분류선이 아주 꼬불꼬불하면 어떤 상황인가?

-모델이 데이터의 이상치, 노이즈 등을 모두 학습해 분류선이 그런 값을 전부 따라가는 상황이므로, 해당 모델이 과적합된 상황이다. 모델이 과적합되면, 해당 모델은 자신의 학습에 쓰인 데이터에 대해서는 매우 좋은 성능을 보여주지만, 그 밖의 데이터에 대해서는 충분히 좋은 성능을 보여주지 못한다.

지도학습, 반지도학습, 강화학습에 대해 설명하시오.

-이 3가지 학습 방식은 AI 모델을 학습시킬 때 사용하는 방식이다. 학습시키는 데이터에 정답이 있는지 없는지에 따라 종류를 나눈다.  
지도학습은 모든 데이터에 정답이 있는 경우이다.  
반지도학습은 일부 데이터만 정답이 있고 나머지는 없는 경우이다. 데이터에 정답이 하나도 없는 상태로 학습하는 것은 비지도학습이라고 한다. 현실에 존재하는 데이터 대부분은 정답이 존재하지 않고, 모든 데이터에 정답을 추가하는 것은 현실적으로 불가능하다. 그런 상황에서 일부 데이터에만 정답을 추가해주어 반지도학습을 하면 비지도학습에 비해 모델 성능이 오른다고 한다.  
강화학습은 데이터에 정답이 없는 대신, AI 모델이 낸 결과가 좋을수록(원하는 결과에 가깝거나, 목표 수치가 높거나 등) AI에게 보상을 주고, 나쁠수록 벌점을 주는 방식이다. 게임 등 뚜렷한 정답이 없는 문제에 이용한다.

경사하강법이 무엇인가?

-모델을 반복학습할 때, 한 번 학습할 때마다 loss 함수 등을 이용해 해당 모델의 파라미터(특히 가중치)를 조정해주는 함수를 optimizer라고 하는데, 경사하강법은 그 optimizer의 한 종류다.  
loss 함수를 미분한 값, 곧 loss 함수 그래프의 기울기를 이용해 가중치를 조절하며 점점 기울기가 0에 가까운 곳으로 가중치를 이끌고 가기 때문에 ‘경사하강’법이라고 부른다.  
수식은 w(다음) = w(이전) – (r \* G)  
여기서 w는 가중치, r은 학습률, G는 loss 함수를 미분한 값이다.

Optimizer를 사용할 때 데이터를 이용하는 방법 3가지를 설명하시오

-optimizer를 사용해 학습할 때, 한 학습에 모든 데이터를 사용할지, 일부 데이터만 사용할지, 어떻게 사용할지에 대한 3가지 방법이 있다. 각각 장단점이 있다.  
batch: 한 번 학습할 때 모든 데이터를 한 번에 묶어서 사용. 안정적이지만 여전히 학습 시간이 김.  
확률적: 한 번 학습할 때 데이터 중 하나를 랜덤하게 선택. 빠르지만 학습에 사용하는 데이터가 1개 뿐이므로 학습이 불안정하고 최적 가중치에 정확히 다다르기 힘듦.  
mini batch: 데이터를 특정 크기 묶음으로 나눈 뒤, 각 묶음을 사용. Batch, 확률적 두 방식의 절충안.

기울기 소실 문제가 무엇인가?

-모델의 가중치를 역전파 방식으로 업데이트할 때, 입력층에 가까워질수록 노드의 값 수정량이 0이 되거나 0에 가까워져 버리는 현상. 특히 sigmoid와 같은 활성화 함수를 사용할 경우, 입력값이 0이나 1에 가까우면 출력값의 기울기가 0에 가까워지는데, 이렇게 되면 역전파 방식에서 특정 노드가 전파시킬 기울기가 0에 가까워져 전파받은 노드의 가중치가 거의 수정되지 않게 된다.

Accuracy, F1, mse에 대해 설명하시오.

-accuracy는 머신러닝 모델을 평가할 때 사용하는 평가 지표이다. 모델이 예측한 데이터 중 맞춘 데이터의 비율이 얼마나 되는지 계산한 값이다. 하지만 데이터 불균형을 전혀 반영하지 못한다는 단점이 있다.  
F1은 그런 accuracy의 단점을 보완하기 위해 만든 평가 지표이다. 모델이 긍정으로 예측한 것 중 실제로 긍정인 비율, 실제로 긍정인 데이터 중 모델이 제대로 긍정으로 예측한 비율 2가지 지표를 이용한다. (**저 2개 지표 이름: Recall(재현율, 실제로 긍정인 것 중 모델이 긍정이라고 예측), Precision(정밀도, 모델이 긍정이라고 예측한 것 중 실제로 긍정)**) 앞 2가지 지표 중 하나라도 0에 가까우면 F1도 0에 가까워지기 때문에, 데이터 불균형을 더 잘 반영한다.  
F1 공식: 2 \* ((재현율 \* 정밀도) / (재현율 + 정밀도))  
mse는 딥러닝 모델을 평가하기 위한 손실 함수다. 평균 제곱 오차라는 의미. 실제 값과 예측 값의 차를 제곱한 값의 평균이다. 값이 제곱되기 때문에 모든 오차값이 양수가 되고 값이 매우 커진다. 결국 모델이 오차를 더 크게 받아들여 한 번에 더 크게 바뀌도록 할 수 있다. 하지만 데이터의 값이 매우 크거나 이상치, 극단값이 있을 경우 mse가 너무 크게 증가할 수 있다는 점을 주의해야 한다.  
mse 공식: (1/n) \* sigma((실제값 – 예측값)^2)

딥러닝과 머신러닝 차이?

-머신러닝은 데이터의 특징이 이미 설정되어 있지만, 딥러닝은 데이터의 특징을 모델이 알아서 추출한다는 점이 가장 큰 차이점이다. 딥러닝은 일반적으로 머신러닝에 비해 훨씬 많은 데이터가 필요하지만, 대신 훨씬 어려운 문제도 해결 가능하다.

Classification, regression의 차이?

-분류는 서로 관련이 없는 독립된 정답 중 하나의 정답을 골라내는 문제(예: 학생이 좋아할 과목 예측), 회귀는 서로 관련이 있는 연속된 값 중 하나의 값을 골라내는 문제(예: 학생의 성적 예측)이다.

차원의 저주란?

-데이터의 차원(특징) 수가 증가할수록 모델 학습이 어려워지는 현상을 말한다. 특히 각 데이터를 공간에 두고 거리 계산을 하는 모델이라면, 데이터의 차원이 늘수록 각 데이터 간 거리가 비슷비슷해지기 때문에 특히 학습이 어려워진다. 이를 해결하기 위해서는 특징 선택 등으로 특징 수를 줄여야 한다.

Ridge, Lasso의 공통점, 차이점?

-둘 모두 AI 모델의 규제 기법 중 하나다. 모델의 가중치 등 파라미터가 너무 높을 경우 그 값을 줄이거나 없애는 역할을 한다. Ridge는 값을 줄이기만 하므로 모든 변수의 영향력이 중요할 때 사용하고, Lasso는 값을 아예 0으로 만드므로 중요하지 않은 변수를 자동 제거하고 싶을 때 사용한다.

과적합, 과소적합은 무엇인가? 둘을 해소하려면 어떻게 해야 하는가?

Numpy, pandas-DataFrame, 배열(Excel)의 차이는?