

#2

Deep learning optimization

Loss function

- 개념
 - 지도학습 시 알고리즘이 예측한 값(추력값)과 실제 정답의 차이를 비교하기 위한 함수
 - =Cost function = Objective function = Energy function
- 특징
 - 출력값과 실제 정답이 일치할수록 손실함수의 값은 작아지고, 불일치할수록 손실함수의 값은 커짐
 - 성능척도와는 다름
- 평균제곱오차(MSE)
 - 오차가 클수록 아래 e값이 커지 → bad
- 경사하강법
 - 손실함수 값을 최저로 만드는 대표적인 최적화 방법
- 교차 엔트로피
- 교차 엔트로피 목적함수
- Softmax
- 로그우도 목적함수

Performance improvement

- Pre-Processing
 - 정규화
 - 표준화
 - one-hot 코
- Weight Initialization
 - 가중치의 초기값을 무엇으로 설정하는냐가 신경망 학습에 영향을 끼치게 됨

- 가장 작은 error가 도출 → 정확한 모델
- Gradient Decent & Momentum
 - Gradient Decent
 - 학습시 step size중요
 - 무작위 초기화로 인해 알고리즘이 전역 최소값이 아닌 지역 최소값에 수렴할 수 있음
 - Momentum
 - 학습 방향이 바뀌지 않고, 일정한 방향을 유지하려는 성질
 - 같은 방향의 학습이 진행된다면 가속을 가지며 더 빠른 학습을 기대할 수 있음
- Adaptive Methods
 - 학습률의 중요성
 - 작으면 학습에 많은 시간소요
 - 크면 진동할 가능성이 높음
 - 매개변수마다 자신의 상황에 따라 학습률을 조절해 사용
 - SGD
 - Adagrad
 - RMSprop
 - Adam
- Epoch
 - 훈련 데이터 셋에 포함된 모든 데이터들이 한 번씩 모델을 통과한 횟수
 - epoch 값이 높을수록 다양한 무작위 가중치를 학습하였다는 의미
 - 지나치게 epoch가 올라가면 학습 data set에 대한 과적합 발생
 - Batch size
 - 연산 한번에 들어가는 데이터 크기
- Batch Normalization
 - Internal Covariate Shift 현상(재부 공변량 이동)
 - 레이어가 깊어질수록 분포 형태가 더 변화가 심함 → Batch Normalization을 이용
 - 장점

- 데이터 scale 통일
- 활성화 함수 맞춤형 분포 변
- ReLU Activation Function
 - 고차원 데이터를 다룰 경우 값이 커질 수 있어 다시 경사가 소실될 수 있음
 - 복잡한 데이터일수록 고차원일 경우 많은데 이를 회피할 수 있는 활성화 함수임
- Stochastic Pooling
 - Pooling