

정규화

다섯 가지 정규화

강민진(출처 : 파이썬 날코딩으로 알고 짜는 딥러닝)

PNU

- 1 정규화
- 2 과적합과 부적합
- 3 5가지 정규화 기법
- 4 L2 손실
- 5 L1 손실
- 6 Dropout
- 7 Noise Injection
- 8 Batch Normalization

정규화

- 일부러 적당한 제약을 가해 학습을 방해하는 기법
- 신경망이 학습 데이터에 지나치게 맞추어지는 과적합 방지가 목적
- 일반화 단계 : 딥러닝 모델이 실제 업무에 투입되어 이용되는 단계.
딥러닝 모델의 개발 목적 : 일반화 단계에서의 다양한 데이터 처리
학습 데이터에 과도하게 익숙 -> 일반화 단계에서의 성능 저하 초래 위험

과적합과 부적합

- 러닝 모델 학습 실패의 두 가지 주된 이유

- 부적합

문제 난도에 비해 모델 용량이 부족해 학습용 예제를 제대로 못 푸는 현상
학습 과정에서 정확도 향상이 잘 안되며 당연히 평가 단계에서도 부진
모델 용량 확대, 학습 횟수 증가, 데이터의 양적, 질적 확대 등의 조치 필요
어떻게 대처해야 할지 방향을 잡기는 비교적 쉽다

과적합과 부적합

- 과적합

전체적인 문제 특성 파악 없이 학습 데이터의 지엽적 특성 암기

학습 단계에서는 좋은 성과, 평가 단계에서 정확도 크게 저하.(일반화 단계에서도 성능이 현저히 떨어지기 때문에 적용 어렵다.

문제 난도에 비해 모델 용량 크면서 데이터가 부족할 때 주로 발생한다.

- 과적합 처리의 어려움

학습 횟수 늘려도 새로운 데이터 공급 없이는 성능 향상 기대 힘들

가장 좋은 방법은 양질의 데이터를 충분히 확보해 학습에 이용하는 것

- 정규화 기법

학습용 예제의 지엽적 특성 암기를 어렵게 만드는 방안 필요 -> 정규화

딥러닝 알고리즘에 일부러 학습을 방해하는 제약 추가 -> 일반화 과정에서의 성능 향상 기대

5가지 정규화 기법

- L2 손실, L1 손실, Dropout, 잡음 주입(Noise Injection), Batch Normalization
- 지원 형태
L2, L1 : 기존 계층들의 순전파 및 역전파 처리 방식 변형
Dropout, 잡음 주입, Batch Normalization : 새로운 계층
- 학습 단계와 평가 단계
L2, L1 : 역전파 처리에 반영되므로 학습 단계에만 영향
Dropout, 잡음 주입 : 학습 단계에서만 동작하도록 제어
Batch Normalization : 학습 단계와 평가 단계에 서로 다른 처리
- 정규화 기법
정규화 기법을 통해 일반화 단계에서의 성능 향상

L2 손실

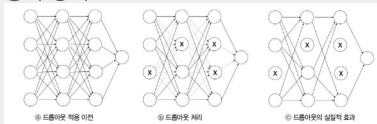
- 절댓값이 큰 파라미터에 불이익을 주는 정규화 기법(제곱 이용)
파라미터 값의 폭주를 막고 기왕이면 작은 절댓값을 갖도록 압박
본연의 학습 효과 때문에 무조건 모든 파라미터의 절댓값이 줄지는 않음
반드시 성능 향상으로 이어진다는 보장은 없음
- L2 손실과 손실 함수
 $\text{L2 손실에서의 손실 함수} = \text{기존의 손실 함수} + \text{L2 페널티}$ (편향 파라미터에는 L2 손실을 적용하지 않음)
- L2 손실과 손실 기울기
파라미터 갱신 때 기존의 손실 기울기에 L2 손실률과 가중치를 곱한 값을 더해서 처리
미니배치 학습 때마다 파라미터 값 일부를 납부

L1 손실

- 절댓값이 큰 파라미터에 불이익을 주는 정규화 기법(절댓값 이용) 전체적으로 L2 손실과 비슷하지만 더욱 강한 압박
- L1 손실과 손실 함수
L1 손실에서의 손실 함수 = 기존의 손실 함수 + L1 페널티 (편향 파라미터에는 L1 손실을 적용하지 않음)
- L1 손실과 손실 기울기
파라미터 갱신 때 기존의 손실 기울기에 L1 손실률과 가중치를 곱한 값을 더해서 처리
미니배치 학습 때마다 파라미터 값 일부를 납부

