

06. 배치 사이즈

2020년 12월 31일 목요일 오전 9:43

- 배치 사이즈는 가중치와 bias 를 수정하는 간격을 의미
- 학습 효율에 영향을 줌

1. 에포크(epoch) 와 배치(batch)

- 모든 훈련 데이터를 1회 학습하는 것을 1 epoch 라 함
- 훈련 데이터의 샘플을 여러 개 묶어 학습에 사용, 이 훈련 샘플 그룹을 배치라 함
- 1 에포크는 m개의 배치로 구성됨



- 배치 사이즈는 하나의 배치에 포함되는 샘플의 수
- 하나의 배치에 포함된 모든 샘플을 사용해 가중치와 bias를 수정
- 배치 사이즈는 학습 중에 일정

2. 배치 학습

- 배치 학습에서 배치 사이즈는 전체 훈련 데이터의 수
- 1 에포크마다 전체 훈련 데이터 오차의 평균을 구해 가중치와 bias를 수정
- 일반적으로 안정된 학습이 진행, 상대적으로 빠른 속도
- local minima에 빠지기 쉽다
- 훈련 데이터의 수를 N, 개별 데이터의 오차를 E_i 라고 할 경우 다음과 같이 정의

$$E = \frac{1}{N} \sum_{i=1}^N E_i$$

- 가중치의 기울기는 다음과 같음

$$\frac{\partial E}{\partial w} = \sum_{i=1}^N \frac{\partial E_i}{\partial w}$$

가중치 기울기를 배치 내의 개별 데이터마다 계산, 그 기울기들의 합계를 구함

- 훈련 데이터의 샘플 수가 1000개 일 경우, 배치 사이즈가 1,000, 1 에포크에 가중치와 bias가 1회 수정

3. 온라인 학습

- 배치 사이즈가 1
- 개별 샘플마다 가중치와 bias를 수정
- 개별 데이터로부터 영향 받기 때문에 안정성이 떨어짐
- 하지만 local minima 에 빠지는 것을 방지

** 이 전 경사하강법 기울기 구하는 방법이 온라인 학습

- 훈련 데이터의 샘플 수가 1000개 일 경우, 배치 사이즈가 1, 1 에포크에 가중치와 bias가 1,000회 수정

4. 미니 배치

- 훈련 데이터를 작은 그룹(배치)으로 분할, 이 그룹마다 가중치와 bias를 수정
- 온라인 학습에 비해 배치 사이즈가 크기 때문에 특이 데이터에 대한 민감도가 적음
- 배치 사이즈를 $n(n \leq N)$ 으로 하면 다음과 같다.

$$E = \frac{1}{n} \sum_{i=1}^n E_i$$

$$\frac{\partial E}{\partial w} = \sum_{i=1}^n \frac{\partial E_i}{\partial w}$$

- 훈련 데이터의 샘플 수가 1000개 일 경우, 배치 사이즈가 50 이면, 1 에포크에 가중치와 bias가 20회 수정

- 배치 사이즈가 학습 시간과 성능에 영향을 준다는 것은 경험적으로 알려져 있음
- 배치 사이즈를 적절하게 설정하는 것은 상당히 어려운 문제
- 일반적으로 10 ~ 100 정도를 설정