

머신러닝/딥러닝입문 가이드

기초 개념부터 실전 프로세스까지 한눈에 이해하기



© 2025 예영 EDU AI 연구소. All Rights Reserved.



목차



1. 큰 그림 이해- ML/DL의 기본 개념과 관계



2. ML의 세계- 활용 사례와 학습 방법 분류



3. 기본 용어- 데이터/모델/학습 용어 정리



4. 실전 프로세스- ML 7단계 및 평가 방법



5. 심화- 데이터 유형별 알고리즘 카탈로그

Part 1: 큰 그림 이해

머신러닝과 딥러닝의 기본 개념과 관계를 이해합니다.



1-1 머신러닝(Machine Learning) 이란?



1-2 딥러닝(Deep Learning) 이란?



1-3 머신러닝 vs 딥러닝 - 핵심 차이점



1-4 머신러닝 vs 데이터마이닝 - 관계와 차이



머신러닝(Machine Learning)이란?



데이터 기반 학습

컴퓨터가 데이터를 이용해 향후 적용 가능한 규칙을 스스로 생성하는 기술



지식 추출

대량의 데이터에서 의미 있는 패턴과 관계를 발견하여 지식을 추출하는 과정



인공지능의 하위 분야

인공지능의 한 계열로, 경험(데이터)을 통해 자동으로 개선되는 알고리즘 시스템



융합 학문 분야

통계학, 인공지능, 컴퓨터 과학이 융합된 학문 분야

머신러닝 관련 학문 분야

통계학
데이터 분석 및 확률 모델

+

인공지능
지능적 행동 및 의사결정

+

컴퓨터 과학
알고리즘 및 계산 효율성

= 머신러닝

머신러닝/딥러닝 입문 가이드

딥러닝(Deep Learning)이란?



인공신경망 기반 기술

다층 신경망(Neural Network)을 활용하여 심층적인 학습을 수행하는 머신러닝의 한 종류



자동 특징 추출

데이터로부터 자동으로 특징(feature)을 추출하여 인간의 수작업 설계 없이 학습 가능



대량 데이터 필요

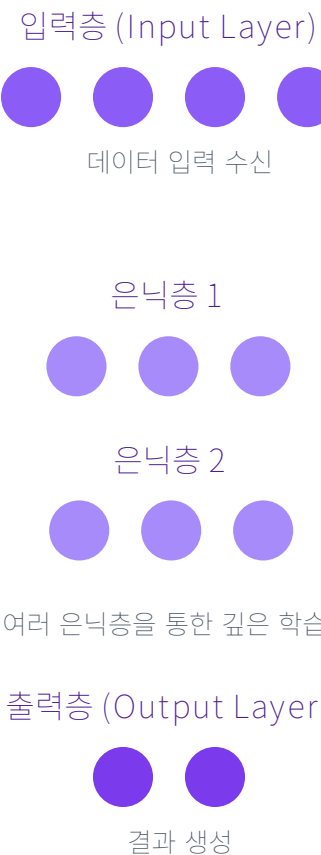
방대한 양의 데이터를 통해 정확도를 높이며, 고성능 하드웨어 (GPU) 요구



복잡한 문제 해결

이미지, 음성, 자연어 등 비정형 데이터 처리에 뛰어난 성능 발휘

딥러닝 신경망 구조



깊은 계층 구조로 복잡한 패턴 학습 가능

머신러닝 vs 딥러닝 비교



관계

딥러닝은 머신러닝의 하위 분야로, 인공신경망을 활용하여 데이터에서 더 복잡한 패턴을 학습합니다.



특징 추출 방식

머신러닝은 수작업으로 특징을 설계하고 추출하는 반면, 딥러닝은 데이터로부터 자동으로 특징을 학습합니다.



데이터 요구량

머신러닝은 상대적으로 적은 데이터로도 학습이 가능하지만, 딥러닝은 대량의 데이터가 필요합니다.



하드웨어 요구사항

머신러닝은 일반 컴퓨팅 환경에서도 실행 가능하나, 딥러닝은 고 사양 GPU 같은 특수 하드웨어가 필요합니다.

