

## 1. 인공지능에서 지능에 해당하는 기능은 무엇인가?

: 인공지능에서 지능은 문제해결, 학습, 추론, 인식, 자연어 처리, 의사결정 등의 기능을 포함한다. 이는 인간의 인지 능력을 모방하여 데이터를 학습하고, 이를 기반으로 예측하거나 패턴을 인식하는 능력을 의미한다.

## 2. 인공지능의 종류 3가지를 다뤄서 설명해라. (지도학습, 반지도학습, 강화학습)

: (1) 지도학습 : Supervised Learning. 정답(Label)이 있는 데이터를 학습하여 새로운 데이터를 예측하는 방식이다.

예.) 이메일 스팸 분류, 손글씨 숫자 인식, 주택 가격 예측

대표적인 알고리즘 - 선형 회귀, 로지스틱 회귀, 의사결정 트리, SVM, 랜덤 포레스트 등

(2) 반지도 학습 : Semi-Supervised Learning. 일부 데이터만 정답(Label)이 주어진 상태에서 학습하는 방식이다. 적은 양의 라벨링된 데이터와 많은 양의 비라벨링 데이터를

함께 학습하여 모델의 성능을 향상시킨다.

예.) 대량의 이미지 데이터에서 일부만 라벨링된 데이터를 이용한 분류

(3) 강화학습 : Reinforcement Learning. 보상(Reward)과 벌점(Penalty)을 기반으로 최적의 행동을 학습하는 방식이다. 환경과 상호작용하며 목표 달성을 위한 최적의 정책(Policy)을 학습한다.

예.) 양파고, 자율주행자동차, 보트 제어

## 3. 전통적인 프로그래밍 방법과 인공지능 프로그래밍의 차이점은 무엇인가?

: (1) 전통적인 프로그래밍 : 사람이 직접 명시적인 규칙을 정의하고 프로그램을 작성한다.

(2) 인공지능 프로그래밍 : 데이터로부터 패턴을 학습하여 규칙을 스스로 학습한다.

예.) 머신러닝 기반의 영문인식 시스템

## 4. 딥러닝과 머신러닝의 차이점은 무엇인가?

(1) 머신러닝 : Machine Learning. 데이터에서 패턴을 학습하는 알고리즘을 포함하는 개념으로, 대표적인 개념에는 SVM, 의사결정트리, 랜덤 포레스트 등이 있다.

(2) 딥러닝 : Deep Learning. 머신러닝의 한 분야로, 다층 신경망(Deep Neural Network)을 사용하여 더 깊은 데이터를 학습하는 방식이다.

딥러닝은 대량의 데이터와 높은 연산 능력이 필요하며, 특징(Feature)을 자동으로 추출하는 것이 장점이다.

## 5. Classification과 Regression의 구분 차이점은?

(1) Classification (분류) : 데이터를 특정 범주 (Label)로 분류하는 문제

예.) 스팸 이메일 분류, 손글씨 인식

(2) Regression (회귀) : 연속적인 숫자 값을 예측하는 문제

예.) 주택 가격 예측, 주가 예측

6. 머신러닝에서 차원의 저주 (Curse of dimensionality)란?

: 데이터의 차원이 증가할수록 학습이 어려워지는 현상이다. 차원이 높아질수록 데이터가 희소해지고, 계산량이 급격히 증가하여, 모델의 성능이 저하될 수 있다.

7. Dimensionality Reduction는 왜 필요한가?

: 차원의 저주를 완화하여 모델의 성능을 향상시킨다. 연산량을 줄이고 학습 속도를 향상시킨다. 데이터의 노이즈를 제거하고 핵심 기능을 추출한다.

Dimensionality Reduction의 대표적인 방법으로는 PCA(주성분 분석), LDA(선형 판별 분석), t-SNE이 있다.

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

: (1) 공통점 : 과적합(Overfitting)을 방지하기 위해 Regularization(규제)을 적용하는 선형회귀모델이다.

