

2024 09 05 수업 정리



#hidden layer 1개일때 단층 퍼셉트론 —ex)mnist

mnist를 다층으로 키워 볼 수가 있나??2층으로 키우면 성능이 오를까?? 일반화 성능이 떨어짐(오버피팅 판단은 어떻게? 테스트 데이터셋이 뭔가에 따라 다름.)

판단하는거 그래프로 판단 loss평선이 0에 가까워지다가 어느순간 올라가는 구간이 있는데 0에 가까워지는 부분이 overfitting 과적합이라고 한다.

로스가 일정한 순간이온다(가중치의 변화가 없다)

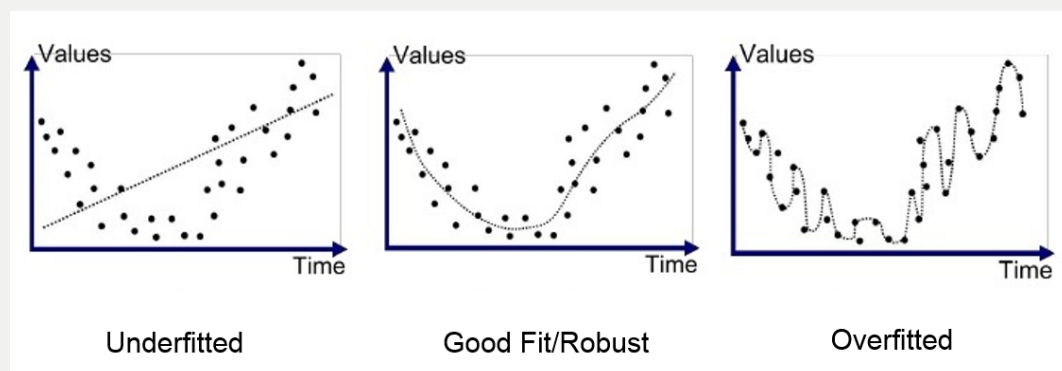
파라미터 업데이트 할 게 많아진다. W (가중치의 의미) 가중치란 데이터의 특성을 담기 위한 그릇을 의미, 그릇이 적으면 입력신호는 그대로 받으나

가중치를 많이하면 된다. 모델을 키운다. 그러면 W가 많아지면서 담을 수 있는 특징의수치가 많아진다.

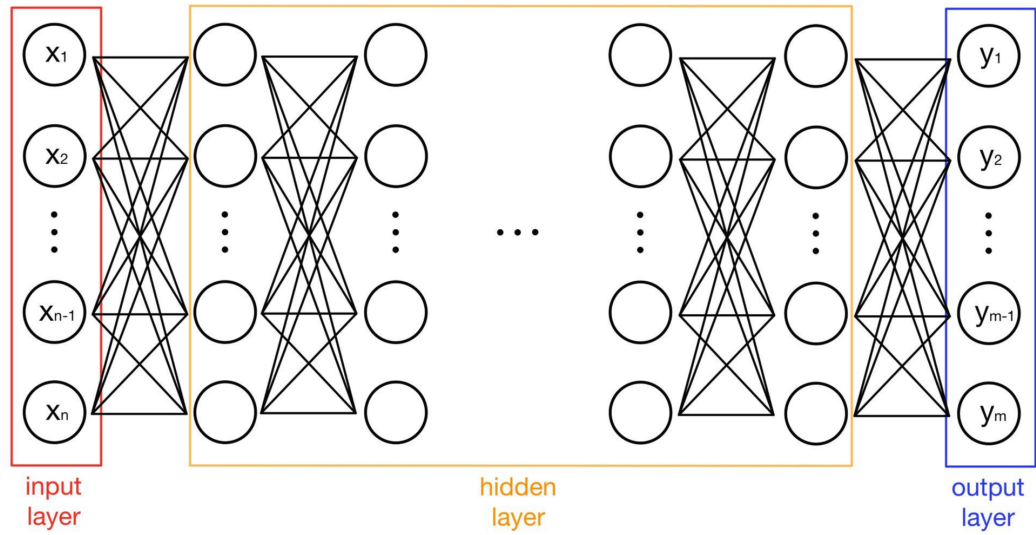
5차식은 학습의 의미가 없기 때문에 1차식으로 들어간다.

→데이터의 특징을 잘봐야 한다.

