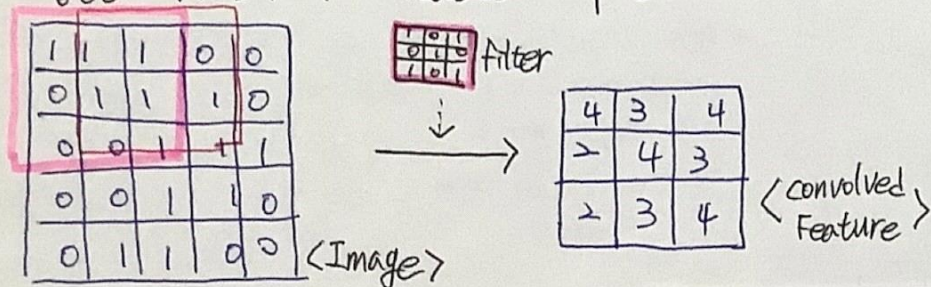


CNN의 주요 용어 정리

* 합성곱 (Convolution)

: 예) 2차원 입력 데이터 (shape: (5,5))를 1개의 필터로 합성곱 연산.
합성곱 처리 결과로부터 Feature Map 만든다.

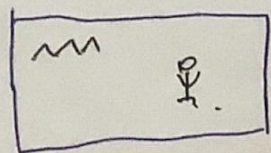
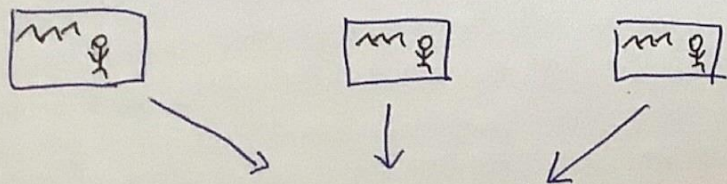


* 채널 (channel)

: 이미지 픽셀 하나 하나는 실수. 컬러사진은 천연색 표현 위해, 각 픽셀을 RGB 3개의 실수로 표현한 3차원 데이터. → 컬러이미지 채널 3개.
흑백 사진은 명암만 표현. 2차원 데이터 → 흑백 이미지 채널 1개.

ex) 높이 39 픽셀, 폭 31 픽셀, 컬러사진의 데이터 shape: (39, 31, 3)³
높이 39 픽셀, 폭 31 픽셀, 흑백사진의 데이터 shape: (39, 31, 1)

Red channel Green channel Blue channel

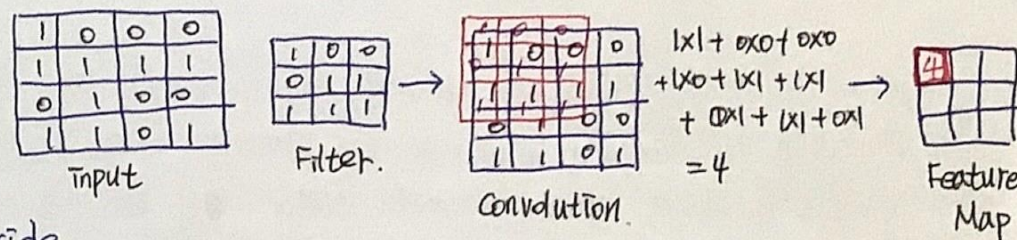


color image

* 필터 (Filter) & stride

: 필터는 이미지의 특징을 찾아내기 위한 공통 파라미터 (= kernel).
일반적으로 정사각 행렬. CNN의 학습 대상은 필터 파라미터.
입력 데이터를 지정된 간격으로 순회하며 채널 별로 합성곱을 하고 모든 채널의 합성곱의 합을 Feature Map으로 만들기. 전체 입력데이터 x Filter. 필터: 3개.

