

1. 인공지능에서 2능에 해당하는 기능은 무엇인가?

2능 $\begin{cases} \text{classification (분류)} - \text{이산적인 값,} \\ \text{regression (회귀)} - \text{연속적인 값} \end{cases}$

EX) 결과값이 정수일때 관계성이 있으면 regression, 관계성이 없으면 classification.

2. 인공지능의 종류 3가지

- 지도학습: 정답을 알려주고 학습하는 방식

- 비지도학습: 일부분 데이터에 정답이, 나머지 데이터 정답 X, 학습하는 방식.

- 강화학습: 보상을 기반으로 스스로 학습하는 방식.

3. 전통적인 프로그래밍과 인공지능 프로그래밍 차이

전통 \rightarrow 사람이 입력, 규칙을 직접 책으로 정함.

2 - 인공지능 \rightarrow " 입력, 정답을 " 규칙을 API로 이용.

4. 딥러닝, 머신러닝 차이점.

- 딥러닝 \rightarrow 모델 안에서 Feature 추출.

\hookrightarrow 우리가 알 수 없음.

- 머신러닝 \rightarrow Feature 추출을 모델 밖에서.

5. Classification, regression 차이

- API가 다르고, 평가기준도 다르다.

$\begin{matrix} \swarrow & \searrow \\ \text{Accuracy} & \text{MSE} \\ \text{(classification)} & \text{(regression)} \end{matrix}$

6. 다변량에서 과잉의 리스크

- 과잉이 증가할 수록 (고차원공간) 데이터가 희박해지고, 모델 정확도가 떨어지는 현상.

7. 과잉축소가 필요한 이유

- 데이터 분식, 신호를 불의하게 하고, 노이즈를 제거하여 모델의 복잡도 ↓

8. Ridge와 Lasso

- 공통점:
- 정규화 기법 사용.
 - 특성 스케일링 필요.
 - 선형계 모델

- 차이점:
- Ridge → 모든 특성 사용, 계수를 작게 조절.
 - Lasso → 불필요한 특성의 계수를 0으로 만들어 특성 선택.

9. overfitting vs underfitting

- overfitting (과적합)

: 아웃리어 같은 노이즈가 데이터로 포함됨.

- underfitting (과소적합)

: 모델 정확도가 충분히 못되어 발생하는 현상.

원인 → 데이터 부족, 모델이 단순.

해결책: 데이터 증대, 복잡한 모델 사용.

10. Feature Engineering과 Feature Selection 차이

- Feature selection: Features의 상관관계를 통해서 중요도를 평가하고 불필요한 Feature 만 선택.
- Feature Engineering: 기존 데이터를 변형 or 새로운 Feature 생성.

11. 전처리 (Preprocessing) 목적, 방법.

목적: 정제되지 않고 많은 데이터보다 잘 정제된 작은 데이터가 편하게 더 사용하기 때문에 데이터를 정제하려는 목적

방법: 결측치 처리, 스케일링, 아웃리어 처리.

Date.

No.

12. EDA (Exploratory Data Analysis)란?

↳ 탐색적 데이터 분석 : 패턴, 선형적 관계의 존재, 그래프, 요약, 통계 등을 이용하여 통계학을 수행하는 방법.

데이터 분포 분석 → 이상치, 정규성, 분포 문제 해결 가능.

상관관계 분석 → 불필요한 변수 제거, 타겟 변수

13. 회귀에서 결과와 기술기 의미 → 결과값과 어떤 연계를

회귀 → 결과 → 원인이 어떤 것인지 기밀성.

↳ 기술기 → 입력 데이터가 변화할 때 결과가 얼마나 변하는지.

결과값 → 결과 → 방향은 의미.

↳ 기술기 → 계수화하여 연관.

14. 교차검증, k-fold 교차검증이 의미와 차이

교차검증 (cross-validation) : 모델의 성능을 평가하는 방법 중 하나.

데이터를 train, test 2로 나누어 여러번 반복하여

모델을 평가하는 방법.

k-Fold 교차검증 : 데이터를 k로 나누고, 각 k를 한번씩 검증용으로 사용하여 k번 반복해서 모델을 평가하는 방법.

차이점 : 교차검증 — 편향성 ↓, 많은 데이터의 검증에 활용, 평가 신뢰성 ↑
k-Fold 교차검증 — 편향성 ↑, 전체 "를 검증에 활용, 평가 신뢰성 ↑

15. 하이퍼파라미터 튜닝이란

하이퍼 파라미터 튜닝 : 머신러닝 모델이 잘 작동하도록 사용자가 직접 조절해야 하는 매개변수를 찾는 과정.

↳ 모델이 학습하면서 자동으로 찾는 것

하이퍼 파라미터 : 학습 자체 사용자에게 직접 설정해야 하는 값

16. 결정트리에서 불순도 - 지니지수란?

- 결정트리 → 불순도 (impurity) : 데이터를 어떻게 나누면 가장 깔끔하게 분류할 수 있는지에 평가 지표.
 ↳ 지니지수 (Gini index) : 결정트리가 가장 좋은 불순도 측정도구로 사용되는 측정도구.
 노드의 불순도를 측정하는 방법.

17. 앙상블이란?

! 여러개의 개별 모델을 결합하여 더욱 강력한 예측 결과를 얻는 방법.

18. 부트스트랩이란?

- ! 주어진 데이터에서 복원추출을 통해 새로운 데이터셋을 생성하는 방법.
 ↳ 뽑은 데이터를 다시 원본 데이터에 넣고 반환해서 추출.

19. 배깅이란?

! 여러개의 모델을 훈련시켜 평균을 내거나 투표해서 최종결정을 내리는 방법.

20. 랜덤 포레스트란?

- ! 많은 데이터를 더 작은 클러스터로 학습하는 방법.
 ⇒ 데이터의 랜덤한 특징만 뽑아서 모델을 학습하는 방법.
 ⇒ 데이터를 임의하게도 랜덤한 특징은 유지할 수 있는 방법.