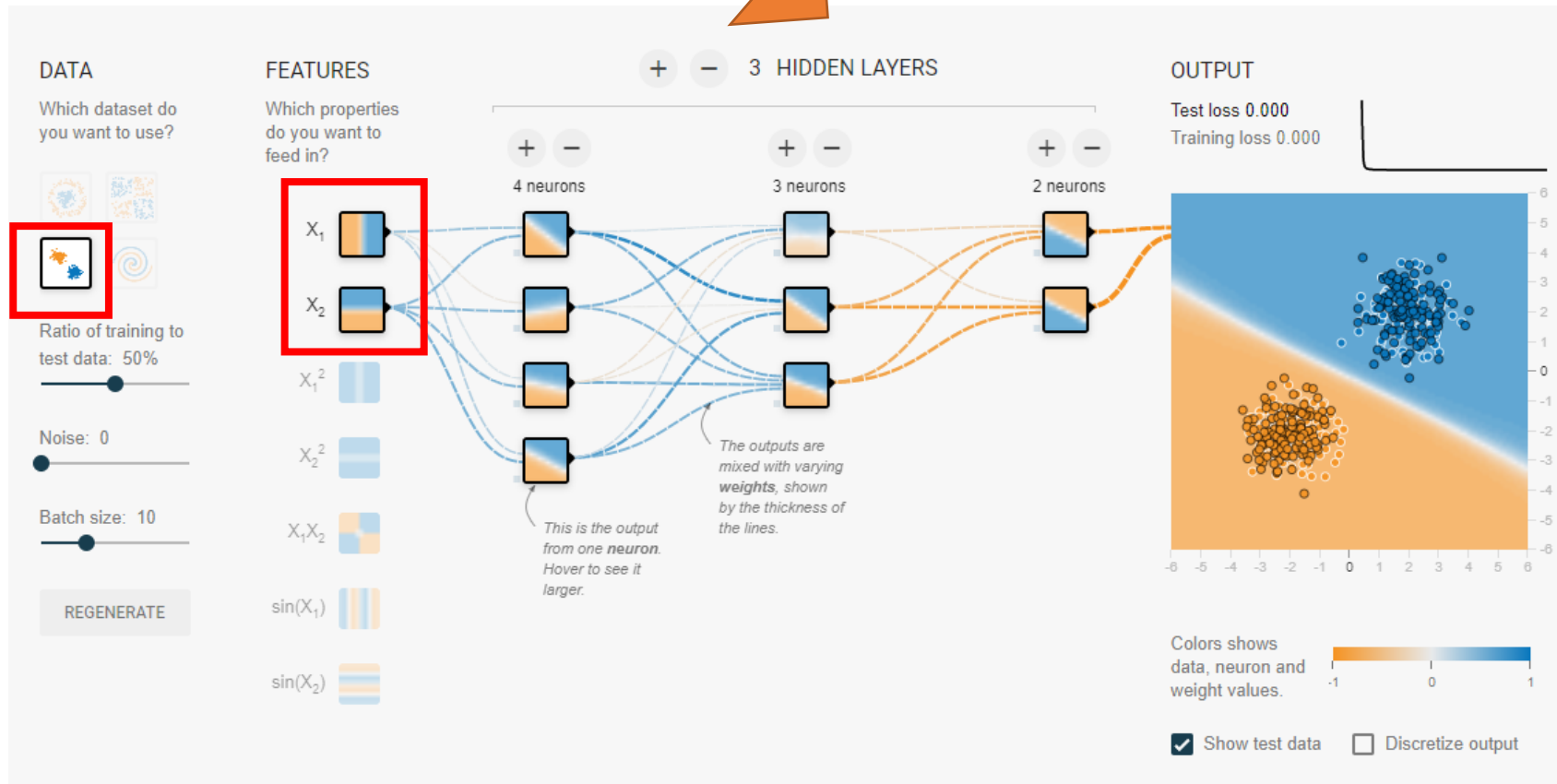


그림 7-12 | 학습이 잘 이뤄져서 명확하게 구분된 모습

이 때 입력값은
 x_1 과 x_2 를 선택



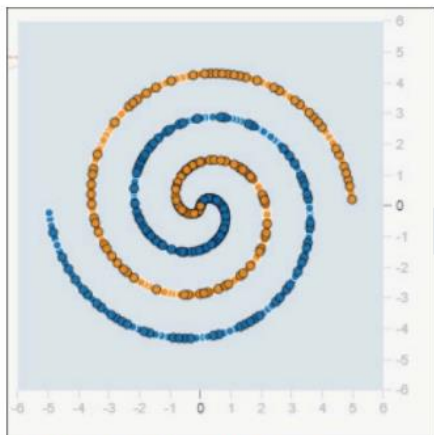


그림 7-14 | 분류하기 어려운 나선형 데이터

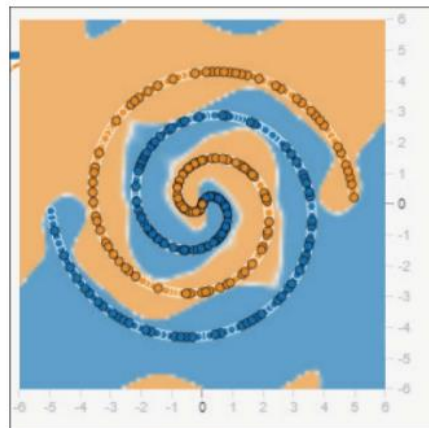


그림 7-15 | 학습이 잘 이뤄진 모습

