

과제 #2 – 다중계층 신경망을 이용한 얼굴 표정 분류기 작성

코드 실행 방법:

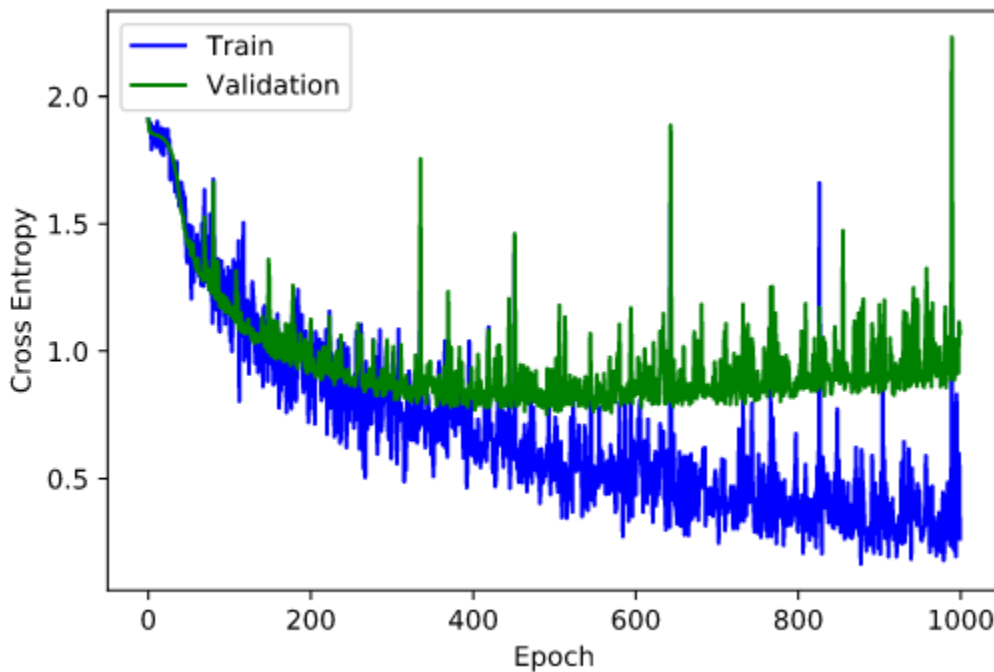
nn.py의 main 함수의 mode 변수 값을 문제 번호로 설정한다.

(예 : 문제 1번 코드를 돌리고 싶을 경우 mode = 1 로 설정)

코드를 수행할 경우, nn.py가 존재하는 경로에 figure 폴더가 생기고 그 안에 문제 번호 폴더가 생성되어 해당 폴더 내에 figure와 실행 결과 텍스트 파일이 저장된다. Error figure은 파일명에 _E가 붙고, Accuracy figure은 파일명에 _A가 붙는다.

1. 기본적인 일반화

오차 그래프



기본 setting:

eps = 0.01, momentum = 0.0, num_epochs = 1000, batch_size = 100

Epoch가 지날수록 Train Error는 계속 줄어드는데 비해 Validation Error는 일정 값으로 수렴하였다가 다시 상승증임을 알 수 있다. 이는 모델이 Train Set에 대해 반복적으로 학습하여 Train Set에 대한 성능은 증가하였지만 과적합이 발생하여 일반화 성능은 점점 떨어지게 되는 것을 나타낸다.

