

2025학년도 1학기

인공지능 기초를 위한 FAQ



과목명	인공지능개론
담당	이명규 교수님
제출일	2025년 3월 22일
학과	컴퓨터공학과
학번	202136017
성명	이준상

1. 인공지능에서는 인간의 인지 능력을 모방하거나 뛰어넘는 학습 능력, 문제 해결 능력, 추론 및 인지 능력 등 다양한 기능을 통해 기능을 구현한다.
2. 지도학습은 정답이 정해진 데이터를 사용하여 모델을 학습시키는 방식이다. 모델이 패턴을 학습하고 새로운 입력에 대한 예측을 수행한다. 비지도학습은 정답이 정해진 데이터와 양을 데이터를 함께 사용하여 모델을 학습시키는 방식이다. 정답이 정해진 데이터로 기본적인 학습하고 정해진 양을 데이터는 데이터의 특성을 파악하고 모델은 예측 결과로 출력한다. 결과학습은 인간이 목표를 달성하기 위해 사용하는 시행착오 학습 과정으로 모방한다. 각 작업에 피드백을 통해 배우고 최종 결과를 달성하기 위한 최상의 처리 경로는 스스로 학습한다.
3. 전통적인 프로그래밍은 프로그래머가 문제 해결을 위한 모든 규칙과 방법을 코드로 작성하지만 인공지능 프로그래밍은 데이터로부터 패턴을 학습하고 스스로 규칙을 생성한다.
4. 머신러닝은 인공지능의 하위 분야로 사람이 데이터에서 통찰을 추출하여 이를 바탕으로 패턴을 학습한다. 비교적 적은 양의 데이터로도 학습이 가능하고 일반적인 컴퓨팅 환경에서도 작동한다. 딥러닝은 머신러닝의 하위 분야로 신경망을 이용하여 데이터로부터 자동으로 통찰을 추출하고 학습한다. 대량의 데이터를 필요로 하고 고성능 컴퓨팅 환경은 필요하다.
5. Classification과 Regression의 주요 차이점은 예측하려는 값의 형태에 따라 구분된다. Classification은 범주형 값으로 예측한다. 데이터가 어떤 범주에 속하는지를 예측한다. Regression은 연속적인 값을 예측한다. 데이터의 수치적인 값을 예측한다.
6. 차원 저하는 데이터의 차원(특성 수)가 증가함에 따라 발생하는 모델의 성능 저하, 학습 시간 증가, 데이터 분석의 어려움 등의 문제들을 해결하는 용어이다. 차원 축소, 데이터 증강, 적절한 모델 선택 등의 방법으로 차원 저하를 해결할 수 있다.
7. 차원 축소는 차원 저하를 해결하고 모델의 복잡도를 줄여 과적합을 방지할 수 있다. 계산 복잡도를 줄여 학습 시간을 단축하고 메모리 사용량을 줄인다. 불필요한 특성을 제거하고 중요한 특성이 집중이 가능하다. 차원 축소는 머신러닝 모델의 성능을 향상시키고 데이터 분석을 용이하게 해준다. 데이터의 특성과 분석 목적에 따라 적절히 차원 축소를 하는 것이 중요하다.
8. Ridge와 Lasso의 공통점은 선형회귀 모델에 규제 항을 추가하여 모델의 복잡도를 제어한다. 규제를 통해 모델이 학습 데이터에 과적합을 방지하고 새로운 데이터에 대한 예측 성능을 향상시킨다. 모두 일정한 변수의 수에 의해 인감한다. 차이점은 Ridge는 가중치의 제곱항을 규제 항으로 사용하는 L2 규제를 사용한다. Lasso는 가중치의 절대값항을 규제 항으로 사용한다. Ridge는 가중치를 0이 안되고 0에 가깝게 만들지만 Lasso는 일부 가중치를 0으로 만들어 특정 특성을 모델에서 제거하는 특성 선택 효과를 가진다.
9. Overfitting은 모델이 학습 데이터를 너무 과도하게 학습하여 노이즈나 불필요한 세부 사항까지 학습하는 현상이다. 모델이 너무 복잡하거나 불필요한 특성이 학습 데이터에 포함된 등의 요인으로 과적합은 발생하며 모델은 더 단순하게 하거나 규제 및 교차 검증을 사용하는 등의 방법으로 해결 가능하다. Underfitting은 모델이 너무 단순하여 데이터의 기본적인 패턴조차 학습하지 못하는 현상이다. 데이터의 특성이 부족하거나 모델이 데이터 복잡성이 너무 단순하다는 등의 이유로 발생한다. 더 복잡한 모델은 사용하거나 특성을 추가하는 등의 해법으로 해결 가능하다.
10. Feature Engineering과 Feature Selection의 차이점은 목적과 방법이 차이가 있다. 특성 인지의 깊은 기존 특성을 기반으로 새로운 특성을 생성하거나 변환하여 모델의 예측력을 높이는 데 목적이 있다. 특성 선택은 모델 학습이 불필요하거나 성능을 저하시키는 특성을 제거하고 중요한 특성을 선택하여 모델의 효율성과 일반화 성능을 향상시키는 데 목적이 있다.
11. 전처리된 머신러닝 모델의 성능을 향상시키기 위해 데이터를 분석 및 모델링이 적합하도록 정리하고 변환하는 과정이다. 전처리는 데이터의 품질을 향상시키거나 모델의 성능 향상, 학습 효율성 증대 등이 목적으로 두고 있다. 노이즈 제거, 불필요한 특성 제거, 결측치 처리, 스케일링 등의 기법으로 노이즈를 제거하거나 원하지 않는 이상치를 제거 또는 대체하는 등의 방법이 있다. 이상치 전체가 범위는 이상치를 포함하는 데이터를 제거하거나 대체하는 방법과 관련하여 영향을 줄이는 방법이 있다. 결국 전처리는 제거, 대체, 값은 유지하고 결측치를 제거할 수 있는 여러 방법은 사용한다.

