

# 인공지능 기초를 위한 FAQ

## 1. 인공지능에서 지능에 해당하는 기능은 무엇인가

- 학습 : 데이터를 통해 패턴을 학습하고 경험을 축적하는 능력
- 추론 및 논리적 사고 : 기존 지식을 바탕으로 새로운 결론을 도출하는 능력
- 문제 해결 능력 : 복잡한 문제를 해결하기 위해 논리적 방법을 적용하는 능력
- 지식 표현 : 정보를 구조화하여 저장하고 활용하는 능력
- 패턴 인식 : 이미지, 음성, 텍스트 등의 패턴을 분석하고 분류하는 능력
- 자연어 이해 및 생성 : 인간의 언어를 이해하고 생성하는 능력
- 계획 및 의사결정 : 특정 목표를達成하기 위한 희선의 행동을 결정하는 능력
- 창의성 : 새로운 콘텐츠를 생성하는 능력

## 2. 인공지능의 종류 3가지에 대해 설명하시오(지도학습, 반지도학습, 강화학습)

- 지도학습 : 레이블이 주어진 상태에서 컴퓨터를 학습시키는 방법

대표적인 예는 본류로 솔씨 인식 모델을 만드는 경우 손으로 쓴 숫자에 해당하는 클래스에 레이블을 붙이고 머신이 각 레이블을 학습한다.

알고리즘으로는 K-최근접 이웃, 선형 회귀, 로지스틱 회귀 등이 있다.

- 반지도학습 : 출력값이 없는 입력데이터만으로 학습을 진행하여 입력데이터의 패턴을 찾는 방법

대표적으로 뉴스기사와 일부 가사에만 주제 레이블이 있는 데이터셋을 학습시켜 새로운 뉴스 기사가 주제를 예측하게 한다.

알고리즘은 자기 학습, 일관성 정규화, 그래프 기반 알고리즘이 있다.

- 강화학습 : 에이전트가 주어진 환경에 대해 어떤 행동을 취하고 이로부터 어떤 보상을 얻며 학습하는 방법

대표적인 예는 알고리가 강화학습의 사례이며 게임에서 이길 시 양의 보상을 얻고 질 경우 음의 보상을 얻는다.

알고리즘으로는 Q-Learning 이 있고 Deep-Q-Network (DQN) 이 있다.

## 3. 전통적인 프로그래밍과 인공지능 프로그래밍의 차이

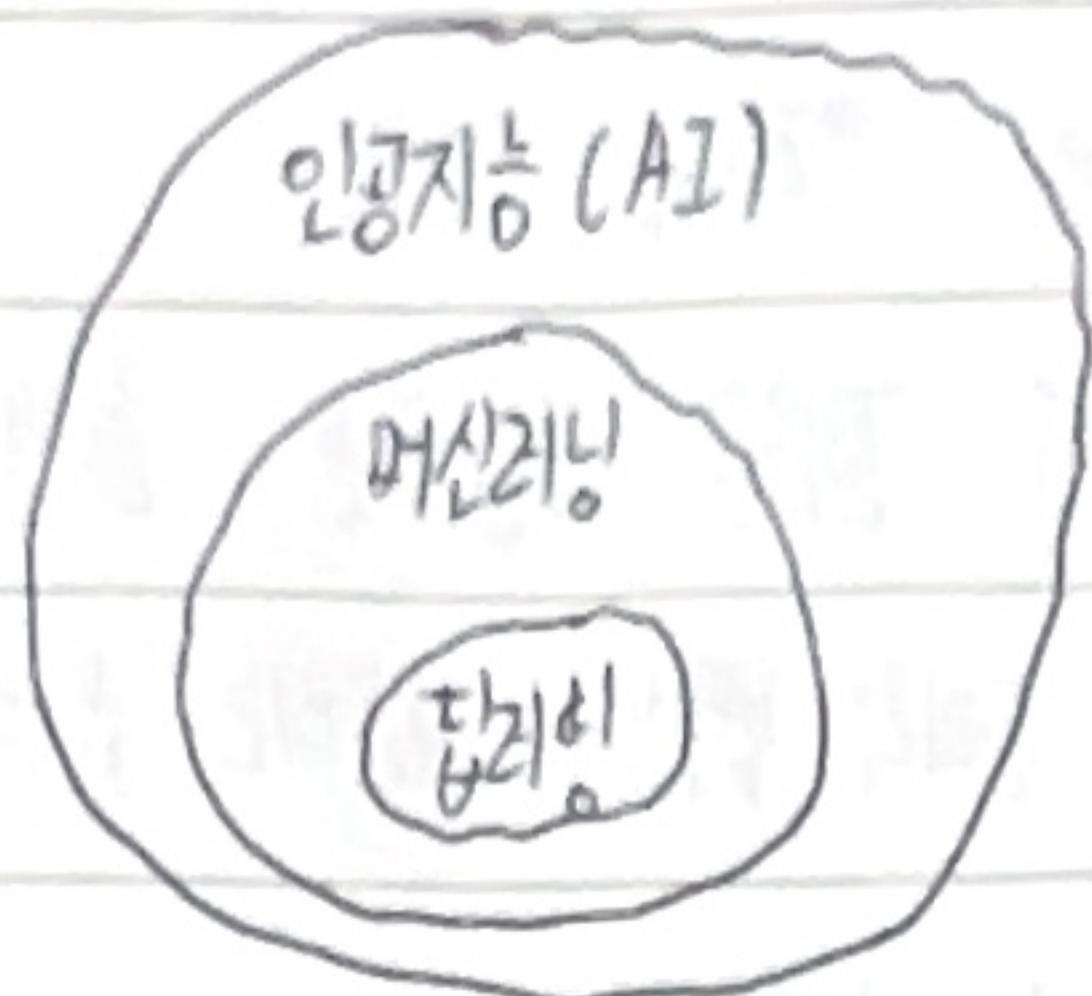
전통적인 프로그래밍은 개발자가 logic을 정의하고 이를 기본으로 프로그램이 동작한다. (입력 → 규칙 → 출력)

