

1. 학습, 추론, 문제 해결, 자동화 언어 이해, 이미지 인식 등을 지능의 해당하는 기능이고 머신러닝 및 딥러닝 기술을 통해 데이터에서 패턴 학습과 이를 기반으로 의사결정 내릴 수 있는 능력이 주요 기능이다.

2. 지도학습: 정답이 있는 데이터를 활용해 데이터를 학습시키는 것을 의미하고
종류는 분류, 회귀가 있다.

분류는 주어진 데이터를 정해진 카테고리에 따라 분류하는 것을 의미
회귀는 어떤 데이터들의 Feature를 기준으로, 연속된 값을 예측하는 문제로
주로 어떤 패턴이나 트렌드 등을 예측할 때 사용

비지도 학습: 정답이 없는 데이터를 비슷한 특징까지 군집화하여 새로운 데이터에 대한
결과를 예측하는 방법을 의미하고 지도학습보다 난이도가 높다 (ex. GAN)

강화학습: 분류할 수 있는 데이터가 존재하는 것도 아니고 데이터가 있어도
정답이 따로 정해져 있지 않으며 자신이 한 행동에 대해 보상을
받으며 학습하는 것을 의미 (ex. DQN, ABC)

3. 전통적인 프로그래밍은 개발자가 직접 코드를 작성하여 컴퓨터에 전달한다.
반면 인공지능 프로그래밍은 머신러닝 및 딥러닝과 같은 기술을 활용하여
모델이 데이터에서 패턴을 학습하고 예측한다
이점으로 데이터 기반 학습과 자동적인 패턴 인식을 강조하는 차이점을 알 수 있다

4. 머신러닝은 데이터를 이용하여 사용자의 특성을 학습하여 예측하는 기술을 의미하고
딥러닝은 사람의 뇌를 모방하여 데이터를 직접 학습하고 문제해결까지
처리하는 차이가 있다.

5. 둘 다 지도학습의 종류로 비슷한 개념이지만 Classification은 예측값으로
이산값을 출력하고 Regression은 예측값으로 연속적인 값을 출력하는
서로 다른 종류의 값을 출력한다는 차이점이 있다.

6. 공간의 차원이 증가함에 따라 데이터의 밀도가 급격히 감소하고 이로 인해
데이터 분석이나 머신러닝 모델의 성능에 부정적인 영향을 미치는 현상을
의미한다. 이로 인해 연산 비용 증가, 과적합, 예측 성능 저하가 발생
할 수 있다.

7. 차원 축소란 어떤 목적에 따라서 데이터 양을 줄이는 여러가지 방법을 말한다.
필요에 맞는 데이터 양을 줄이면 시간 복잡도와 공간 복잡도가 줄어들어
더 적은 자원으로 같은 목적에 달성 할 수 있고 오히려 데이터의 차원을 줄여서
학습 시키면 모델이 간단해지고, 적은 데이터로 안정적인 결과를 내놓을 수 있다.
마지막으로 모델이 간단 할수록 내부구조 이해가 편하고 사람이 결과를 알아보기
편하다.

8. 교차검정: 둘 다 모델의 복잡성을 줄이기 위해 정규화를 사용하고
모델의 가중치를 축소하여 과적합을 방지하고 모델의 일반화 성능을 향상 시킨다.
차이점: Ridge는 L2 정규화를 사용하고, Lasso는 L1 정규화를 사용한다.
Ridge는 모든 특성이 어느정도 영향을 미치게 하지만
Lasso는 특정 특성 모델에 영향을 미침

9. Overfitting은 모델이 학습 데이터에 과도하게 적응하여 훈련 데이터에
존재하는 문제까지 학습하는 현상을 의미하며 훈련 데이터는 높은 정확도를
보이지만, 새로운 데이터에 대해서는 일반화 성능이 낮아질 수 있음
해결 방법으로는 데이터를 더 수집하거나, 모델을 간단하게 만들거나
정규화 기법을 사용하여 모델의 복잡성을 줄이는 방법 사용
Underfitting은 모델이 학습 데이터에 대해 적절히 학습하지 못하고
훈련 데이터 나 새로운 데이터에 대한 성능이 낮은 현상을 의미함
모델이 너무 간단하거나 데이터에 중요한 패턴을 학습하지 못함
해결 방법으로는 모델의 복잡성을 늘리거나, 더 많은 특성을 추가하는 방법을
사용함

10. Feature Engineering은 원본 데이터를 변형하여 모델의 성능을 향상 시키는 것이
목적이고, Feature Selection은 주어진 특징 중에서 중요한 특징만을
선택하여 모델을 구축하는 것이 목적이다.
Feature Engineering은 데이터의 특성을 변형하거나 새로운 특성을 생성하는 작업
Feature Selection은 주어진 특징 중에서 어떤 것을 선택할지 결정하는 작업

11. 전처리의 목적은 데이터 품질 향상, 모델 성능 향상, 적절한 입력 형태로 모델이 이해하는 쉬운 형태로 데이터를 변환하여 입력데이터를 최적화 시키는 것입니다.