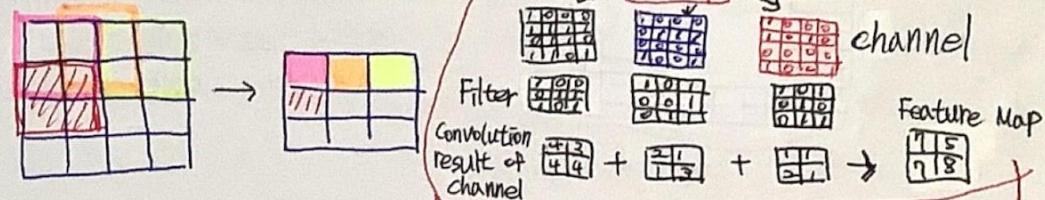
ex) 채널이 1개인 입력데이터를 (3,3)크기의 필터로 합성곱 과정.



Stride

: 필터가 입력데이터를 지정된 간격으로 순회하면서 합성곱 계산할 때, 지정된 간격 (필터 순회 간격).

ex) stride가 1로 설정되었을 때.



입력데이터가 여러 채널을 가지고 있는 경우, 필터는 각 채널을 순회하며 합성곱 계산해 채널별 피쳐 맵 만든다. 각 채널의 피쳐 맵을 합산해 최종 피쳐 맵 반환. 입력데이터는 채널 수와 상관 없이 필터 별로 1개의 피쳐 맵 생성

하나의 convolution Layer에 크기가 같은 여러개의 필터 적용가능. 이때, Feature Map은 필터 개수만큼 채널 만들어짐. 입력 데이터에 적용한 필터의 개수는 출력 데이터인 Feature Map의 채널 수가 됨.

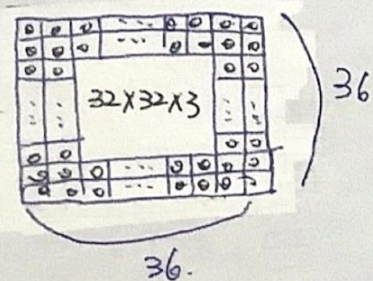
* Feature Map

: Convolution layer의 입력데이터를 필터가 순회하며 합성곱을 통해 만든 출력을 Feature Map (= Activation Map). Feature Map은 합성곱 계산으로 만들어진 행렬. Activation Map은 Feature Map에 활성화 함수 적용 결과. 즉, Convolution layer의 최종 출력 결과는 Activation Map.

* 패딩 (Padding)

: Convolution layer에서 Filter와 stride의 작용으로 Feature Map 크기는 입력 데이터보다 작음. Convolution layer의 출력 데이터가 줄어드는 것을 방지하는 방법이 패딩임. 패딩은 입력데이터의 외곽에 지정된 픽셀 만큼 특정값으로 채워넣는 것. 보통 패딩 값으로 '0'으로 채워넣음.

ex) (32, 32, 3) 데이터를 외곽에 2pixel 추가해 (36, 36, 3) 행렬 만들기.



* 패딩을 통해

- ① Convolution layer와 출력데이터 크기 조정
- ② 외곽을 '0'으로 둘러싸는 특징에서 인공신경망이 이미지의 외곽 인식 효과.

cf. Pooling layer와 Convolution layer 비교.

Pooling layer는

- ① 학습 대상 파라미터 X.
- ② Pooling layer 통과시 행렬 크기 ↓
- ③ Pooling layer 통과시 채널수 변경 X.

→ 추가설명: Convolution layer에서 입력 데이터의 채널이 3개여도 Feature Map은 하나였음. 따라서 Pooling layer가 채널수 변경 X가 중요하다고 생각할 수 있지만 Convolution layer를 3번 실행하고 Pooling layer 실행하면 이득점 좋아!

* Pooling layer

→ Activation map

: 풀링레이어는 Convolution layer의 출력데이터를 입력으로 받아 출력데이터의 크기를 줄이거나 특정데이터 강조하는 용도. Pooling layer 처리 방법으로 Max Pooling, Average Pooling, Min Pooling 있음. 각각 행렬의 특정 영역에서의 최댓값, 평균 구하기. 일반적으로, Pooling 크기나 stride를 같은 크기로 설정하여 모든 원소가 한 번씩 처리 되도록!

ex)

12	20	30	0
8	12	2	0
34	10	31	7
112	100	22	12

20	30
112	31

Max Pooling. CNN에서 자주 사용

13	8
79	18

Average Pooling.