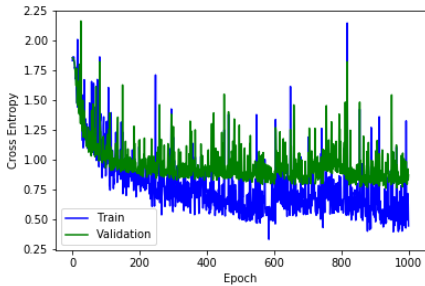
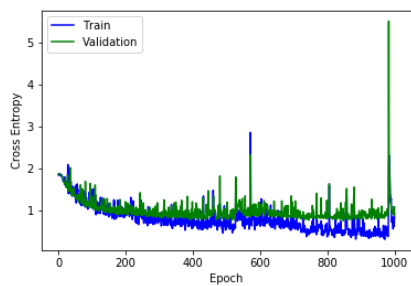
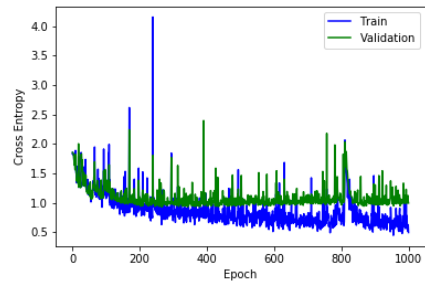
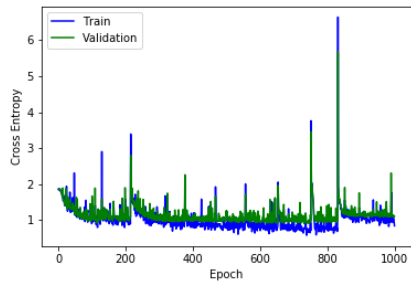
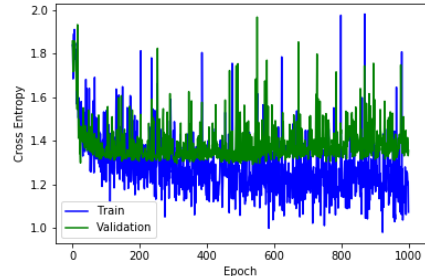
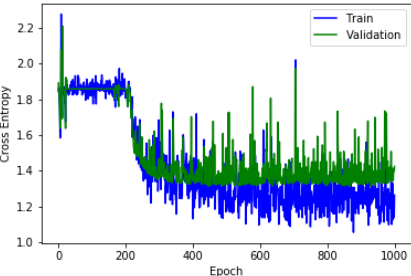
2. 최적화

Learning rate(LR)를 키우면 최저점까지 도달하기 힘들게 되고, LR을 작게 하면 수렴 속도가 낮아진다. 하지만 LR이 작을 경우, 지역 최저점에 갇히는 문제가 있다. 따라서 LR은 여러 경우를 테스트하기 위해 0.05, 0.01, 0.005, 0.001 네 가지 경우를 시험하였다.

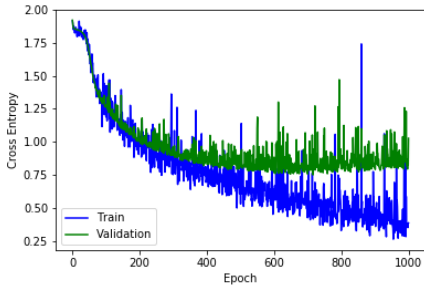
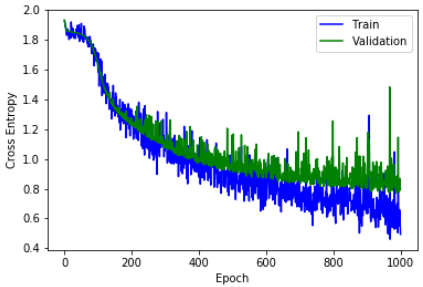
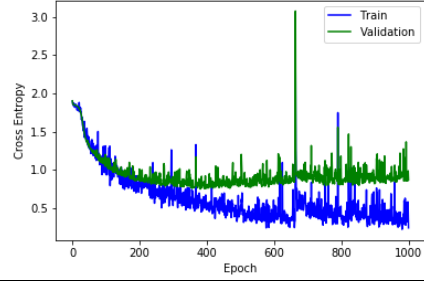
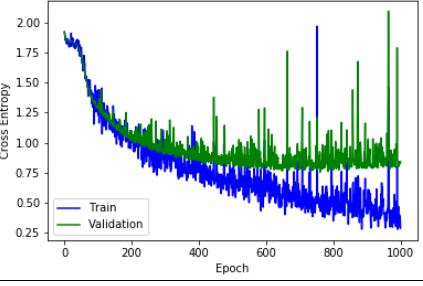
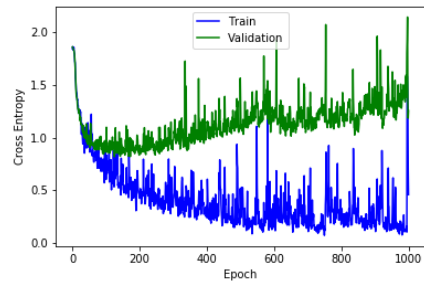
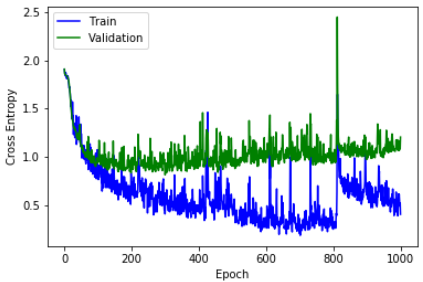
Momentum은 이전의 gradient를 반영하는 비율이다. 이 값이 적절하면 학습 속도가 더욱 빨라질 것으로 예상된다. Momentum은 0.3, 0.6, 0.9의 세 가지 경우를 시험하였다.

