

MNIST 학습 실험 보고서

빅데이터마이닝 과제

소프트웨어학과

32203598 이창수

#1 기준을 잡기 위해 baseline(100, 32, 0.05, 50)으로 실험한 결과 0.694의 test accuracy가 나왔다.

#2 시드 값을 고정하고 하는 실험이 아니기에 동일한 조건에서 얼마나 편차가 나는지 알아보고자 같은 조건으로 한번 더 실험한 결과 0.634가 나왔다. 시드도 문제지만 epoch에 비해 iteration과 batch size가 턱없이 작은 것도 영향이 있는 듯하다.

수행 시간에 제약이 있기 때문에 리소스에 대한 파악이 우선이었다. baseline 코드 진행은 1시간이 조금 안되는 실행시간이 나왔다. 수치미분 기반의 계산이기 때문에 연산량은 iteration과 batch size에는 정비례하지만 hidden size에는 제곱으로 비례한다.

#3 수행시간에 가장 큰 영향을 미치는 것은 hidden size이므로 이에 대한 적절한 크기를 찾는 것이 우선이라고 생각했다. 50에서 75로 증가시켰지만 코스트에 비해 큰 효과는 없었다.

#4 hidden size가 영향이 크지 않은 것 같아 이번엔 iteration과 batch size를 건드려봤다. 꽤나 큰 성과가 있었다.

#5 같은 조건에서 hidden size를 증가시켜보았다. 유의미한 변화가 전혀 없었다.

#6 혹시싶어 또 해봤지만 의미없었다. 아무래도 hidden size는 남은 리소스 내에서만 다루어야 할 것 같다.

#7 지금까지의 진행상황으로 GPT에게 고견을 구하여 시도해보았다. 크게 상승했다.

#8 아무래도 학습을 할 수 있는 충분한 횟수가 필요한 것 같아 극단적으로 시행해 보았지만 무의미했다. 600 이후로는 개선이 없었다.

#9 iteration의 한계가 600이라 생각하며 남은 리소스를 hidden size에 할당해보았으나 더 안좋은 결과가 나왔다.

#10 batch size가 64일 때 더 좋은 결과가 나오는 것 같아 비슷한 학습 횟수에서 리소스를 최대한으로 끌어올려 마지막 실험을 한 결과 가장 좋은 성능을 보였다.

#1 100, 32, 0.05, 50 : 0.694
#2 100, 32, 0.05, 50 : 0.634
#3 100, 32, 0.02, 75 : 0.721
#4 150, 64, 0.03, 50 : 0.820
new code
#5 160, 64, 0.03, 75 : 0.816
#6 160, 64, 0.03, 75 : 0.792
#7 320, 64, 0.02, 75 : 0.864
#8 1000, 32, 0.02, 50: 0.843
#9 600, 32, 0.03, 60 : 0.838
#10 400, 64, 0.02, 75 : 0.874

파라미터 조합에 따른 결과값 로그