이 때 입력값은 x_1 과 x_2 , $\sin(x_1)$ 과 $\sin(x_2)$ 를 선택

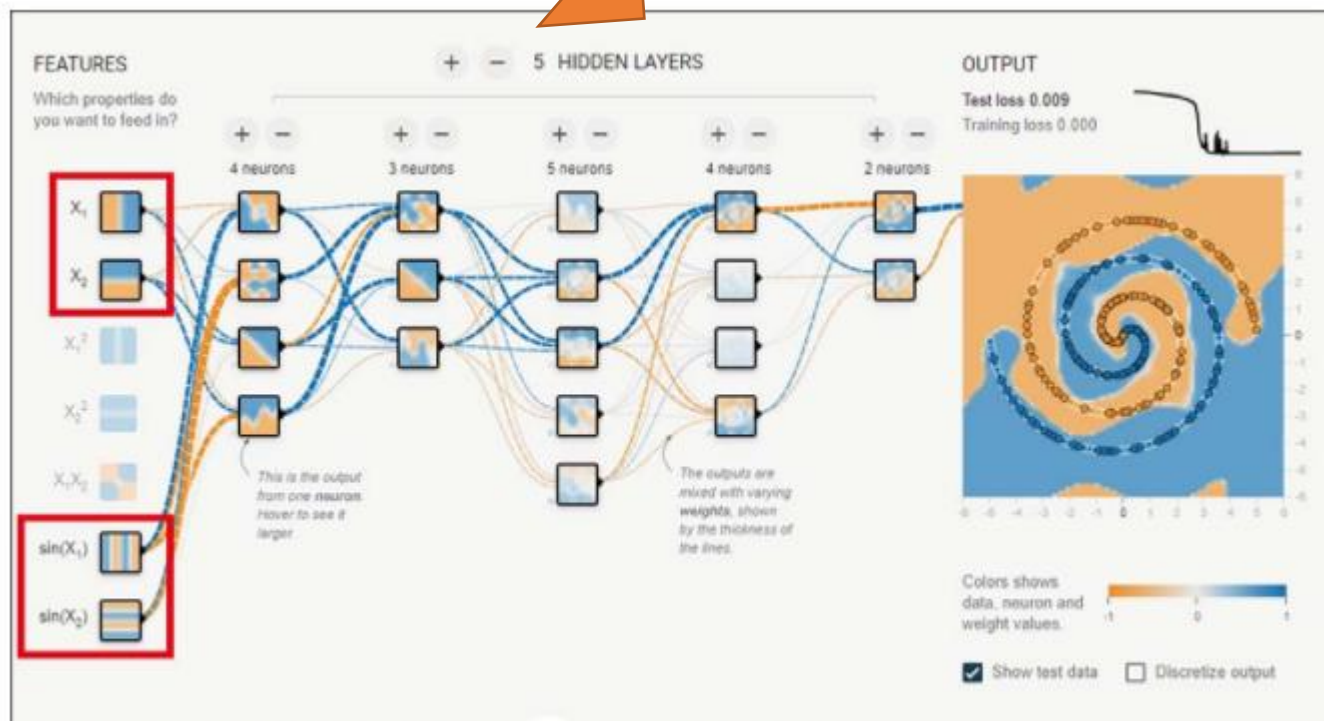


그림 7-16 | 입력 데이터, 층의 수, 뉴런의 수에 따라 생성되는 딥러닝 모델

7. 텐서플로 플레이그라운드 2배로 즐기기

Learning rate	Activation	Regularization	Regularization rate	Problem type
0.03	Tanh	None	0	Classification

