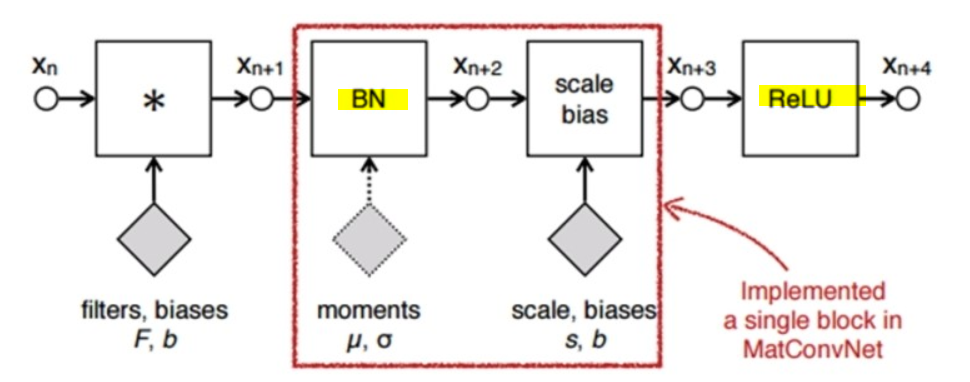
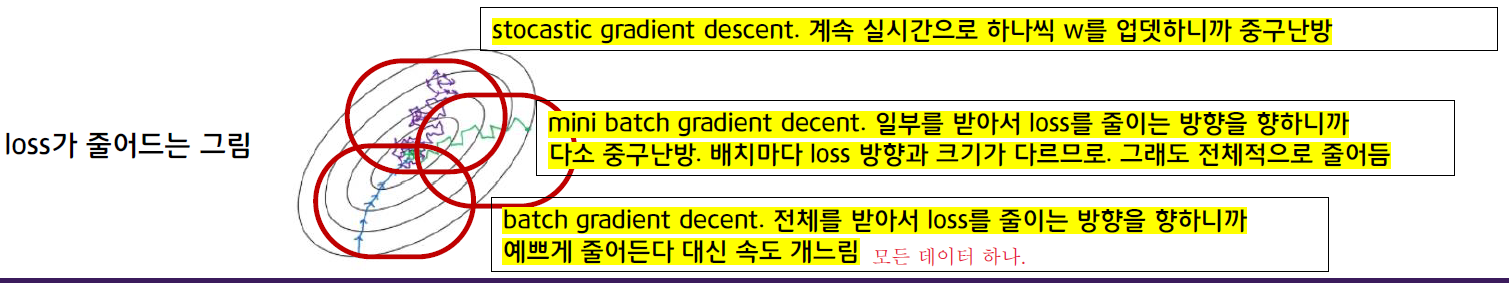
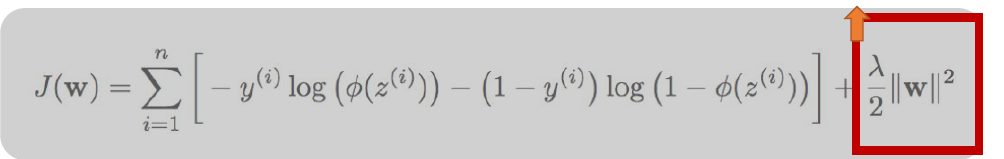
1. 그래디언트 소실 or 그래디언트 폭주

* 해결법) 활성화 함수 변경, 가중치 초기화, batch normalization

그래디언트 소실-> 역전파 알고리즘으로 심층 신경망을 학습시키는 과정에서, 출력층에서 멀어질수록 신경망의 출력오차가 반영되지 않는 현상.(layer가 너무 많아지면, 그 값이 소실될 수가 있다. 왜? 층이 지날수록 오류값이 적어지기 때문.)  
이에 대한 대처 방안 -> 활성함수를 Rellu

그래디언트 폭주-> 소실과는 반대로, 역전파에서 그래디언트가 점점 커져 입력층으로 갈수록 가중치 매개변수가 기하급수적으로 커지게 되는 경우.

* Mini batch 단위로 파라미터 조정 // batch 기준으로, 평균 분산 다 새로 계산하는 것.  
  단순 정규화 보다, 훨씬 더 강력해짐.   
  활성화 함수 전에 들어간다.  
  

1. 속도의 문제 ( 대규모 신경망에서 훈련이 극단적으로 느려짐)   
   batch 기준으로 w를 업데이트 한다.  
   stochastic-> 한 행씩 업데이트 하는 과정.  
     
     
     
     
     
   Gradient descent 를 효율적이게 하는 법 (속도와 방향을 조정한다.)  
   방향)  
   Momentum – 이 전에 하강이 많이 된 w에 더 가중치를 줘서 하강.  
   과거에 이동했던 방식을 기억하면서, 그 방향으로 일정 정도를 추가적으로 이동하는 방식.  
   속도)  
   RMSProp- w의 하강 속도가 빠르지도 느리지도 않게 해주는 것.   
     
   이 두개를 섞은 것이 Adam!!
2. Over fitting의 문제 )  
   L2 reguralization 은 loss funtion을 큰 값이 많이 존재하는 가중치에 제약을 두고, 특정 몇 개의 입력데이터에 강하게 적용되기 보다는, 모든 입력데이터에 약하게 적용되도록 하는 것.  
     
     
   비용 함수에, 모든 가중치의 제곱 합을 더 해줌으로써 ,특정 가중치에 쏠리지 않게 한다.