Activation map.

CNN 주요 용어 정리 2

• 알고리즘이 iterative 하다.

• gradient descent와 같이 결과를 내기 위해서 여러번의 최적화 과정을 거쳐야 하는 알고리즘.
→ 다뤄야 할 데이터가 너무 많고, 메모리 부족해 한번의 계산으로 최적화된 값 찾기는 힘들다. 따라서, ML에서 최적화 (optimization)를 할 때는 일반적으로 여러번의 학습과정도, 한번의 학습과정 역시 사용하는 데이터를 나누는 방식으로 세분화.
(epoch, batch size, iteration)

* epoch

One epoch is when an ENTIRE dataset is passed forward and backward through the neural network only ONCE.

한번의 epoch는 인공신경망에서 전체 데이터셋에 대해 forward/backward 과정 거친 것. 즉, 전체 데이터셋에 대해 한 번 학습을 완료한 상태

→ 신경망에서 사용되는 역전파 알고리즘 (backpropagation algorithm)은 파라미터를 사용하여 입력부터 출력까지의 각 계층의 weight를 계산하는 과정을 거치는 순방향 패스 (forward pass) forward pass를 반대로 거슬러 올라가며 다시 한번 계산 과정을 거쳐 각층의 weight를 수정하는 역방향 패스 (backward pass)로 나뉜다. 이 전체 데이터셋에 대해 forward pass + backward pass가 완전히 완료되면 한 번의 epoch가 진행됨.

ex) epoch 40 : 전체 데이터를 40번 사용해 학습 거치기

⇒ 적절한 epoch 값을 설정해 모델 만들어야 underfitting과 overfitting 방지

epoch 값 ↓ : underfitting 현상 有

epoch 값 ↑ : overfitting 현상 有

* batch size

: Total number of training examples present in a single batch.

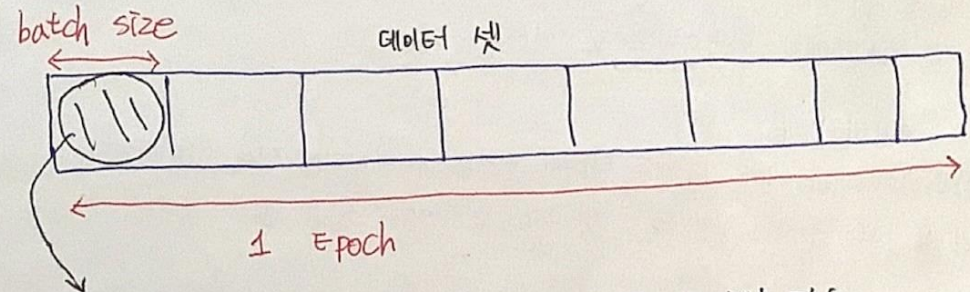
* iteration

: The number of passes to complete one epoch.

⇒ batch size는 한 번의 batch마다 주는 데이터 샘플의 size.

여기서 batch는 나뉜 데이터셋을 뜻하며 iteration은 epoch를 나뉜 실행하는 횟수.
↳ 보통 mini-batch로 표현.

⇒ 메모리의 한계와 속도 저하 때문에 대부분의 경우에는 한 번의 epoch에서 모든 데이터를 한꺼번에 읽어낼 수 X. 그래서 데이터 나눠주게 되는데, 이때 몇 번 나눠주는가를 iteration. 각 iteration마다 주는 데이터 사이즈 batch size



• minibatch : 데이터 셋을 batch size 크기로 쪼개서 학습.

예시

전체 2000개 데이터셋. epoch = 20, batch size = 500 일 때,

1 epoch = 4 iteration, 전체 데이터 셋에 대해 20번 학습.
이테레이션 기준 80번 학습.
각각 데이터 size 500인 batch 들어감.