주요 차이점 비교

머신러닝		딥러닝
해석이 비교적 쉽고 투명함	vs	'블랙박스' 형태로 해석이 어려움
표 형식의 정형 데이터에 적합 (ex: 신용평가, 마케팅 예측)	vs	비정형 데이터 처리에 뛰어남 (ex: 이미지, 음성, 자연어)
SVM, 결정트리, 랜덤포레스트 등 다양한 알고리즘 활용	vs	CNN, RNN, Transformer 등 신경망 구조 활용
명시적 특징 정의가 필요	vs	특징을 자동으로 추출하고 학습

머신러닝과 데이터마이닝의 관계

-  데이터마이닝이란?
대용량 데이터에서 패턴, 규칙을 탐색하고 모델화하여 유용한 지식을 추출하는 과정
-  상호 연관성
데이터마이닝은 복잡성 높은 데이터 분석에 머신러닝 이론이 적용됨
-  주요 초점 차이
머신러닝: 예측에 집중
데이터마이닝: 숨겨진 속성 발견에 집중
-  활용 분야
머신러닝: 이미지 인식, 추천 시스템, 자연어 처리
데이터마이닝: 장비구니 분석, 이상 감지, 시장 세분화



Part 2. ML의 세계

머신러닝/딥러닝의 주요 활용 사례와 학습 방법의 다양한 유형을 소개합니다.



머신러닝은 다양한 분야 에서 활용되며, 손글씨 인식부터 자율주행까지 우리 일상 곳곳에 적용되고 있습니다.



지도학습, 비지도학습, 강화학습 등 다양한 학습 방법을 통해 컴퓨터는 데이터로부터 규칙을 발견합니다.



회귀, 분류, 군집화 등 문제 유형에 따라 적합한 알고리즘을 선택하여 최적의 결과를 도출합니다.

머신러닝/딥러닝 주요 활용 사례

이미지/시각 인식



손글씨 우편번호 인식

우편물의 손글씨 주소 자동 인식 및 분류



의료 영상 진단

MRI, X-ray 등에서 종양 감지 및 질병 진단



객체 및 얼굴 인식

영상에서 얼굴, 사물 등 다양한 객체 탐지



문자 인식

손글씨, 인쇄 텍스트 자동 판독



의료 진단

질병 감지 및 예측 의료 지원



보안 및 이상 탐지

침입 탐지, 사이버 보안



자연어 처리

언어 이해, 번역, 대화 시스템



자율주행

객체 감지, 경로 계획, 신호 인식



창작 및 예술

이미지, 음악, 영상 생성

머신러닝/딥러닝 주요 활용 사례

데이터 분석 및 보안



부정 거래 감지

신용카드 이상 거래 패턴 탐지 및 실시간 방어



고객 군집화

유사한 취향/행동 패턴으로 고객 그룹화



언어/생성형 AI



텍스트 분류 및 요약

블로그, 뉴스 글 주제 분류 및 콘텐츠 요약



생성형 AI

ChatGPT, DALL-E 등 대화/이미지 생성 기술



문자 인식

손글씨, 인쇄 텍스트 자동 판독



의료 진단

질병 감지 및 예측 의료 지원



보안 및 이상 탐지

침입 탐지, 사이버 보안



자연어 처리

언어 이해, 번역, 대화 시스템



자율주행

객체 감지, 경로 계획, 신호 인식



창작 및 예술

이미지, 음악, 영상 생성

머신러닝 학습 방법 분류

머신러닝은 학습 방식에 따라 크게 세 가지로 구분됩니다. 각각의 방식은 다른 문제 유형에 적합하며 서로 다른 장단점을 가지고 있습니다.



