**Model explanation**

1D CNN model

number of conversion layer : 3

filters : 128 / 256 / 128

kernel\_size : 9 / 6 / 3

padding : same(크기동일하게 패딩해줌)

activation : relu

Batch Normalization

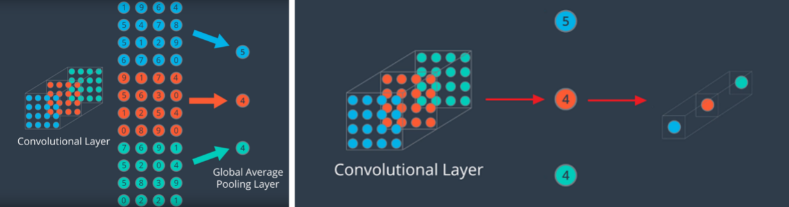
* 학습을 더 빨리 하기 위해서/ 로컬 최적화 문제에 빠지는 가능성을 줄이기 위해서 사용
* 배치 정규화는 단순하게 평균과 분산을 구하는 것이 아니라 **감마(Scale), 베타(Shift) 를 통한 변환을 통해 비선형 성질을 유지 하면서 학습 될 수 있게 해줌**  
  (배치 정규화가 신경망 레이어의 중간 중간에 위치하게 되어 **학습을 통해 감마, 베타를 구할 수 있음)**
* **Internal Covariate Shift 문제**로 인해 신경망이 깊어질 경우 학습이 어려웠던 문제점을 **해결**
* gradient의 스케일이나 초기 값에 대한 dependency 가 줄어들어 Large Learning Rate 를 설정할 수 있기 떄문에 결과적으로 **빠른 학습 가능함**, 즉, **기존 방법에서 learning rate 를 높게 잡을 경우 gradient 가 vanish/explode 하거나 local minima 에 빠지는 경향이 있었는데 이는 scale 때문이었으며, 배치 정규화를 사용할 경우 propagation 시 파라미터의 scale 에 영향을 받지 않게 되기 때문에 learning rate 를 높게 설정할 수 있는 것**임
* regularization 효과가 있기 때문에 **dropout 등의 기법을 사용하지 않아도 됨** (효과가 같기 때문)
* 학습 시 Deterministic 하지 않은 결과 생성
* Learning Rate Decay를 더 느리게 설정 가능
* 입력의 범위가 고정되기 때문에 saturating 한 함수를 활성화 함수로 써도 saturation 문제가 일어나지 않음, 여기서 saturation 문제란 가중치의 업데이트가 없어지는 현상임

‘

Dropout

rate: 0.3, 0.4, 0.5

GlobalAveragePooling1D()



* GAP(global average pooling)은 앞에서 설명한 Max(Average) Pooling 보다 더 급격하게 feature의 수를 줄**여 1차원 벡터로 만들어줍니다.**
* 위 그림을 보면 같은 채널 (같은 색)의 feature들을 모두 평균을 낸 다음에 채널의 갯수(색의 갯수) 만큼의 원소를 가지는 벡터로 만듭니다.
* 이런 방식으로 GAP는 (height, width, channel) 형태의 feature를 (channel,) 형태로 간단하게 만들어 버립니다.
* 그러면 이렇게 극단적인 벡터 형태로 나타내는 GAP는 왜 고안되었을까요?
* GAP는 CNN + FC(Fully Connected) Layer에서 classifier인 **FC Layer를 없애기 위한 방법으로 도입**되었습니다.
* FC Layer는 마지막 feature와 matrix 곱을 하여 feature 전체를 연산의 대상으로 삼아서 결과를 출력합니다. 즉, feature가 이미지 전체를 함축하고 있다고 가정하면 이미지 전체를 보고 출력을 만들어 내는 것입니다.
* 하지만 FC layer를 classifier로 사용하는 경우 파라미터의 수가 많이 증가하는 단점이 있으며 feature 전체를 matrix 연산하기 때문에 위치에 대한 정보도 사라지게 됩니다. 더구나 FC Layer 사용 시 반드시 지정해 주어야 하는 FC layer의 사이즈로 인해 입력 이미지 사이즈 또한 그에 맞춰서 고정되어야 하는 단점이 있습니다.
* 반면 GAP는 어떤 크기의 feature 라도 같은 채널의 값들을 하나의 평균 값으로 대체하기 때문에 벡터가 됩니다. 따라서 **어떤 사이즈의 입력이 들어와도 상관이 없습니다.** 또한 단순히 (H, W, C) → (1, 1, C) 크기로 줄어드는 연산이므로 파라미터가 추가되지 않으므로 학습 측면에서도 유리합니다. 또한 파라미터의 갯수가 FC Layer 만큼 폭발적으로 증가하지 않아서 over fitting 측면에서도 유리합니다.
* 따라서 GAP 연산 결과 1차원 벡터가 되기 때문에 **최종 출력에 FC Layer 대신 사용**할 수 있습니다.
* 경우에 따라서 FC layer와 같이 사용 되기도 합니다. FC layer에 전달하기 전에 GAP를 이용하여 차원을 줄여서 벡터로 만든 다음에 FC layer로 전달 하면 FC Layer에서 쉽게 사이즈를 맞출 수 있기 때문입니다.

Dense

activation : softmax