* Batch Normalization

학습의 효율 높이기 위해 도입. 배치 정규라는 Regularization 해준다고 볼 수 있음

-장점

- ① 학습 속도 개선 : 학습률 높게 설정할 수 있기 때문
- ② 가중치 초기화 선택의 의존성 ↓ : 학습할 때 마다 출력값 정규화
- ③ 과적합 (overfitting) 위험 ↓ : 조금 다른 같은 가법 대체 가능
- ④ Gradient Vanishing 문제 해결

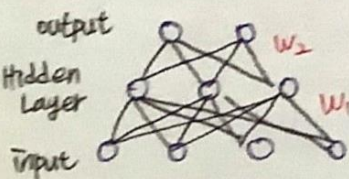
-개념

배치 정규라는 활성화 함수의 값 또는 출력값을 정규화 하는 작업을 말한다. 배치 정규화를 활성화 함수 이전에 하는지, 이후에 하는지 논의/실험중. 신경망의 각 Layer에서 데이터(배치)의 분포를 정규화 하는 작업. 인공의 노이즈를 추가하는 방법. 배치마다 정규화 함으로써 전체 데이터에 대한 평균과 분산의 값이 달라질 수 있음. 학습할 때 마다 활성화함/출력값 정규화 하기 때문에 초기값에서 자유로워질 수 있다.

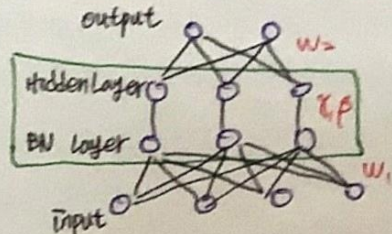
-BN in Neural Net

각 hidden layer에서 정규화를 하면서 임계값이 일정하게 되고, 이에 따라 Learning rate을 크게 설정해도 괜찮아진다. 결과적으로 학습 속도가 빨라진다.

Neural Net without BN

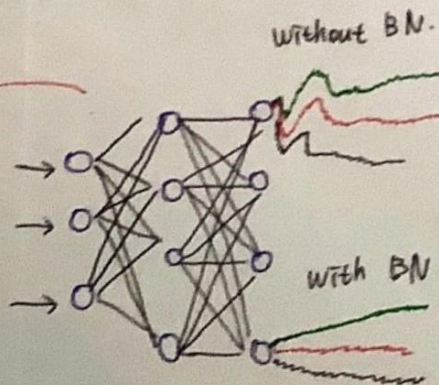


Neural Net with BN



-입력분포의 균일화

학습할 때, hidden layer의 중간에서 입력분포가 학습할 때 마다 변화하면서 가중치가 일정한 방향으로 정해지는 문제 발생할 수 있음. 신경망 층이 깊어질수록 학습에 부정적인 입력분포가 변화하여 일정한 학습이 될 수 있다.



-배치 정규화 알고리즘

간단하게 배치 정규라는 학습 시 미니 배치를 한 단위로 정규화 하는 것으로, 분포의 평균이 0. 분산이 1이 되도록 정규화 하는 것.

input : values of x over a mini-batch : $b = \{x_1, \dots, x_m\}$
Parameters to be learned : γ, β

output : $\{y_i = BN_{\gamma, \beta}(x_i)\}$

$$\mu_b \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad \# \text{ mini-batch mean}$$

$$\sigma_b^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_b)^2 \quad \# \text{ mini-batch variance}$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu_b}{\sqrt{\sigma_b^2 + \epsilon}} \quad \# \text{ Normalize}$$

$$y_i \leftarrow \gamma \hat{x}_i + \beta \quad \# \text{ scale and shift}$$

~정규화했다!

input으로 제공된 mini batch의 평균 분산

hidden layer의 활성화함/출력값에 대해 평균이 0, 분산이 1인 정규화.