Mini-batch는 batch마다 학습이 이루어진다. 일반적으로 Mini-batch를 낮게 설정하면 빠른 수렴을 기대할 수 있지만, 너무 학습이 많이 이루어지면 과적합이 발생할 수 있다. 1번 기본적인 일반화에서 500 epoch에서 수렴함을 확인했으므로 Mini-batch size는 200, 400 두 가지 경우를 시험하였다.

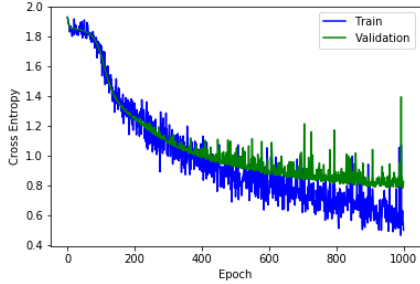
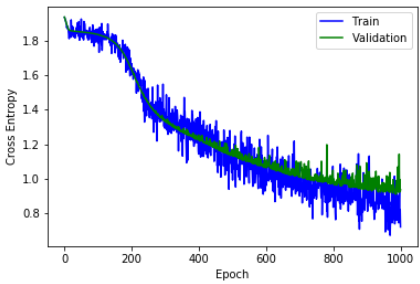
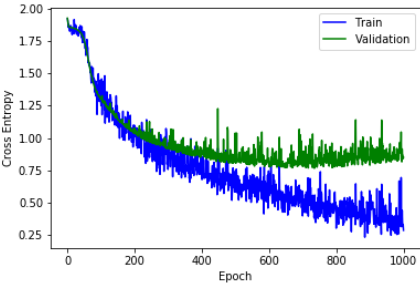
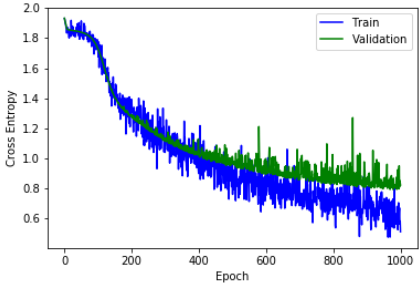
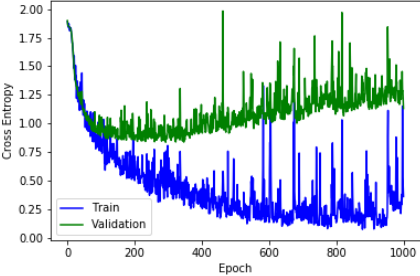
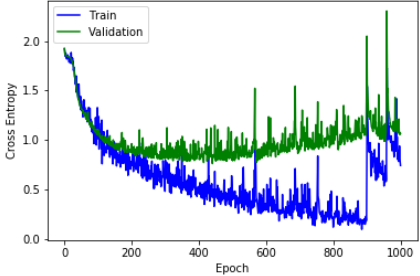
LR : 0.05

Momentum	Batch size = 200	Batch size = 400
0.3	Test Error : 0.893 	Test Error : 1.067 
0.6	Test Error : 0.992 	Test Error : 1.062 
0.9	Test Error : 1.363 	Test Error : 1.408 