AI는 입력과 출력을 함께 주고 규칙을 스스로 도출하게 한다. (입력 + 출력 → 규칙 (API) → 데이터 예측)

## 4. 딥러닝과 머신러닝의 차이점은 무엇인가?

머신러닝은 전통적인 알고리즘을 사용하여 수동으로 각 데이터의 특징을 알아내고 다양한 학습 방식으로 이를 진행한다.

딥러닝은 인공 신경망을 사용하여 자동으로 데이터의 특징을 알아내고 기계학습을 가장 많이 사용한다.



## 5. Classification과 Regression의 주된 차이점은?

분류(Classification)은 범주형 데이터를 예측하는 문제로 이메일 스팸 필터링, 질병진단 등이 포함된다.

회귀(Regression)는 연속적인 숫자 값을 예측하는 문제로 주택가격 예측, 주식 시장 예측 등이 포함된다.

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

데이터 학습을 위해 차원이 증가하면서 학습 데이터 수가 차원의 수보다 적어져 성능이 저하되는 현상

→ 차원이 증가함에 따라 모델 성능이 안 좋아지는 현상

## 7. Dimensionality Reduction는 왜 필요한가?

차원축소(Dimensionality Reduction)는 고차원의 데이터로부터 저차원의 데이터로 변환하는 방법

- 비용, 시간, 자원, 용량 문제
  - 차원이 높으면 과적합 문제
  - 군집화 분석 결과가 좋지 않은
  - 차원이 높으면 설명력이 떨어짐
- ) 이런 문제를 해결하기 위해 필요

## 8. Ridge와 Lasso의 공통점과 차이점(Regularization, 규제, scaling)

공통점으로는 풀다 과적합을 방지하기 위해 가중치에 제한리를 부여하고 Ridge, Lasso 모두 선형 회귀 모델에 적용되는 정규화 기법이다. 정규화 과정에서 가중치 크기가 중요하기 때문에 정규화를 적용하는 것이 일반적이다.

차이점은 Ridge는 L2 정규화를 적용하여 모든 회귀 계수를 0에 가깝게 줄이지만 0으로 만들지 않고 모든 변수를 사용하면서 모델 복잡도를 조절한다. Lasso는 L1 정규화를 적용하여 일부 회귀계수를 0으로 만들어 변수 선택 가능을 수행하며 불필요한 변수를 제거하여 모델을 더 간결하게 만든다.