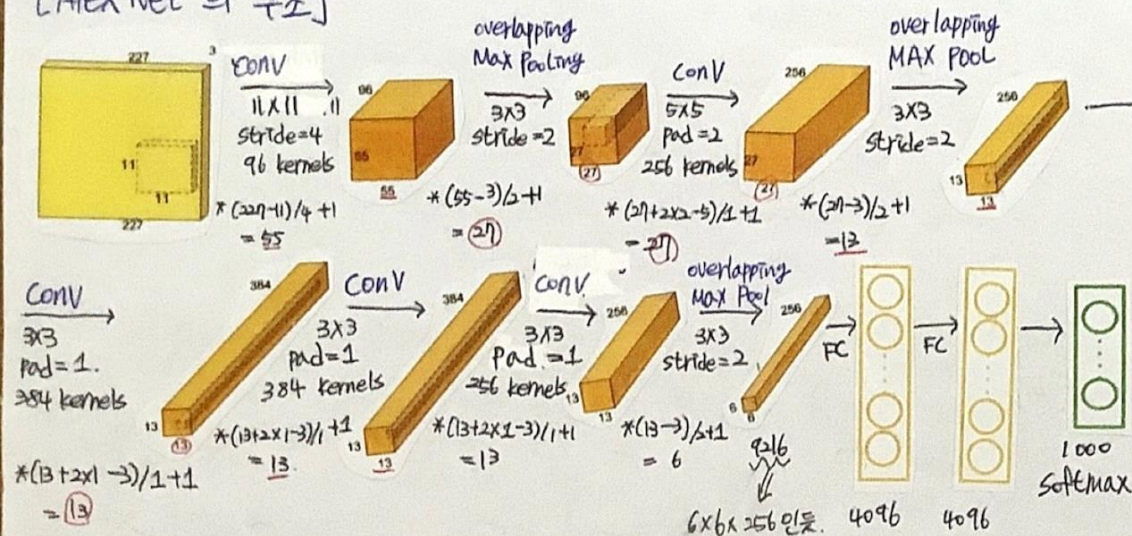
↓

데이터 분포가 덜 좌우치게 되고 배치 정규화 단계마다 한대 scale과 이동 수를 변할 수 있음.

CNN 에서 Parameter 개수, tensor 계산

- Alex Net 기준 예시

[Alex Net 의 구조]



↓
설명

- Input : 227x227x3 크기의 컬러 이미지
- Conv-1 : 11x11 크기의 커널 96개, stride=4, padding=0.
- Max Pool-1 : stride=2, 3x3 max pooling layer.
- Conv-2 : 5x5 크기의 커널 256개, stride=1, padding=2
- Max Pool-2 : stride=2, 3x3 max pooling layer
- Conv-3 : 3x3 크기의 커널 384개, stride=1, padding=1.
- Conv-4 : 3x3 크기의 커널 384개, stride=1, padding=1.
- Conv-5 : 3x3 크기의 커널 256개, stride=1, padding=1
- Max Pool-3 : stride=2, 3x3 max pooling layer
- FC-1 : 4096 개의 fully-connected layer.
- FC-2 : 4096 개의 fully-connected layer
- FC-3 : 1000 개의 fully-connected layer

* Convolution layer 의 output tensor size

$$\text{size (width) of output Image} = \frac{I-k+p}{S} + 1$$

- I : size (width) of input image
- k : size (width) of kernels used in the conv layer.
- p : padding size
- S : stride of the convolution operation.
- + N : Number of kernels. → 출력 이미지의 채널 수는 커널의 개수와 같다 (N)

ex) Alex Net 에서 conv-1

입력 이미지 : 227x227x3, conv layer : 11x11x3 크기, 커널 96개, stride=4, padding=0
 output image size = $\frac{227-11+2 \times 0}{4} + 1 = \frac{216}{4} + 1 = 55 \Rightarrow 55 \times 55 \times 96$

* Max Pooling layer 의 output tensor size

$$\text{size (width) of output Image} = \frac{I-P_s}{S} + 1 \rightarrow P_s : \text{Pooling size.}$$

→ 출력의 채널 수는 입력의 개수와 동일 (conv. layer와 다름!)

conv layer의 수식에서 커널 크기 (k)를 P_s 로 대체. $P=0$ 으로 설정하면 동일함.

ex) AlexNet 에서 maxPool-1

stride=2, 사이즈 : 3x3, 이전 단계 (conv-1) 출력 크기 : 55x55x96.
 $\Rightarrow \frac{55-3}{2} + 1 = 27$

* Fully Connected layer 의 output tensor size.

