인공지능 기초를 위한 FAQ 컴퓨터공학과 202334735 임서경

1. 인공지능에서 지능에 해당하는 기능은 무엇인가?

: 학습하는 능력, 추론하는 능력

1. 인공지능의 종류 3가지에 대해서 설명하시오 (지도학습, 반지도학습, 강화학습)

: 지도학습 – 입력과 정답 lable을 주고 그 데이터를 컴퓨터가 학습하면서 입력과 출력에 대한 일반적인 규칙이나 모델을 만드는 것

EX) 여러 장의 이미지에 ‘고양이’라는 라벨을 붙여두고 컴퓨터가 그 이미지를 ‘고양이’로 인식할 수 있는 모델을 만다는 학습 방식으로 그 이후 새로운 이미지가 입력되면 컴퓨터가 스스로 ‘고양이’인지를 판단

반지도학습 – 일부 데이터에만 레이블을 주고 학습하는 것

강화학습 – 경험을 통해 기대하는 보상이 최대가 되도록 하는 .. > 컴퓨터가 환경의 상태를 감지하여 행동을 하면 이에 대한 ‘보상’을 제공하면서 학습시키는 방식

+ 비지도학습 – lable을 주지 않고 학습시키는 방법으로, 컴퓨터가 데이터를 토대로 데이터의 특성을 파악하여 그룹으로 분류하는 방식으로 학습한다.

1. 전통적인 프로그래밍 방법과 인공지능 프로그램의 차이점은 무엇인가?

: 전통적인 프로그래밍은 직접 코드를 하나씩 다 짜는 것이고 인공지능 프로그램은 규칙을 생성해서 자동으로 코드를 짤 수 있다.

1. 딥러닝과 머신러닝의 차이점은 무엇인가?

: 딥러닝은 데이터 처리를 자체적으로 해서 예측을 하고, 머신러닝은 외부적으로 한다.

/ 딥러닝은 비정형 데이터, 머신러닝은 큰 데이터를 다룸

1. Classification과 Regression의 주된 차이점은?

: Classification은 이산적인 데이터를 처리하고 예측하는 것, Regression은 연속적인 데이터 중 하나를 예측하는 것. 또한 손실 함수의 측면에서도 차이가 있음

1. 머신러닝에서 차원의 저주(curse of dimensionality)란?

: 차원이 높아지면 학습의 성능이 떨어질 수 있음. 데이터가 부족하기(적어지기) 때문이다.

1. Dimensionality Reduction는 왜 필요한가?

: 차원의 저주를 방지하기 위해서 > 모델의 성능을 높임(데이터의 크기를 줄여서)

1. Ridge와 Lasso의 공통점과 차이점? (Regularization, 규제 , Scaling)

: 둘 다 과적합(Overfitting)을 방지하기 위해 사용하는 Regulation 기법이다. 오버피팅이란 모델에 대해 잘못된 데이터를 포함하여 (과한) 학습을 했을 때 발생하는 현상이다. Ridge는 가중치의 크기를 줄이는(0과 가깝지만 0이 되지 않음) 방식이며 모든 변수를 유지해야 하는 경우에 쓰인다. Lasso는 특정 가중치를 완전히 0으로 만드는 방식으로, 불필요한 변수를 아예 제거할 수 있으며 일부 변수를 제거하고 싶을 때 쓰인다.

1. Overfitting vs. Underfitting

: 오버피팅은 모델에 대해 너무 많이 알아서(잘못된 데이터까지 학습해서 노이즈가 낌) 데이터에 대한 일반화가 잘못된 현상이고, 언더피팅은 모델에 대해 너무 몰라서(=데이터가 부족함 또는 모델이 너무 단순함) 데이터를 제대로 학습하지 못하는 현상이다. 오버피팅의 경우 정규화를 하고, 언더피팅의 경우 더 복잡한 모델을 쓰거나 데이터를 늘려야 한다.

1. Feature Engineering과 Feature Selection의 차이점은?

: Feature Engineering은 기존의 데이터에서 의미있는 정보를 기반으로 새로운 Feature를 만드는 것

Feature Selection는 불필요한 Feature를 제거하고 중요한 피쳐만 남기는 것

1. 전처리(Preprocessing)의 목적과 방법? (노이즈, 이상치, 결측치)

: 전처리는 모델링의 정확도를 높이고 예측값의 오차를 줄이기 위해서 필요하다. 수집한 데이터에 결측치가 있는지 확인하고 결측치를 제거하고 이상치를 처리한다, 불필요한 데이터(노이즈)를 제거한다.

1. EDA(Explorary Data Analysis)란? 데이터의 특성 파악(분포, 상관관계)

: 데이터 분석 전, 데이터의 분포, 상관관계 등을 시각화하여 파악하는 과정

1. 회귀에서 절편과 기울기가 의미하는 바는? 딥러닝과 어떻게 연관되는가?

: ^​y=wx+b

절편(b)은 x=0일 때 y값이고, 기울기(w)는 변수 x가 증가할 때 y의 변화량이다.

딥러닝에서 기울기는 가중치(w)이고 절편은 편향(b)이다.

1. 교차검증, K-fold 교차검증의 의미와 차이

: 교차검증은 데이터를 훈련/테스트 세트로 나누어 모델 성능을 평가하는 것 K-fold 교차검증은 데이터를 K개의 부분으로 나누고 여러 번 학습하여 평균 성능 측정하는 것

1. 하이퍼파라미터 튜닝이란 무엇인가?

: 모델의 성능을 최적화하기 위해 학습률, 트리 개수 등 학습 과정의 설정값을 조정하는 과정.

1. 결정트리에서 불순도(Impurity) – 지니 계수(Gini Index)란 무엇인가?

: 결정트리는 데이터를 반복적으로 나누어 예측(Classification or Regression)하는 알고리즘이고 데이터를 나눌 때 어떤 기준으로 할 지가 중요한데 그 기준 중 하나가 불순도이다. 불순도는 데이터가 얼마나 섞여 있는지 나타내는 지표이고 지니 계수는 임의의 데이터를 뽑았을 때 잘못 분류될 확률이다.(0이면 불순도 0 ,최적의 분할 / 1이면 완전 섞임, 최악의 분할) > 지니 계수는 0에 가까울수록 좋음

1. 앙상블이란 무엇인가?

: 여러 개의 모델을 결합하여 성능을 향상시키는 기법이다

1. 부트 스트랩핑(bootstraping)이란 무엇인가?

: 데이터 샘플을 반복적으로 랜덤하게 선택하여 여러 개의 학습 데이터를 생성하는 방법이다

1. 배깅(Bagging)이란 무엇인가?

: 부트스트랩 샘플을 이용하여 여러 모델을 학습한 후 평균을 내는 방식

1. 주성분 분석(PCA) 이란 무엇인가?

: 고차원의 데이터를 저차원으로 축소하는 기법으로, 가장 중요한 특징(주성분)을 추출하여 데이터의 구조를 단순화한다