지도학습 (Supervised Learning)

정답(레이블)이 있는 데이터로 학습하는 방법
분류, 회귀 문제에 주로 활용
예: 이미지 분류, 스팸 필터링, 주가 예측



비지도학습 (Unsupervised Learning)

정답 없이 데이터의 패턴과 구조를 찾는 방법
군집화, 차원 축소, 연관 규칙에 활용
예: 고객 세그먼트, 이상치 탐지



강화학습 (Reinforcement Learning)

환경과 상호작용하며 보상을 최대화하는 방향으로 학습
시행착오를 통한 경험 기반 학습
예: 게임 에이전트, 자율주행, 로봇 제어

학습 방법 비교			
	지도학습	비지도학습	강화학습
데이터	입력값 + 정답(레이블)	입력값만 존재	상태 + 보상 신호
목표	예측/분류 모델	데이터 구조 파악	최적 행동 정책
알고리즘	SVM, 결정트리, 신경망	K-means, 계층적 군집	Q-learning, DQN
결과물	예측기(Predictor)	군집/패턴 발견	에이전트(Agent)
난이도	★★	★★★	★★★★★
핵심 차이점			
지도학습은 예측에, 비지도학습은 패턴 발견에, 강화학습은 의사결정에 최적화되어 있습니다.			

머신러닝 문제 유형



회귀 (Regression)

연속적인 값을 예측하는 문제 유형으로, 입력 데이터와 출력 값 사이의 상관관계를 학습합니다.

주택 가격 예측

매출액 예측

기온 예측



분류 (Classification)

데이터를 미리 정의된 범주로 구분하는 문제 유형으로, 이메일 스팸 필터링 같은 예/아니오 판단이나 다중 카테고리 분류에 활용됩니다.

스팸 메일 분류

질병 진단

이미지 인식



군집 (Clustering)

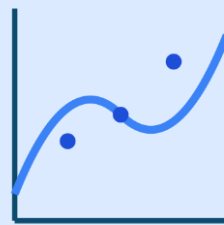
레이블이 없는 데이터를 유사한 특성을 가진 그룹으로 자동 분류하는 비지도학습 문제 유형입니다.

고객 세분화

유사 문서 그룹화

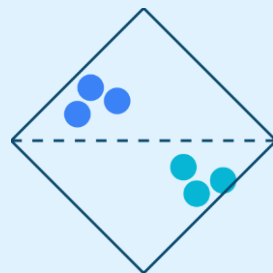
이상치 탐지

회귀 (Regression)



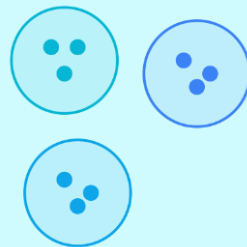
연속적인 숫자 값을 예측하는 회귀 모델은 데이터 포인트들 사이의 관계를 표현하는 선형 또는 비선형 함수를 찾습니다.

분류 (Classification)



분류 모델은 데이터를 서로 다른 카테고리로 구분하는 경계를 학습하여 새로운 데이터 포인트가 어떤 범주에 속하는지 결정합니다.

군집 (Clustering)



군집화는 데이터 포인트들을 유사성에 따라 그룹화하여 숨겨진 패턴을 발견하고 데이터의 구조를 이해하는데 도움을 줍니다.

Part 3. 기본 용어

머신러닝/딥러닝 분야에서 꼭 알아야 할 핵심 용어들을 알기 쉽게 정리했습니다.



1. 데이터 용어- 데이터셋, 샘플, 특성, 레이블



2. 모델 용어- 모델, 모델링, 알고리즘, 학습



3. 학습 용어- 학습 데이터, 검증 데이터, 테스트 데이터



4. 평가 용어- 정확도, 오차, 과적합, 과소적합

머신러닝/딥러닝 기본 용어



데이터셋 (Dataset)

머신러닝 모델을 학습하고 평가하는 데 사용되는 데이터 모음. 학습 데이터와 테스트 데이터로 분리하여 사용



샘플, 데이터 포인트 (Sample)

데이터셋의 개별 항목 또는 행. 하나의 개체에 대한 관측값



특징/특성 (Feature)

입력 데이터에서 모델 예측에 사용되는 개별 변수 또는 속성. 예측 인자, 독립변수, 인풋이라고도 함



레이블 (Label)