: FC layer는 layer의 뉴런 수와 동일한 길이의 벡터를 출력.

ex) AlexNet

Input : 227x227x3 → conv-1은 max pool 1 가지여서 55x55x96에서 27x27x48 변환
 → conv-2는 27x27x128에서 max pool 2 가지여서 13x13x64으로 변경 → conv-3은 13x13x128
 → conv-4는 크기 유지 → conv-5는 27x27x256 변환 → MaxPool-3은 6x6x256 줄임
 → 이 이미지는 크기 4096x1 크기의 벡터로 변환되는 FC-1에 feed. → FC-2 크기 유지.
 → FC-3은 size를 1000x1로 변환

* Convolution layer의 parameter 개수.

- CNN의 각 layer는 weight parameter와 bias parameter 존재.
- 전체 Network의 parameter의 수는 각 convolution layer 파라미터 수의 합.

Number of weights of the conv layer : $W_c = k^2 \times C \times N$

Number of biases of the conv layer : $B_c = N$

Number of parameters of the conv. layer : $P_c = W_c + B_c$

→ k : size (width) of kernels used in the conv. layer.

N : Number of kernels

★ kernel = filter. ★

C : Number of channels of the input image

- conv. layer에서 모든 커널의 값은 항상 입력 이미지의 채널 수와 같음.
- 따라서, 모든 커널에는 $k^2 \times C$ 개의 parameter들이 있으며, 그러한 커널들이 N 개 존재.

ex) AlexNet의 Conv-1 : 입력 이미지의 채널 수 $C=3$, kernel size $k=11$.

전체 커널 개수 $N=96$. → $W_c = 11^2 \times 3 \times 96 = 34,848$, $B_c = 96$ ∴ $P_c = 34,848 + 96 = 34,944$

conv-2, 3, 4도 동일한 방법으로 614, 656, 888, 120 / 1,320, 488 / 884, 992 개의

parameter 값을. AlexNet의 conv. layer parameter 개수는 3,141,200개.

FC layer의 parameter 수가 더러지지 않았으므로 전체 네트워크의 parameter 개수 \times .

→ Conv. layer의 장점은 weight parameter가 공유되므로 FC layer에 비해 매개변수 축소

* Max Pool layer의 parameter 개수.

: Pooling, stride, padding은 hyper parameter. ← 계산 \times .

* Fully Connected layer의 parameter 개수.

: CNN에는 2종류의 FC layer 존재

- ↳ 마지막 conv. layer의 바로 뒤에 붙는 FC layer
- ↳ 다른 FC layer에 연결되는 FC layer.

1. FC layer connected to a Conv. layer.

$W_{cf} = O^2 \times N \times F$. (Number of weights of a FC layer which is connected to a conv. layer)

$B_{cf} = F$ Number of biases of a FC layer which is connected to a conv. layer

$P_{cf} = W_{cf} + B_{cf}$ (Number of parameters of a FC layer which is connected to a conv. layer)

→ O : size (width) of the output image of the previous conv. layer

N : Number of kernels in the previous conv. layer

F : Number of neurons in the FC layer.

ex) AlexNet

FC-1 layer : $O=6$, $N=256$, $F=4096$ → $W_{cf} = 31,748,126$ $B_{cf} = 4096$
 $P_{cf} = 31,752,832$

2. FC layer connected to a FC layer.

- Number of weights of a FC layer which is connected to a FC layer → $W_{ff} = F_1 \times F$.

- Number of biases of a FC layer which is connected to a FC layer → $B_{ff} = F$

- Number of parameters of a FC layer which is " → $P_{ff} = W_{ff} + B_{ff}$

→ F : Number of neurons in FC layer, F_1 : Number of neurons in the previous

→ $F_1 \times F$ 는 이전 FC layer에서의 neuron과 현재 FC layer의 neuron 사이의 중 연결 개수.

FC layer.

→ Bias parameter의 개수는 neuron의 개수 F 와 같음.

ex) AlexNet : FC-3.

$F_1 = 4096$, $F = 1000$ → $W_{ff} = 4096 \times 1000 = 4,096,000$ $B_{ff} = 1,000$
 $P_{ff} = 4,097,000$