방법

노이즈 처리: 평균화, 이동평균, 이동중앙값 등을 사용하여 노이즈를 제거하는 방식

이상치 처리: 통계적인 방법, 시각적 탐지, 클러스터링을 활용하여 이상치를 식별하고, 필요에 따라 제거하거나 대체

결측치 처리: 삭제, 대체, 예측 모델을 활용하여 결측치를 추정하여 처리

12. EDA는 주어진 데이터를 탐색하고 이해하기 위해 다양한 통계적, 시각적 방법을 사용하는 과정

데이터의 분포, 패턴, 특성을 파악하여 가설을 세우고 모델링에 도움을 준다
변수 간의 상관 관계를 조사하여 데이터 간의 패턴을 이해하고 모델링에 활용

13. 회귀에서 절편은 모델의 편향을 나타내고 기울기는 모델의 가중치를 의미합니다.

단리방에서는 모델의 파라미터로 존재한다. 신경망에서는 가중치가 기울기, 편향이 절편에 해당한다.

단리방은 다중선형방을 사용하고, 각층의 가중치와 편향을 조절하여 입력과 출력 사이의 관계를 학습한다.

14. Activation Function을 사용하는 이유는

신경망의 층을 쌓을 때 비선형성을 도입하여 모델이 더 복잡한 패턴을 학습할 수 있게 해주고, 입력력 간의 복잡한 비선형 관계를 모델링할 수 있도록 만들어주며 출력 범위를 제한하여, 신경망의 안정성과 수렴속도를 향상 시킨다.

Softmax는 다중 클래스 분류에 사용되며, 클래스 간 상대적인 확률을 나타낸다.

Sigmoid는 이진 분류에 사용되며, 해당 클래스에 속할 확률을 나타낸다.

15. Forward Propagation는 신경망이 입력데이터를 받아 출력까지의 값을 계산
입력에서 출력까지의 값 계산을 통해 모델이 주어진 데이터에 대한 예측을 수행
Backward Propagation은 신경망이 학습 데이터를 기반으로 가중치를 업데이트하기
위한 기울기를 계산하는 과정

모델이 학습 데이터에 대한 오차를 최소화하기 위해 가중치를 조정

16. 손실함수란 모델의 예측 값과 실제 값 사이의 차이를 측정하는 함수로,

학습 중에 모델을 조정하기 위한 기준으로 사용

4가지 종류

평균 제곱 오차: 회귀 문제에서 주로 사용되며 예측 값과 실제 값 사이의 제곱 오차를 계산
교차 엔트로피 손실: 이진 분류 및 다중 클래스 분류 문제에서 주로 사용

훈련 속도 증가 손실: 회귀 문제에서 이상치에 민감하지 않으면서 MSE보다는 덜
민감한 손실 함수를 사용할 때 활용

훈련 속도 감소 손실: 서포트 벡터 머신과 관련이 있으며, 이진 분류 문제에서 사용

17. 옵티마이저란 모델의 가중치 및 편향을 조정하여 손실 함수를 최소화하는 최적의 모델
파라미터를 찾는 알고리즘이다.

손실 함수의 기울기를 계산하고 이를 사용하여 가중치를 업데이트하여 모델을 향상
차이

손실 함수는 모델의 성능을 측정하고, 옵티마이저는 손실을 최소화하기 위해 모델의
파라미터를 조정.

18. 경사 하강법은 손실 함수의 기울기를 사용하여 손실의 최소화하는 방향을 모델 파라미터
에 업데이트하는 최적의 알고리즘

확률적 경사 하강법: 각 훈련 데이터 포인트에 대해 경사를 계산하고 파라미터를 업데이트

배치 경사 하강법: 전체 훈련 데이터 세트에 대해 경사를 계산하고 파라미터를 업데이트

미니 배치 경사 하강법: 훈련 데이터를 작은 미니 배치로 나누어 각 미니 배치에
대해 경사를 계산하고 파라미터를 업데이트

19. 교차 검증: 모델의 성능을 신뢰할 수 있게 평가하기 위한 기법으로

데이터를 여러 부분집합으로 나누어 모델을 여러 번 훈련하고 평가

k-fold 교차 검증: 데이터를 k개의 폴드로 나누고, 각 폴드에 대해 모델을 k번 훈련하고
평가

차이 교차 검증은 단일 분할로 인해 훈련 데이터와 검증 데이터의 성능의 영향을 받을 수 있고

k-fold 교차 검증은 데이터가 작을 때 유용하며, 모델이 어떻게 일반화 될지에 더 신뢰할 수 있는 정보를
제공

20. 머신러닝 모델의 학습 과정에서 사용되는 하이퍼 파라미터들을 최적화하는 과정

방식은 하이퍼 파라미터 공간 정의 후 평가 기준을 선택하고 하이퍼 파라미터
조합 탐색 후 최적의 하이퍼 파라미터를 선택한다.

모델의 성능을 극대화하거나 일반화 성능을 향상시키기 위해 적절한 하이퍼 파라미터
설정을 찾아내는 것이 목표