그림 7-17 | 텐서플로 플레이그라운드 상단 화면

- Learning rate: Learning rate(학습률)은 딥러닝의 핵심 개념인 경사 하강법과 관련한 용어
- Activation: 활성화 함수
- Regularization rate: 정규화 할 때 어느 정도로 값을 수정할지 정해줄 때 사용하는 값
- Problem type: 분류와 회귀



UNIT08

다양한 딥러닝 기술 살펴보기

모두의
인공지능

이영호 지음

이미지 인식 인공지능 과정 살펴보기

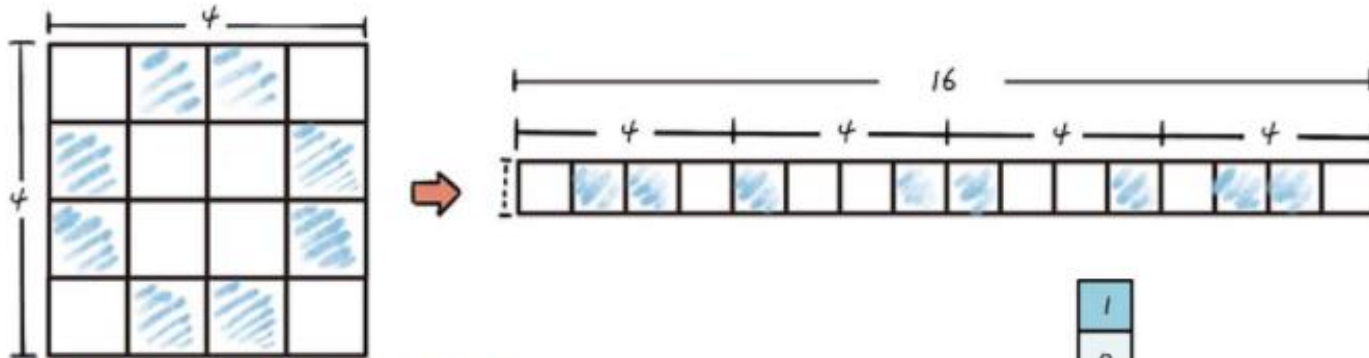


그림 8-1 | 4×4 이미지를 1×16 형태로 바꾸기

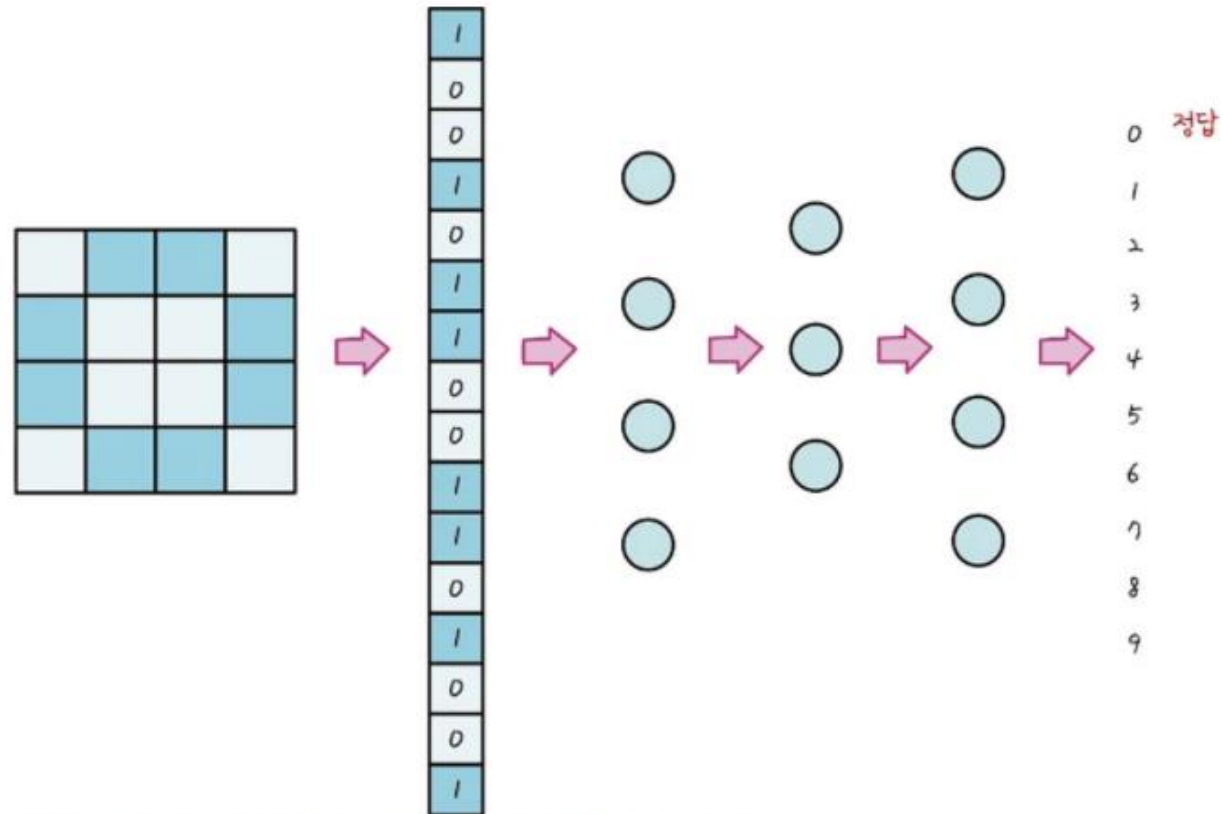


그림 8-2 | 16개의 숫자를 인공 신경망에 넣어 학습한 후 정답 결과 도출



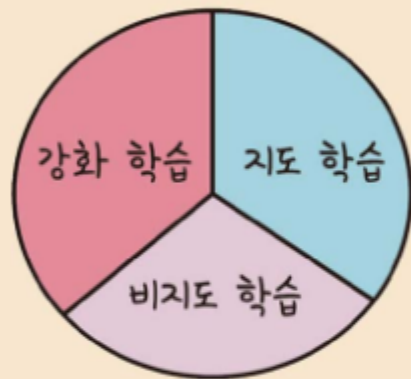
인공 신경망의 대표적인 기법

- 합성곱 신경망 살펴보기
- 순환 신경망 살펴보기
- 스케치 RNN 으로 순환 신경망 체험하기
- 생성 신경망 살펴보기
- deepart로 생성 신경망 체험하기



머신러닝의 학습 방법과 알고리즘이 어떻게 다른가요?

머신러닝의 학습 방법



머신러닝의 알고리즘



“머신러닝의 학습 방법은 지도 학습, 비지도 학습, 강화 학습이 있다고 했는데, 그럼 딥러닝은 또 무엇인가요?”