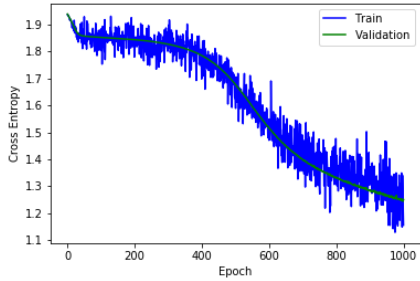
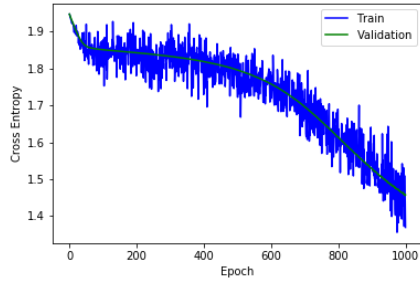
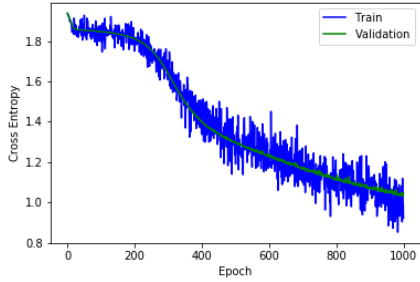
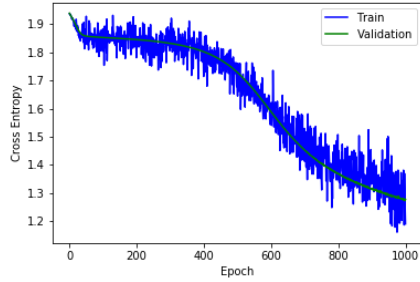
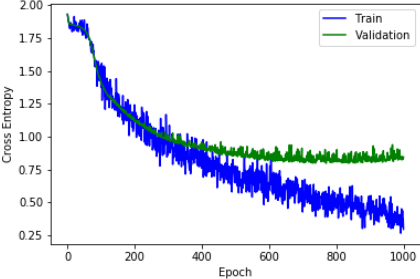
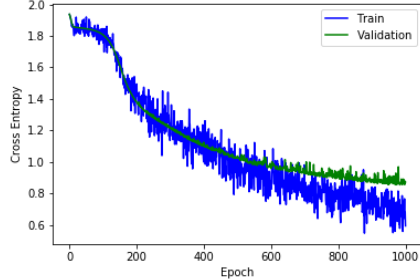
LR : 0.01

Momentum	Batch size = 200	Batch size = 400
0.3	Test Error : 0.929 	Test Error : 0.798 
0.6	Test Error : 0.728 	Test Error : 0.735 
0.9	Test Error : 1.263 	Test Error : 1.051 

LR : 0.005

Momentum	Batch size = 200	Batch size = 400
0.3	Test Error : 0.827 	Test Error : 0.940 
0.6	Test Error : 0.776 	Test Error : 0.838 
0.9	Test Error : 0.965 	Test Error : 1.003 

