
Deep Learning

Generative Deep Learning

Keras와 백엔드 라이브러리 Tensorflow

- Keras

- 신경망을 만들기 위한 고수준 파이썬 라이브러리
- 매우 유연하고 사용하기 쉬운 API

- TensorFlow

- 구글에서 만든 오픈소스 파이썬 머신러닝 라이브러리
- 텐서 조작에 강점이 있음

keras로 기본 신경망 틀을 만들고 순수한 Tensorflow를 내용을 채워 넣는것이 가장 좋은 옵션

* 텐서(Tensor): 데이터를 저장하고 네트워크를 통해 전달되는 다차원 배열

대부분 딥러닝 시스템은 여러개의 은닉층을 쌓은 인공 신경망이다.

- **Training(훈련)**

- 정확한 예측을 만들기 위해 필요한 각 층의 가중치 조합을 찾는것

- **Backpropagation(역전파)**

- 예측 에러는 네트워크를 통해 거꾸로 전파되어 예측을 가장 많이 향상시킬 수 있는 방향으로 가중치를 조금씩 수정함

- **Batch**

- 데이터셋 전체를 의미하거나 데이터셋을 동일한 개수로 나누어 놓은 것

입력 신호의 총합을 출력 신호로 변환, 입력 받은 신호를 얼마나 출력할지 결정하고 Network에 층을 쌓아 비선형성을 표현할 수 있도록 해줌

• 리키렐루(LeakyReLU)

- 입력이 0 이상이면 그대로 출력, 0 이하이면 작은 음수 값 출력
- 렐루 기반의 함수가 심층 신경망의 층을 안정적으로 훈련할 수 있는 가장 믿을만한 활성화 함수

• 시그모이드(sigmoid)

- 층의 출력을 0에서 1 사이로 조정하고 싶을 때 유용
- 다중 레이블 분류에 유용

• 소프트맥스(softmax)

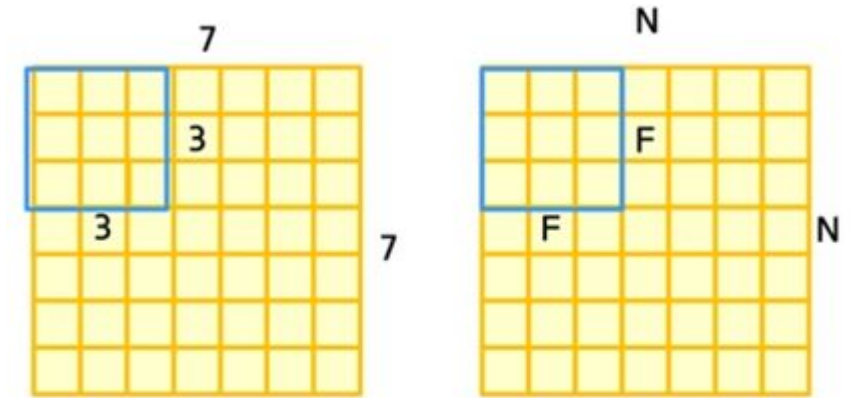
- 전체 출력의 합이 1이 되어야 할 때 사용
- 샘플이 정확히 하나의 클래스에만 속하는 다중분류에 사용

합성곱은 필터를 이미지의 일부분과 픽셀끼리 곱한 후 결과를 더하는 것
이미지 영역이 필터와 비슷할수록 큰 양수가 출력되고 필터와 반대일수록 큰 음수가 출력 됨

• 스트라이드(Stride)

- strides 매개변수는 필터가 한 번에 입력 위를 이동하는 크기
- 스트라이드를 크게 하면 출력 텐서의 크기가 줄어듦

2차원 배열에서 스트라이드 이해



• 패딩(padding)

- padding = “same” 매개변수는 입력 데이터를 0으로 패딩하여 stride = 1일 때 출력의 크기를 입력 크기와 동일하게 만듦

Output size :
 $(N - F) / \text{stride} + 1$

예) $N = 7, F = 3$
stride 1 $\Rightarrow (7-3)/1+1 = 5$
stride 2 $\Rightarrow (7-3)/2+1 = 3$
stride 3 $\Rightarrow (7-3)/3+1 = 2.33$

