

## Practice#3

2017029970 우원진

**Table1(batch\_size = 128)**

	BinaryCrossentropy	MeanSquaredError
Accuracy(with train set)	0.9997	0.9995
Accuracy(with test set)	1.0	1.0
Train time(sec)	56.6353	53.7645

**Table2(batch\_size = 128)**

	SGD result	RMSProp	Adam
Accuracy(with train set)	0.9997	0.9996	0.9997
Accuracy(with test set)	1.0	1.0	1.0
Train time(sec)	56.6353	59.6991	59.7737

**Table3**

	Mini-batch = 4	Mini-batch = 32	Mini-batch = 128
Accuracy(with train set)	0.9991	0.9996	0.9998
Accuracy(with test set)	1.0	1.0	1.0
Train time(sec)	1429.185	212.1505	57.1142

## Discussion

1. Table1은 같은 조건에서 loss function을 다르게 하여 실험을 한 것이다. “BinaryCrossentropy”와 “MeanSquaredError”를 비교하여 실험했는데, 속도는 MeanSquaredError가 조금 더 빨랐다. 이는 BinaryCrossentropy에서 log가 사용되어 그런것으로 생각된다. 그리고 Test Set에 대한 정확도는 둘다 100%로 같았지만, Training Set에 대한 정확도는 BinaryCrossentropy가 엄청 조금 더 높았다. 이는 MeanSquaredError는 Min값에 가까이 갈수록, 변화량이 작아져서, BinaryCrossentropy보다 최솟값에 가까이 도달하지 못해 그런것으로 생각된다.
2. Table2는 같은 조건에서 Optimizer만 SGD, RMSProp, Adam으로 나누어 실험한 것이다. 이 실험에서는 Train Set에 대한 정확도와 Test Set에 대한 정확도 모두 거의 같은것으로 나타났다. 이론대로면 항상 그런것은 아니지만 Adam의 결과가 가장 좋게 나와야할것 같았는데, 이 실험은 문제의 난이도에 비해 충분히 많은 데이터셋과, Epoch가 있기 때문에 셋다 높은 정확도를 보이는것 같았다. Train time에는 SGD가 가장 빨랐고, RMSProp과 Adma은 거의 비슷했다. 하지만 Andrew Ng 박사의 강연을 유튜브를 통해 잠시 들은적이 있는데, 거기서 안전한 모델을 만들기 위해 Optimizer로 Adam을 쓰라는 말을 하였다.
3. Table3은 같은 조건에서 Mini-batch의 크기만 다르게 해서 실험을 진행했는데 batch의 크기

는 4, 32, 128 이었다. 이 셋 모두 Test Set에 대한 정확도는 100%로 같았지만, Train Set은 아주 근소한 차이로, batch size가 128인것이 가장 정확도가 높고, 그 다음으로 32, 그 다음 4로 나타났다. Train time에서는 엄청난 차이를 보였는데, batch size가 작을수록 시간이 많이 걸렸고, batch size = 4 인것과 128 인것의 차이는 4인것이 약 25배 정도의 시간이 걸렸다.