서술 숙제

1. Library와 Framework의 차이 간단하게 서술 (100자 내외)

라이브러리와 프레임워크의 차이는 2가지 측면에서 발견할 수 있습니다. 첫번째, 포함 관계로 두 대상을 살펴볼 때, 프레임워크가 라이브러리를 포함함을 알 수 있습니다. 또한 두번째, 개발자의 선택 측면에서도 명백한 차이를 보입니다. 개발자는 해당 프레임워크의 규칙과 구조를 따라야 합니다. 즉, 선택권이 없습니다. 반면에 개발자는 필요에 따라 라이브러리의 함수를 선택하여 호출할 수 있습니다.

1. 딥러닝과 머신러닝의 관계 및 특징, 차이 서술 (200자 내외)

머신러닝과 딥러닝의 특징 및 차이를 이해하기 앞서 딥러닝이 머신러닝에 해당한다는 점을 알아야 합니다. 머신러닝과 딥러닝의 개념을 살펴보면 각각의 특징과 차이점을 발견할 수 있습니다. 머신러닝은 문자 그대로 기계나 컴퓨터가 주어진 데이터를 학습하여 미래를 예측하는 기술로 개략적으로 정의할 수 있으며, 딥러닝은 인공신경망을 머신러닝에 적용한 기술을 의미합니다. 즉, 실질적으로 딥러닝은 머신러닝의 하위개념이므로 두 개념이 비슷한 방식으로 기능함은 명백합니다. 다만 딥러닝의 경우 머신러닝보다 요구되는 데이터 수가 압도적으로 많으며 이에 따라 학습시간도 훨씬 오래 걸리고 비용 역시 고가입니다. 그러나 최근 폭발적으로 증가하는 데이터와 이를 둘러싼 변수들로 인해 딥러닝에 대한 연구가 매우 활발하게 진행되고 있습니다.

1. 주석달기
2. RandomForest example

* Regressor 사용 (y label이 연속형 수치이므로 regressor 사용)
* sklearn에서 유명한 알고리즘인 random forest를 사용한 예시.

: ensemble, 즉 다수결을 이용한 알고리즘으로 bagging에 해당한다.

: 배깅은 의사결정나무가 서로 독립적

1. XGBoost regressor 사용

: ensemble은 크게 bagging과 boosting으로 이루어지는데 XGBoost는 부스팅에 해당한다.

: 초반의 의사결정나무에서 발생한 에러를 줄이는 방향으로 추후의 의사결정 나무가 생성된다.

* 알고 싶어하는 값(y)가 정해져 있을 때 사용하는 지도학습 적용.

1. DCGAN 예시 -> 딥러닝으로 비지도 학습에 해당합니다. 비지도 학습은 주로 MAIN 지도 학습 전, 더 나 결과 혹은 시각적인 효과를 얻기 위해 주로 사용합니다. 또한 Y\_label, target label을 이용하지 않는 알고리즘입니다. 비지도 학습은.
2. EDA 숙제

파이썬을 이용하여 전처리 및 EDA를진행해주세요.

1) 결측치, 이상치 검토

\* isnull(), isna 보기

\* 평균과 표준편차 수치, 데이터 표준화, 박스플롯

\* 결측치 = 제거, 평균대체, 중앙값 대체, 최빈값 대체, 선형보간법

\* 이상치 = 표준점수로 변환 후 제거, 수치형 변수 범주화

2) 유의미한시각화 5개 이상

\* 먼저 HEAD(), Tail(), infor(), describe()로 기본정보 체크

\* 박스플롯, 히스토그램, 모자이크 플롯(범주-범주), side-by-side box plot(수치-범주) 등

3) 수치형 변수 간 상관관계 파악

\* 산점도(scatter plot), 선형 관계(corr plot) 등 체크

4) 파생변수생성