

강원지역혁신플랫폼

기계학습

Machine Learning

컴퓨터가 데이터에서 배우는 원리



▶ 학습목표

📁 컴퓨터가 데이터에서 배우는 원리를
설명할 수 있습니다.





01 | 컴퓨터는 데이터에서 배움



컴퓨터는 데이터에서 배움

△ 현대 기술 시대에는 정형 또는 비정형 데이터가 매우 풍부함

◆ 20세기 후반에 데이터에서 지식을 추출하여 예측하는 자가 학습(self-learning)

알고리즘과 관련된 인공지능(Artificial Intelligence, AI)의 하위 분야로 기계학습이 출현했음

› 전통적인 접근 방식에서는 사람이 수동으로 대량의 데이터를 분석하여 규칙을 유도하고 모델을 만들

› 기계학습은 데이터에서 더 효율적으로 지식을 추출하여

예측 모델과 데이터 기반의 의사결정 성능을 점진적으로 향상시킬 수 있음



01 | 컴퓨터는 데이터에서 배움



컴퓨터는 데이터에서 배움

△ 컴퓨터 과학 연구에서 기계학습은 점점 더 중요해지고 있음

◆ 우리 일상생활 전반에서도 아주 큰 역할을 하고 있음

▶ 기계학습 덕택에 다음과 같은 견고한 애플리케이션 프로그램을 이용할 수 있음

- 견고한 이메일 스팸 필터
- 편리한 텍스트와 음성 인식 소프트웨어
- 믿을 수 있는 웹 검색 엔진
- 체스 대결 프로그램
- 안전하고 효율적인 자율 주행 자동차
- 의료 애플리케이션



02 | 지도학습으로 미래 예측

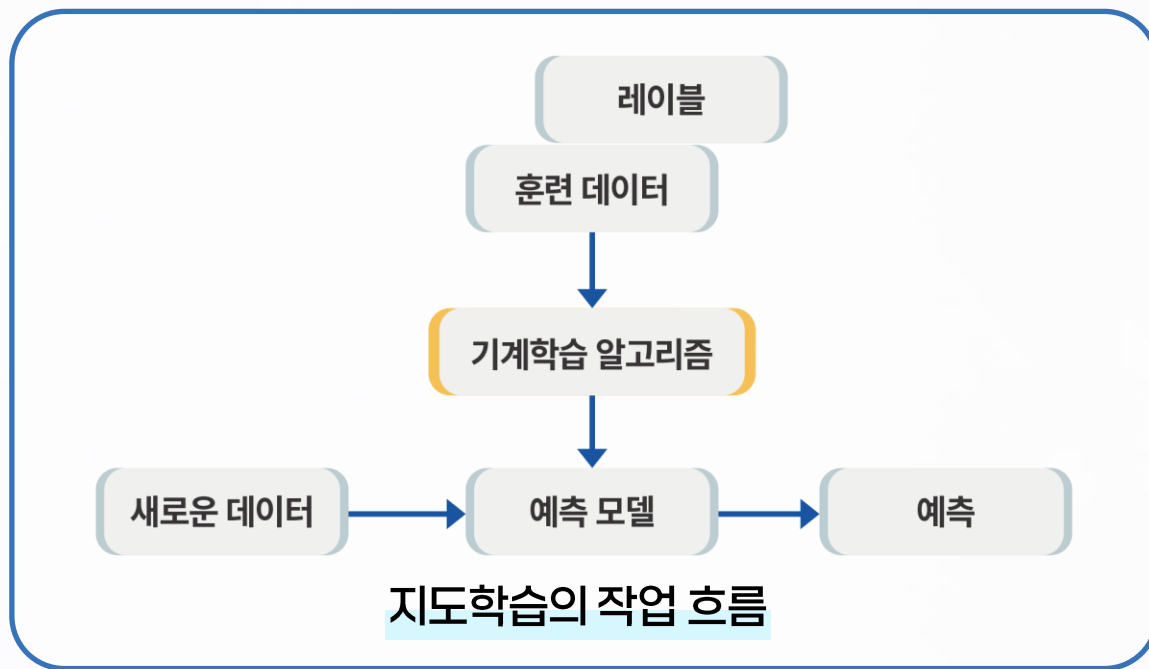


지도학습으로 미래 예측

△ 지도학습(supervised learning)의 주요 목적은 레이블(label) 된 훈련 데이터에서 모델을 학습하여 본 적 없는 미래 데이터에 대해 예측을 만드는 것임

◆ 여기서 지도(supervised)는 희망하는 출력 신호(레이블)가 있는 일련의 샘플(데이터)을 의미함

➤ 아래의 그림은 전형적인 지도학습 작업 흐름을 나타냄





02 | 지도학습으로 미래 예측



지도학습으로 미래 예측

△ 스팸 메일을 필터링하는 예를 생각해 보자.