딥러닝은 인공 신경망을 사용한 머신러닝의 알고리즘 중 하나입니다. 머신러닝의 학습 방법과 머신러닝의 알고리즘은 다른 영역입니다. 이렇게 머신러닝의 다양한 영역 중 하나인 딥러닝에도 다양한 학습 방법이 있습니다. 바로 지도, 비지도, 강화 학습처럼 말이죠.

앞에서 예를 든 것은 정답이 있는 데이터로 학습하였기 때문에 지도 학습 방법의 딥러닝입니다. 아주 기본적인 딥러닝 방법이죠. 딥러닝에도 비지도 학습 방법의 딥러닝과 강화 학습 방법의 딥러닝이 있습니다.

비지도 학습 방법의 딥러닝에는 오토인코더, 적대적 생성 신경망(GAN) 등의 기법이 있습니다. 강화 학습 방법을 사용하는 Deep Q-Net work 딥러닝 모델도 있습니다. 이 책의 넷째 마당에서는 지도 학습 방법의 딥러닝인 기본적인 인공 신경망(ANN)과 순환 신경망(RNN)의 인공 신경망을 실습하고 비지도 학습 방법의 딥러닝인 GAN을 실습해 봅니다.

※ RNN, GAN은 딥러닝의 여러 알고리즘 중 하나입니다. 순환 신경망(RNN)은 연속된 값으로 특정한 값을 예측할 때 사용하며, 적대적 생성 신경망(GAN)은 무엇인가를 만들 수 있는 생성 신경망의 일종입니다. 자세한 내용은 다음 장에서 다루겠습니다. 단, 강화 학습은 이 책에서 따로 다루지는 않습니다. 강화 학습을 제대로 공부하고 싶은 사람은 중급서 이상의 딥러닝 서적을 볼 것을 권장합니다.