12. EDA는 데이터 분석의 초기 단계에서 데이터를 이해하고 주요 특징을 파악하기 위해 수행하는 과정이다. 전반적인 구조와 특성을 파악하고 품질을 평가하여 오류나 이상치를 발견한다. 데이터의 분포 파악 방법에는 평균, 중앙값 등을 계산하여 데이터의 중심 경향과 산포도를 파악하거나 히스토그램, 박스 플롯과 같은 시각화준 용하여 파악이 가능하다. 변수 간 상관관계는 선형, 비선형 상관관계 여부를 파악가능한 산점도나 모든 변수 간의 상관 계수를 계산하여 이후 시각적으로 표현하는 등의 방법으로 파악이 가능하다.

13. 회귀에서 절편은 독립 변수가 아닌 때에러항 값으로 모델의 기본값이라고 해석할 수 있다. 회귀식에서 절편은 상수항으로 표현되며, 모델이 전체적인 수준을 조정하는 역할을 한다. 기울기는 독립변수가 증가함에 따른 변화량이다. 기울기의 절댓값이 클수록 독립변수와 종속변수 간의 관계가 강하다는 것을 의미한다. 부호는 관계의 방향을 나타낸다. 여러 결과에 대한 답변의 가장 기본적인 가정 모인 뉴턴은 선형 회귀와 유사한 방식으로 작동한다. 각 뉴턴은 기울기와 편향을 가진다. 입력 값이 증가할수록 급하고 편향은 더하여 출력값을 생성하는 답변에서 가중치는 회귀에서 기울기와 유사하고 편향은 절편과 유사하다는 것은 안다.

14. 교차검증은 모델의 성능을 평가하기 위해 데이터를 여러 개의 외부 집합으로 나눈다. 이 외부 집합들을 반복적으로 학습과 검증에 사용하는 방법이다. 교차검증은 양지하고 모델의 일반화 성능을 평가할 수 있다. K-이브 교차검증은 교차검증의 한 종류로, 데이터를 K개 그룹으로 크기의 부분 집합으로 나눈다. K번 반복하여 각 집합을 학습 및 검증에 사용하고, 나머지 K-1개의 집합을 학습에 사용한다. 최종적인 K번의 학습 결과를 평균하여 모델의 성능을 평가한다. K-이브 교차검증의 한 종류이며 더 구체적인 절차로 가진다 할 수 있다.

15. 하이퍼파라미터 튜닝은 머신러닝 모델의 성능을 최적화하기 위해 모든 학습 전에 사용자가 직접 설정하는 파라미터인 하이퍼파라미터의 최적값을 찾는 과정이다.

16. 결정트리에서 불순도는 데이터 집합 내에서 서로 다른 종류의 데이터들이 얼마나 섞여있는지를 나타내는 지표이다. 지니 계수는 결정 트리에서 데이터의 불순도를 측정하고, 최적의 분할 기준을 찾는 데 중요한 역할을 한다. 지니 계수는 데이터 집합에서 임의로 선택된 두 개의 데이터가 서로 다른 클래스에 속할 확률을 나타낸다.

17. 앙상블은 서로 다른 알고리즘을 사용하거나 동일한 알고리즘이라도 다른 데이터셋으로 학습된 모델들의 예측 결과를 결합하여 최종 예측을 한다. 여러 개의 머신러닝 모델을 결합하여 단일 모델보다 더 나은 예측 성능을 얻는 기법이다.

18. 부트스트래핑은 표본 데이터로부터 모델의 특성을 추정하거나 모델의 성능을 평가하는데 사용되는 재표본추출 기법이다. 주어진 데이터셋에서 복원추출을 통해 여러 개의 새로운 데이터셋을 생성한다.

19. 배깅은 원래 학습 데이터셋에서 복원추출을 함으로써 여러 개의 독립 데이터셋을 생성하고, 각 데이터셋으로 모델을 학습한다. 학습된 모델들의 예측 결과를 투표 또는 평균을 계산하여 최종 예측을 수행한다.

20. 랜덤 포레스트는 고차원 데이터에서 불확실성을 줄이고 중요한 정보를 유지하는 데 사용되는 다중 의사결정 기법이다. 여러 변수 간의 상관관계를 분석하고 데이터의 변수를 최대의 보충하는 새로운 축(주성분)을 찾아내는 방법이다.