LR : 0.001

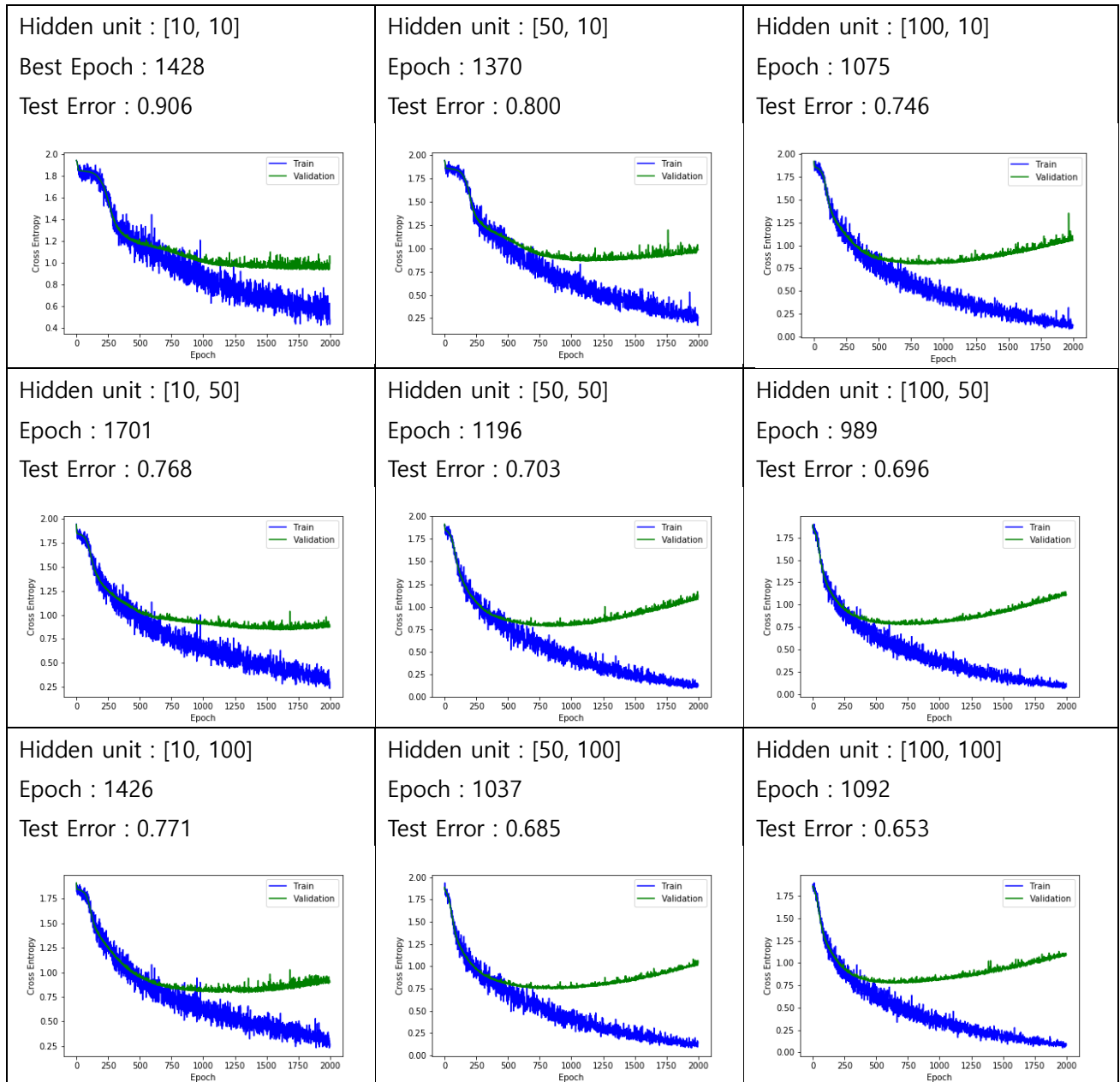
Momentum	Batch size = 200	Batch size = 100
0.3	Test Error : 1.275 	Test Error : 1.461 
0.6	Test Error : 1.071 	Test Error : 1.298 
0.9	Test Error : 0.824 	Test Error : 0.874 

LR : 0.01, Momentum = 0.6, Batch size = 200일 때 Test Error에 대해 가장 높은 성능을 보였다.

3. 신경망 모델 구조 변경

코드의 Neural Network에는 2개의 Hidden layer가 존재한다. 각각의 Hidden layer의 유닛을 10, 50, 100에 대해 테스트하여 9개의 Cross Entropy 그래프를 비교하였다. LR은 0.001, Batch size는 400으로 고정하였다.

각각의 경우, 수렴하는 epoch가 다르고 코드를 재실행 할 때마다 수렴하는 epoch가 달라지기 때문에 원활한 비교를 위해 2000 epoch 중 가장 낮은 Test Error의 성능을 비교하였다.



Hidden unit의 개수가 늘어날수록 빨리 수렴되고 성능이 좋아지는 경향을 보인다. 또한 hidden unit = [10, 100]인 경우보다 [100, 10]인 경우에 성능이 더 높고, 두번째 hidden unit의 수가 첫번째 hidden unit의 수보다 적은 경우, 많은 경우에 비해서 빨리 수렴되는 경향을 보인다.