## 9. Overfitting Vs Underfitting

Overfitting은 모델이 데이터 사이의 의존성뿐만 아니라 랜덤 노이즈까지 모두 학습했을 때 발생한다. 데이터를 너무 과도하게 학습한 경우를 말한다.

Underfitting은 데이터 사이에서 의존성을 정확하게 잡아내지 못할 때 발생한다. 새로운 데이터에 적용할 때 일반화 능력이 떨어진다.

## 10. Feature Engineering과 Feature Selection의 차이점은?

Feature Engineering은 주어진 데이터를 변형하거나 새로운 특성을 생성하여 모델의 성능을 향상시키는 방법이다.

Feature Selection은 주어진 특성 중 불필요한 변수나 중요도가 낮은 변수를 제거하여 모델의 성능을 최적화하는 방법이다.

⇒ Feature Engineering은 새로운 특성을 만들거나 변형하는 과정이고 Feature Selection은 기존 특성 중 중요한 것을 선택하는 과정이다.

## 11. 전처리(Preprocessing)의 목적과 방법?(노이즈, 이상치, 결측치)

전처리는 머신러닝 모델이 최적의 성능을 발휘하도록 데이터를 정리하는 과정이다.

1) 노이즈 처리: 평균 또는 중앙값 필터 사용, 신호 처리 기법 사용

2) 이상치 처리: 이상치는 데이터에서 극단적인 값으로 특정적인 영향을 줄 수 있어 IQR, Z-score를 사용하여 제거한다.

3) 결측치 처리: 결측치는 데이터셋에서 특정 값이 없는 경우로 이를 처리하지 않으면 모델 학습에 오류가 발생한다.

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

EDA는 데이터셋을 분석하여 그 특성과 구조를 파악하는 과정이다. 데이터의 분포, 상관관계, 이상치, 결측치 등을 시각화하고 통계적으로 분석하여 데이터에 대한 주관적 이해를 풍습니다.

- 데이터의 분포: 각 변수의 분포를 이해하여 데이터의 특성을 파악한다

- 상관관계 분석: 변수 간의 상관관계를 파악하여 어떤 변수들이 함께 변하는지 확인한다.

## 13. 회귀에서 절편과 기울기가 의미하는 바는? 딥러닝과 어떻게 연관되는가?

선형 회귀 모델은 다음과 같은 형태로 표현된다.  $y = \beta_0 + \beta_1 x$

- $\beta_0$ (절편): 모델이 시작하는 기본값을 나타냄

- $\beta_1$ (기울기): 데이터의 경향성이나勾配 강도를 나타냄

딥러닝과는 가중치와 편향으로 확장되어 사용된다.

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

불순도는 결정트리에서 각 노드가 얼마나 혼합되어 있는지 나타내는 척도이다. 불순도가 낮을수록 해당 노드는 더 순수하며 각 데이터 포인트가 하나의 클래스에 속할 가능성이 높다.

지니계수는 결정트리에서 노드의 불순도를 측정하는 방법으로 분류 문제에 사용되며 특정노드가 얼마나 순수한지 나타낸다.

## 29. 앙상블이란 무엇인가?

앙상블이란 여러 약한 분류기를 결합하고 강 분류기로 만드는 것. 여러 개의 분류기를 생성하고 그 예측을 결합함으로써 보다 정확한 최종 예측을 도출하는 기법

유형으로는 Voting, Bagging, Boosting으로 나뉨

## 30. 부트 스트랩(bootstrap)이란 무엇인가?

부원 추출을 이용한 표본 재추출 방법을 뜻한다. 쉽게 말하면 뽑은 데이터를 다시 넣는 것

장점은 원본데이터와 비슷한 분포를 가진 데이터를 만들 수 있다.

### 31. 배깅(Bagging)이란 무엇인가?

기존 학습 데이터로부터 단순화하여 복원력을 하여 동일한 시그널을 만들어 여러 개의 앙상블을 구성하여 여러 모델을 학습시키는 방법이다.

여러 모델을 평균화하여 모델의 분산을 줄이고, 예측의 안정성을 높인다.

### 32. 주성분 분석(PCA)이란 무엇인가?

PCA는 차원축소에 쓰이는 기법으로 고차원 데이터를 저차원 공간으로 주영하여 중요한 정보를 보존하면서 데이터의 특성을 파악하는데 사용된다.