지도학습에서 알고리즘이 예측하려는 목표값. 분류 문제에서는 범주, 회귀 문제에서는 연속값



특성 추출/엔지니어링

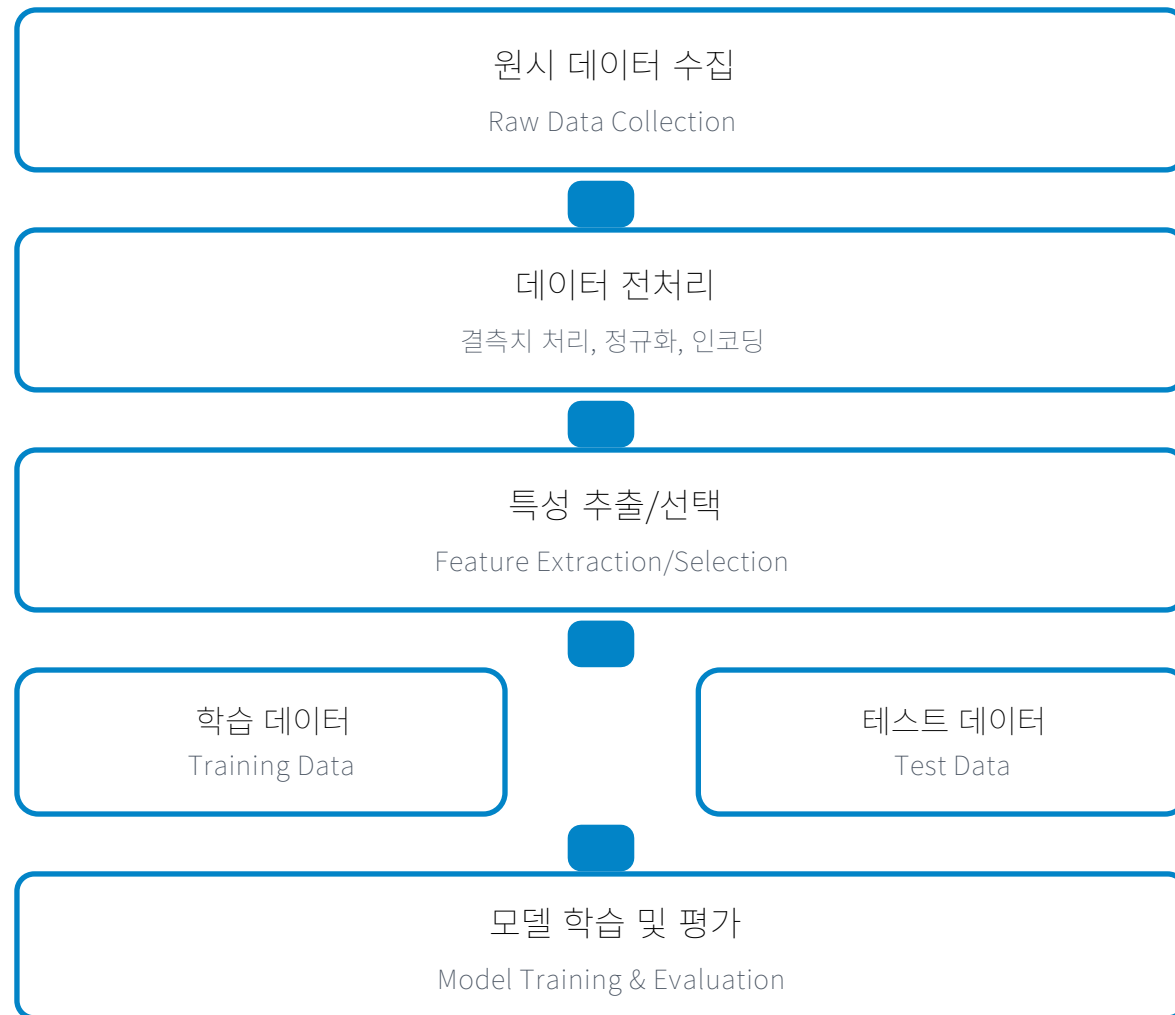
원시 데이터에서 유의미한 특성을 수동 또는 자동으로 추출하고 변환하는 과정



