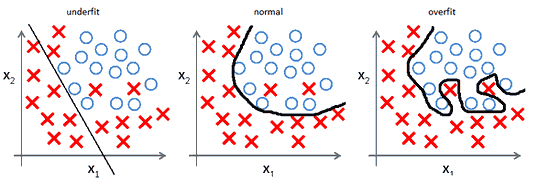
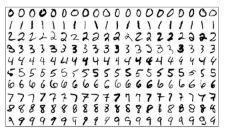
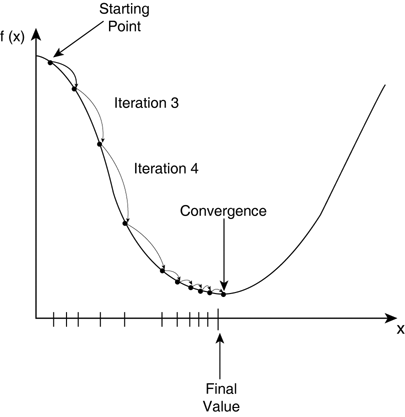
1. Overfitting의 개념과 Regularization의 개념을 조사해 설명하고, 실습 시간에 진행한 MNIST 내용으로 예를 들어라.  
     
   Overfitting(과적합)은 주어진 데이터에 대해 과하게 적합하게 학습된 경우를 뜻한다. 즉, 주어진 데이터에 한해서만 높은 정확도를 보이고 실제 전체 데이터에서는 낮은 정확도를 보이게 된다. 학습된 데이터가 아닌 새로운 데이터가 들어오면 제대로 인식을 못 할 가능성이 높다. 예를 들어, 아래 그림의 제일 오른쪽 그림을 보면 complexity가 너무 증가하여 주어진 데이터에만 적합하게 고려되었다는 것을 볼 수 있다. 이런 경우를 overfitting이라고 한다.  
   [](https://blog.naver.com/PostView.nhn?blogId=cattree_studio&logNo=220703210504&redirect=Dlog&widgetTypeCall=true&topReferer=https%3A%2F%2Fwww.google.com%2F&directAccess=false)  
     
   이러한 Overfitting 현상을 해결하기 위한 방법 중 하나가 Regularization (정규화) 이다. 이것은 cost function을 수정하여 complexity를 낮추는 과정이다. Weight가 높을수록 overfitting이 발생할 가능성이 높으므로 weight 값이 클 경우에 penalty를 부과한다.   
   아래의 공식을 보자.  
   [https://postfiles.pstatic.net/20160507_227/cattree_studio_1462576806699mWyB2_GIF/render_%283%29.gif?type=w773](https://blog.naver.com/PostView.nhn?blogId=cattree_studio&logNo=220703210504&redirect=Dlog&widgetTypeCall=true&topReferer=https%3A%2F%2Fwww.google.com%2F&directAccess=false)  
   여기서 C0는 기존의 cost, lambda는 regularization strength, w는 weight이다. Lambda 값으로 얼만큼의 penalty를 부과할 것인지 결정할 수 있다. 여기서 lambda 값을 어떻게 설정해야 할 지가 중요하다. 만약 Lambda 값이 너무 작으면 overfitting 현상을 해결할 수 없으며, lambda 값이 너무 크면 cost function이 과하게 smooth하게 되어서 underfitting 현상이 발생하게 된다.   
     
   MNIST data set은 우편번호 자동 처리를 위해 만들어 놓은 손 글씨 모음 data set이다.   
     
   위에 주어진 MNIST data set으로 예시를 들면, 만약 ‘7’이라는 숫자에 대한 training을 하는데 주어진 data에 대해서만 과하게 학습이 되어있다고 하자.  
      
   이 때 위에서 주어진 글씨가 아닌 아래와 같은 새로운 손글씨의 7 data가 들어오면 어떻게 될까?  
     
     
   처음에 주어진 ‘7’ data set에 대해서만 학습되어 있기 때문에 이 숫자를 7이라고 인식하지 못 하고 9라고 인식할 수도 있다. 이러한 현상을 overfitting이라고 하는 것이다.  
   그렇다면 이를 해결하기 위해 7과 9을 구분할 때 주어진 data set에 너무 특정되게 구분하지 않고 function을 smoothing 하는 과정이 필요한데 이것이 regularization이다. 정규화 식에서 lambda 값을 적절히 조절하여 regularization 하면 새로운 형태의 7이 들어와도 이것을 7이라고 인식할 수 있게 된다.
2. Learning rate 값을 어떻게 정하느냐에 따라 학습의 정확도가 바뀔 수 있다. 일반적으로 Learning rate는 여러 번의 테스트를 통해 결정되는 경우가 많다. Learning rate와 학습 정확도의 관계에 대해 설명하여라.  
     
   learning rate 값이 적절하게 set 되면 학습의 정확도가 높아진다. 이때 learning rate의 값은 너무 커서도, 너무 작아서도 안된다. 이를 만족하는 learning rate를 실험적으로 적절하게 결정하는 것이 중요하다. 아래 그림을 보자.  
   [](https://blog.naver.com/PostView.nhn?blogId=cattree_studio&logNo=220703210504&redirect=Dlog&widgetTypeCall=true&topReferer=https%3A%2F%2Fwww.google.com%2F&directAccess=false)  
   만약 Learning rate 값을 너무 크게 설정하면 최적 값으로 수렴하지 않고 그래프 사이에서 진동하거나 아예 그래프 밖으로 벗어나 발산(Overshooting)하는 경우가 발생한다. 만약 값이 터무니없이 크게 나온다면 이런 경우를 의심해 볼 수 있다. 그러나 learning rate가 너무 작게 설정되면 수렴하는 속도가 매우 느리기 때문에 시간 제한이 있으면 제대로 된 최적 수렴 값을 찾기 전에 멈추게 될 것이며, 또한 이동 범위가 너무 좁아서 local minimum에 빠질 확률이 증가한다.