Dropout

- 계층 입력 중 일부만 이용해 신경망을 학습시키는 규제화 기법 간단하면서도 큰 효과
- 동작 방식



- 처리 배제 성분 : 미니배치마다 다르게 랜덤 지정
순전파 처리 : 마스크 처리 통해 배제 성분 값을 0으로 변경
순전파 처리 : 마스크 처리 통해 배제 성분 손실 기울기를 0으로 변경
- 과적합 방지를 위한 기법이지 계산량을 줄이는 기법이 아님
오히려 기존 처리 과정에 마스킹 처리가 추가되어 계산량 증가할 수 있다.

Noise Injection

- 두 계층 사이에 삽입되어 은닉 벡터에 잡음을 추가하여 전달
- 순전파 처리
학습 단계에 한해 랜덤한 잡음을 생성해 주입
잡음이 주입되면 이후의 단계들은 학습에 혼란을 겪게 되지만 그 결과 강건한 학습이 이루어져 다양한 입력의 변이를 더 잘 처리할 수 있게 된다.
잡음 주입의 형태, 빈도, 강도는 보통 신경망 설계자가 설정 가능
- 역전파 처리
별도의 처리 불필요 : 출력 손실 기울기를 바로 입력 손실 기울기로 반환
- 잡음 주입 기법의 의미
입력 변이에 대한 강건한 학습 통해 평가 단계에서의 품질 제고
매번 다른 잡음 주입으로 인해 과적합 방지

Batch Normalization

- 미니배치 내 데이터들에 대해 벡터 성분별로 정규화를 수행
- 정규화(Normalization)
동일한 선형 변환으로 대상값들을 평균 0, 표준편차 1 분포로 만드는 처리
입력 성분 간 분포 차이로 인한 학습 불균형을 방지하기 위해 도입(regularization과 다름)
- 정규화의 필요성
역전파 처리 중 종종 파라미터 손실 기울기의 급격한 소멸/폭주 현상 발생
입력의 성분별 분포가 심하게 다를 때 더욱 쉽게 발생
- 정규화와 배치 정규화
정규화 : 모집단 전체 데이터를 대상으로 하는 처리
- 통계 처리나 다른 딥러닝 연구에서도 많이 이용하는 방법
배치 정규화 : 모집단이 아닌 미니배치 데이터를 정규화 대상으로 삼음
- 매번 작은 규모의 미니 배치 데이터에 대해 별도로 정규화 수행

Batch Normalization 처리

- 학습 단계에서만 미니배치 데이터를 대상으로 적용
- 미니배치 데이터로 $[m, n]$ 형태의 입력이 주어졌을때
 - 1단계 : 벡터 성분별 그룹 n 개에서 각 그룹 안의 데이터값 m 개를 정규화
 - 먼저 처리 대상 데이터들의 평균과 표준편차를 구한 후
 - 각각의 데이터값에서 평균을 뺀 후 표준편차로 나누어 진행
 - 미니배치 정규화 과정을 거치면
 - 벡터 각 성분 데이터들은 모두 평균 0, 표준편차 1의 분포를 갖게 되어 성분별 불균형이 적어도 미니배치 내에서는 사라지게 된다.
 - 2단계 : 정규화된 값에 크기 요소 곱하고 이동 요소 더하는 선형 연산 처리
 - 크기 요소(scale factor)와 이동 요소(shift factor) : 처음에 각각 1과 0으로 초기화되어 학습 진행에 따라 문제 처리에 적합한 값 갖게됨
 - 없던 불균형을 다시 만드는 과정이라고 생각할 수 있으나 문제풀이에 적합한 새로운 분포를 찾아 대치(앞이나 뒤에 선형 연산이 있는 경우 실행 불필요, 원하는 경우에만 실행)

Batch Normalization 효과

- 데이터셋 전체 정규화

장점 :

- 한 번에 일괄 적용 후 결과를 계속 재활용하므로 더 효율적
- 정규화된 값이 일정하므로 안정된 학습 기대

단점 :

- 데이터셋이 아닌 은닉 벡터에는 적용 불가

- 미니배치 정규화

장점 :

- 데이터셋이 아닌 은닉 벡터에도 적용 가능
- Epoch마다 무작위로 바뀌는 미니배치 구성이 학습 과정 교란 -> overfitting 방지
- 모든 데이터를 학습에 충분히 이용(일부 성분만을 결정에 참여시키는 Dropout보다 유리)