배치 정규화(Batch Normalization)

신경망 안에 포함되어 학습 시 평균과 분산을 조정하는 과정

각 레이어마다 정규화 하는 레이어를 두어, 변형된 분포가 나오지 않도록 조절하게 하는 것

- 훈련 시 유의점

네트워크의 가중치를 일정한 범위 안에서 유지해야 함

네트워크가 안전하게 훈련되다가 갑자기 손실함수가 NaN을 반환하여 그래디언트 폭주 발생할 수 있음

- 개선

1. 입력층 0~255 픽셀 값 대신 -1 에서 1사이로 스케일 조정

2. 배치 정규화

: 배치 정규화 층은 배치에 대해 평균과 표준 편차를 계산한

다음 평균을 빼고 표준편차로 나누어 정규화 함

배치 정규화 층은 완전 연결 층이나 합성곱 층 직후에 두어 층의 출력을 정규화 함

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_1 \dots x_m\}$;
Parameters to be learned: γ, β
Output: $\{y_i = \text{BN}_{\gamma, \beta}(x_i)\}$

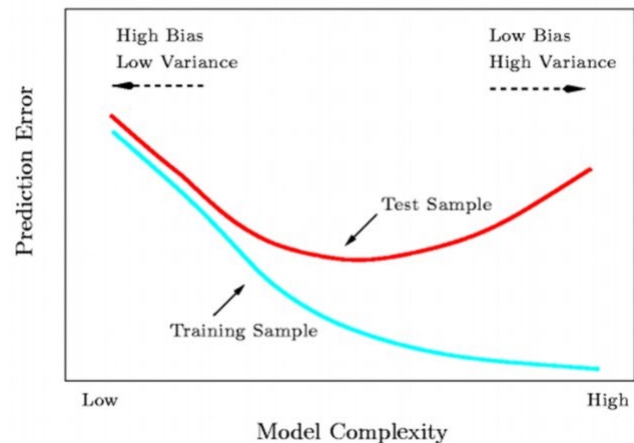
$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad // \text{ mini-batch mean}$$
$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\mathcal{B}})^2 \quad // \text{ mini-batch variance}$$
$$\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \quad // \text{ normalize}$$
$$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i) \quad // \text{ scale and shift}$$

드롭 아웃(Dropout)

훈련과정에서 드롭아웃 층은 이전 층의 유닛 일부를 랜덤하게 선택하여 출력을 0으로 지정

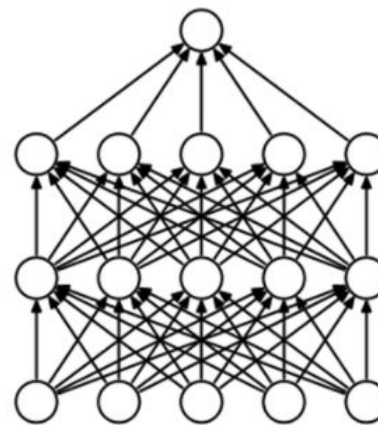
– Over fitting

: train set에 과도하게 의존하여
test set에서는 좋은 결과를 내지 못하는 경향

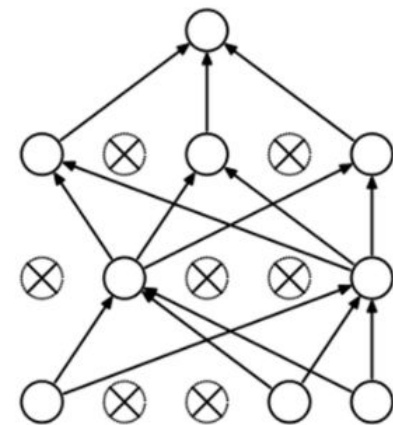


– Dropout

:랜덤하게 유닛을 끄는 식으로 진행
가중치 개수가 많아 과대적합되기 쉬운
Dense층 다음에 주로 사용



(a) Standard Neural Net



(b) After applying dropout.