(2) 차이점 : Ridge Regression - L2 규제. 가중치(weight)의 제곱합을 패널티로 적용하여 가중치를 작게 만든다. 모든 특징을 일부 반영한다.

Lasso Regression - L1 규제. 가중치의 절댓값 합을 패널티로 적용하여 불필요한 특징의 가중치를 0으로 만든다.

9. Overfitting vs Underfitting

: (1) Overfitting : 과적합. 훈련데이터에 너무 맞춰져 일반화 성능이 낮아지는 현상이다. 해결 방법으로는 Regularization, 엘리트 증강, Dropout이 있다.

(2) Underfitting : 과소적합. 모델이 너무 단순하여 데이터의 패턴을 제대로 학습하지 못하는 현상이다. 해결방법으로는 더 복잡한 모델을 사용하거나, 충분한 학습을 진행하여야 한다.

10. Feature Engineering과 Feature Selection의 차이점은?

: (1) Feature Engineering : 새로운 특징을 생성하여 모델의 성능을 높이는 과정이다.

(2) Feature Selection : 기존 특징 중 불필요한 것을 제거하여 모델 성능을 개선하는 과정이다.

11. 전처리 (Preprocessing)의 목적과 방법?

: (1) 목적 : 데이터 품질을 개선하여 모델의 성능을 향상시키는 것이다.

(2) 방법 : 노이즈 제거 - 데이터에서 불필요한 또는 잘못된 값을 제거하여 모델의 성능을 높이는 과정이다.

이상치 처리 - 다른 데이터와 크게 차이나는 값을 찾아서 제거하거나 수정하는 과정이다.

결측치 처리 - 누락된 데이터를 보완하는 과정으로, 평균/중앙값 대체 또는 여러 모델을 활용하여 채우는 방법이 있다.

12. EDA (Exploratory Data Analysis)란? 데이터의 특징 파악 (분포, 상관관계)

: ~~데이터~~ 데이터 분석으로, 데이터의 분포, 패턴, 상관관계를 파악하는 과정이다.

e.g.) 히스토그램, 상한 행렬, 박스 플롯 등을 활용한 데이터의 특징 분석

19. 선형회귀에서 절편과 기울기가 의미하는 바는? 절편과 어떻게 연관되는가?

: 절편 (Intercept)이란 그래프가 y축과 만나는 값이며, 기울기 (Slope)는 입력값이 변할 때 출력값의 변화율이다.

절편에서는 가중치 (Weight)와 편향 (Bias)이 절편과 기울기에 해당하며, 역전파 (Backpropagation)를 통해 최적의 값을 학습한다.

28. 불순도 (Impurity) - 지니계수 (Gini Index)란 무엇인가?

: 데이터가 얼마나 혼합되어 있는지를 측정하는 지표이다. 값이 작을수록 데이터가 한 클래스에 집중되어 있어 불순도가 낮다.

\* 지니계수 =  $1 - \sum (P_i^2)$  (각 클래스의 확률  $P_i$  사용)

29. 앙상블이란 무엇인가?

: 여러개의 모델을 결합하여 성능을 향상시키는 방법이다. 대표적인 기법으로는 배깅 (Bagging)과 부스팅 (Boosting)이 있다.

30. 부트스트래핑 (bootstrapping)이란 무엇인가?

: 유한 데이터에서 복원 추출을 통해 여러 개의 샘플을 생성하는 기법이다. 배깅 (Bagging)과 랜덤 포레스트 (RF)에서 활용된다.

31. 배깅 (Bagging)이란 무엇인가?

: 부트스트래핑 샘플링을 이용해 여러 개의 모델을 훈련하고 평균을 내어 예측하는 기법이다. 대표적인 예로는 랜덤 포레스트가 있다.

32. 주성분 분석 (PCA)이란 무엇인가?

: 다량의 데이터를 저차원으로 변환하여 계산 효율성과 시각화를 용이하게 하는 방법이다. 데이터의 분산을 최대화한 윗차원에서 축을 변환하여 주요 성분 (Principal Component)을 추출한다.