1. 합성곱 신경망 살펴보기

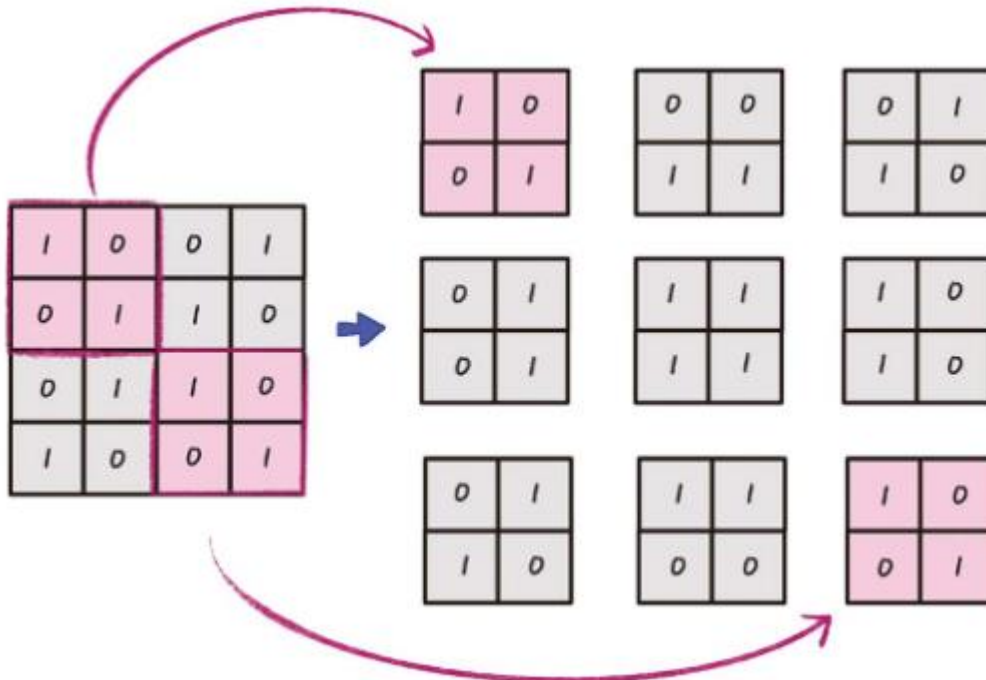


그림 8-3 | 4×4의 이미지를 2×2의 모습으로 추출하는 모습

- 실제 이미지 인식인공지능을 딥러닝 기법으로 만들 때 합성곱신경망(CNN, Convolutional Neural Network) 기법 사용
- 합성곱 신경망은 시각세포의 작동 원리를 본떠 만듦
- 이미지를 특정 영역별 추출하여 학습

합성곱 신경망: 하나의 데이터를 부분부분으로 나누어서 나타내는 것.
전체를 부분으로 나누어서 나타내는 것



1. 합성곱 인공 신경망

- 그런 다음 추출한 데이터를 인공 신경망에 넣음
- 그리고 마지막 결과값이 0이라고 알려주면 인공 신경망은 스스로 가중치와 편향을 바꿔가며 이미지가 숫자0이라고 학습

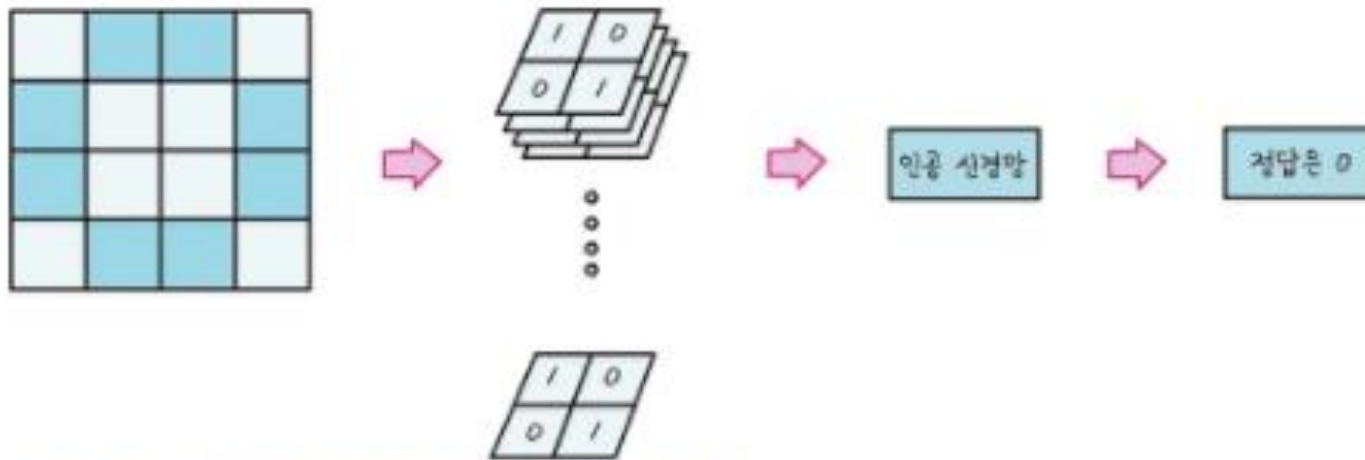


그림 8-4 | 전체와 부분을 학습하는 CNN