모델/모델링

데이터의 추상적 수학적 표현으로, 입력을 출력으로 변환하는 함수 또는 알고리즘

머신러닝 데이터 흐름



용어는 실제 분석 흐름에 맞게 이해하는 것이 중요

Part 4. 실전 프로세스

 ML/DL 프로젝트의 실제 분석 진행 단계(7단계), 평가 방법, 과적합 등 주의사항 안내

머신러닝 7단계 프로세스

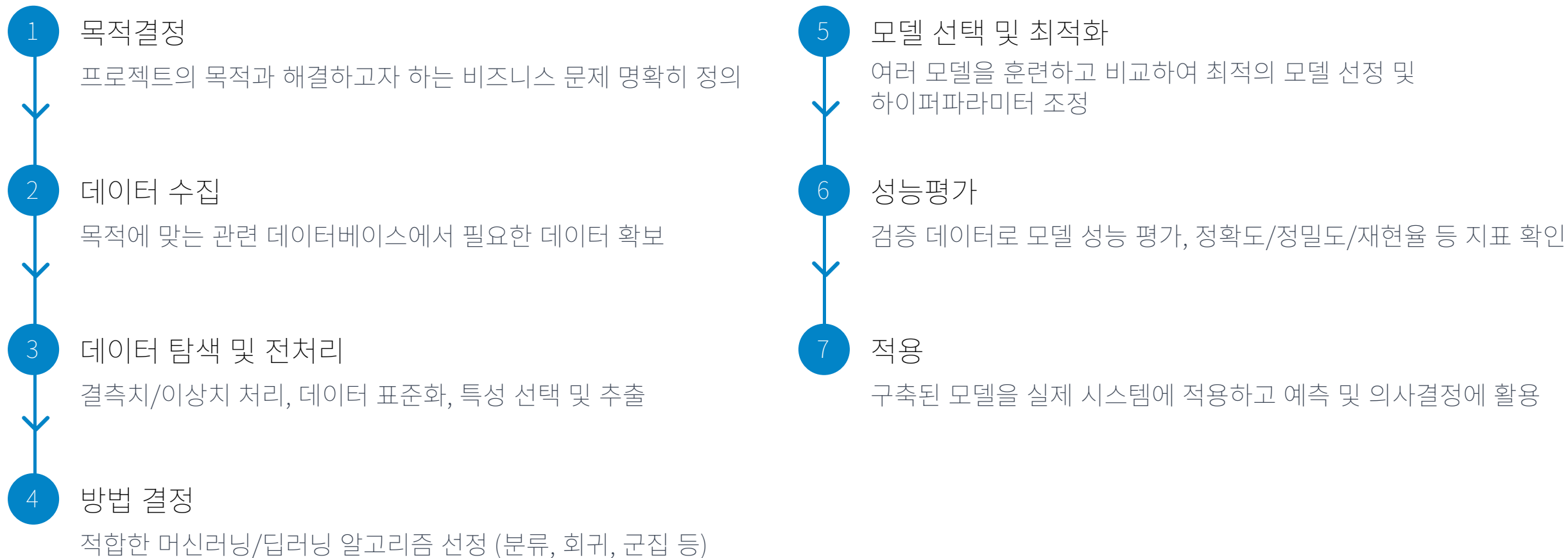
-  1. 목적결정
-  2. 데이터 수집
-  3. 데이터 전처리
-  4. 방법 선정

평가 및 적용

-  5. 모델 선택/최적화
-  6. 성능평가
-  7. 적용 및 의사결정
-  주의: 과적합(Overfitting) 방지

머신러닝/딥러닝 실전 7단계

머신러닝/딥러닝 프로젝트의 체계적인 단계별 진행 프로세스



실무 팁

각 단계는 반복적으로 수행될 수 있으며, 모델의 성능 개선을 위해 데이터 전처리부터 모델 선택까지 여러 번 진행하는 것이 일반적입니다.