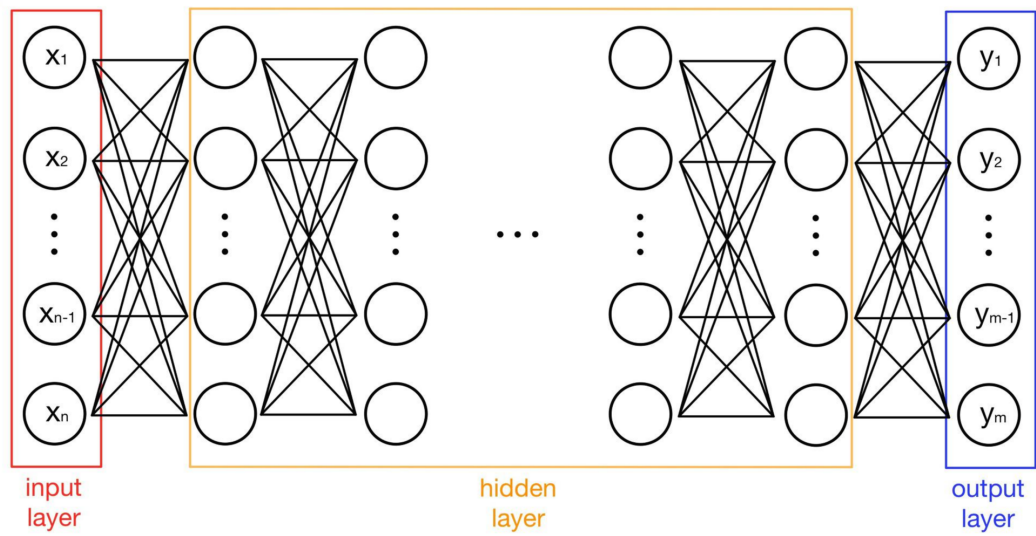
다차그래프로 그리면 ...



#hidden layer 여러개 다층 퍼셉트론 (뉴럴 네트워크-hidden layer여러개)-옵티마이저워



hidden이 여러개라서 다층 퍼셉트론



hidden이 여러개라서 다층 퍼셉트론



특징

컨볼루션 필터

픽처 익스트랙터 -CNN (마인크래프트로 된 얼굴을 사람얼굴로 인식하게끔 하는 거) 특징이 겹쳐지는 것들을 파악

ex)쪼개서 뽑아내서 뉴럴 네트워크에 넣으면 학습을 잘한다.

컨볼루션 필터 먹이면 마인크래프트 얼굴을 사람얼굴로 인식

→근데 특징을 너무 잘게 잘라놓으면 형체를 알아 볼 수가 없다.(오히려 안겹치는 게 많아진다.)(쌀알을 다 잘게 자르면 밥이 아니라 죽이듯이) →ps. 적당히 분해 하는 것도 중요하다.

적당히 분해를 잘하는 방법은 직접 실험을 해 봐야 한다.

node는 특징 담는것

다차로 가는거는 layer늘리는 것

node수도 조절 해야한다. 보통 피라미드 많이 쓴다.

노드가 보통 중간에 많다(중간에 사선으로 짝 그인 그림)

A(1-100특성 겹치는)

B(50-150)

C(100-200특성이)

눈 한쪽 없는 사람??은 어떻게 판단

특징



loss function :실제값과 예측값의 차이를 나타내는 함수 ,
미분 가능 해야 한다.

컴백스 파다

왜 컴백스여야 하나?? W형 그래프이면 안된다.(loss fuction 의 최적점이 여러 개서 찾을 수 없어서 학습이 제대로 안됨)어디가 로컬 미니멈인지 글로벌 미니멈인지 모른다.

→ 컴백스는 최저점이 한개 라서 학습이 잘된다. (내가 짤)

초보자들은 미니멈이 잘짜여진 함수를 사용하는게 좋다. 로스 평선을 적재적소에 잘쓰는게 중요 ex)segmentation(분할) 작업이 있는데 사진(사람+ 공사진,라벨 마스크) →(공,프레디얼 마스크) 픽셀어큐러시 학습이 잘 안됨.보통 크로스 엔터피를 많이쓴다.(픽셀 하나하나에 대해 하기 때문에)



사람 공 사진 → 공 사진(label mask)A = 공사진 테두리(predic mask)B

$$A-B=0$$

$$y^{\wedge}-y^{\wedge}=ho(x)$$

*크로스 엔트로피는 (전체 차이 /현재 (그부분의)차이)를 의미
MSL(NS1?? MS1?)을 쓰이유를 말할 수 있어야 한다.

하나하나 다 구현해야

모델을 크게 만들 수록 좋을 까요? 아니요 상황에 따라 다릅니다.

너무 커도 안좋은 gradient소실이라는 문제가 일어남 적절한 크기를 맞추는 것이 좋다. 데이터를 기억 못하는 경우가 발생

transformer가 등장하면서 모델크기를 무한정으로 늘릴수있다.

transtformer 맵을 이용 (모델크기가 크면 클수록 좋음), (관계도라고 생각하면 된다.)

self attention 몸통과 다리의 관계 입과 허리의 관계 픽셀과 픽셀과의 관계를 계속 파악하며 데이터의 크기가 커짐(특징들을 다 담을 수 있다.)LLM의 경우 다담음→어텐션 때때 다담음

