NN 심화 서술형 과제)

1. **He 초기화를 사용하여 무작위로 선택한 값이라면 모든 가중치를 같은 값으로 초기화 해도 괜찮을까요?**   
   A: 어떠한 방법을 이용하여 가중치를 초기화 하던 간에. 모든 가중치를 같은 값으로 초기화하면, 선형 결합이든 함수의 계산 값이든 그 값이 같기 때문에 가중치 업데이트에 영향을 줄 수 없다.
2. **ReLU는 왜 좋은가요? ReLU보다 ELU가 더 나은 점은 무엇인가요?**  
   vanishing gradient 문제를 줄일 수 있다.  
   ReLU 은 input 음수나 0이면 0을 반환하는데, 이는 학습을 시키지 않게 된다. 따라서 정보 손실을 야기할 수 있기 때문에 그 점을 보안한 것이 ELU이다.
3. **어떤 경우에 ELU, LeaklyReLU, ReLU, tanh, sigmoid, softmax 같은 활성화 함수로 사용해야 하나요?**  
   ELU와 LeaklyReLU는 Input 이 0이하일 때도 노드를 활성화시키고 싶을 때 (학습시키고 싶을 때) 사용하면 좋다.  
   그에 반해 ReLU는 input 이 0이하 일 땐 노드를 비활성화하고 학습을 시키지 않음으로, 특정 노드에 집중을 하거나 프로그램의 계산 량을 줄이고 싶을 때 사용하면 좋고, 보편적으로 ReLU를 많이 사용하는 편이다.  
   Softmax는 전체 확률에서 그 노드의 확률 값을 계산한다. 즉 모든 확률의 합이 1이 되기 때문에 전체 노드에서 그 노드가 활성화된 정도를 보기 좋다.  
   Sigmoid는 SoftMax와 극단 값에서 미분 값이 0이되기 때문에 Layer가 많은 경우엔 결국 0으로 수렴하는 문제가 발생하여 좋지 않다.  
   Tanh는 sigmoid를 활성화함수로 썼을 때 보다 학습 수렴속도가 빠르다.
4. **Momentum Optimizer을 사용할 때 Momentum 하이퍼파라미터 너무 1에 가깝게 하면 어떤 일이 일어날까요?**   
     
   v(t-1)이 이전 벡터인데, 감마를 1로 너무 가깝게 하면 이전 벡터 방향으로 추가적으로 과하게 이동하기 때문에, 과거에만 의존할 수가 있다.
5. **드롭 아웃이 훈련 속도를 느리게 만드나요? Predict도 느리게 만드나요?**  
   Dropout은 망에 있는 모든 layer에 대해 학습을 수행하는 것이 아니라, 일부뉴런을 dropout하고 줄어든 신경망을 통해 학습을 수행하여 ovefitting을 막는 방법 중 하나이다. 그렇기 때문에, 훈련속도를 느리게 만들지 않는다.
6. **Adam optimizer가 좋은 이유는 무엇인가요?**딥 러닝의 고질적인 문제 하나가, 데이터 양이 늘어나면 연산이 많아져 수행속도가 늘어지는 것인데, 역전파 과정에서 Adam은 Gradient descent 진행방향과 속도를 적절하게 조정해줌으로써 전체적인 분석 속도를 줄여준다.