◆ 레이블된 이메일 데이터셋에서 지도학습 기계학습 알고리즘을 사용하여 모델을 훈련할 수 있음

➤ 이 데이터셋은 스팸 또는 스팸이 아닌 이메일로 정확하게 표시되어 있음

— 훈련된 모델은 새로운 이메일이 두 개의 범주(category) 중 어디에 속하는지 예측함

➔ 이메일 스팸 필터의 예처럼 개별 클래스 레이블이 있는 지도 학습을 분류(classification)라고 부름

➔ 연속적인 값을 출력하는 지도학습을 회귀(regression)라고 부름



02 | 지도학습으로 미래 예측: 분류



지도학습으로 미래 예측

△ 분류 : 클래스 레이블 예측

◆ 분류는 지도학습의 하위 카테고리임

▶ 과거의 관측을 기반으로 새로운 샘플의 범주형 클래스 레이블을 예측하는 것이 목적임

– 클래스 레이블은 이산적(discrete)이고 순서가 없어 샘플이 속한 그룹으로 이해할 수 있음

➔ 앞의 스팸 메일 필터는 전형적인 이진 분류(binary classification) 작업의 예임

➔ 스팸과 스팸이 아닌 이메일 두 개의 클래스 사이를 구분하려고
기계학습 알고리즘이 일련의 규칙을 학습함

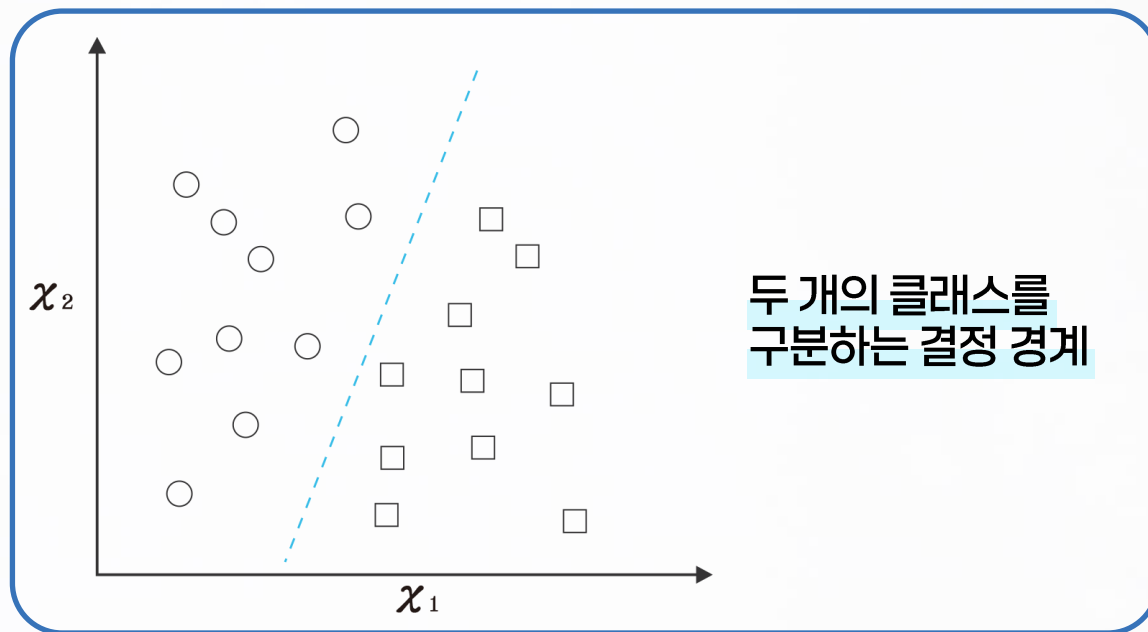


02 | 지도학습으로 미래 예측: 분류



지도학습으로 미래 예측

- ◆ 아래 그림은 20개의 훈련 샘플이 있는 이진 분류 작업의 개념을 나타냄
 - 10개의 샘플은 동그라미로 레이블 되어 있고, 다른 10개의 샘플은 네모로 레이블 되어 있음
 - ▬ 각 샘플이 두 개의 x_1 , x_2 값에 연관되어 있으므로 2차원 데이터셋임
- ➔ 지도학습 알고리즘을 사용하여 두 클래스를 구분할 수 있는 규칙을 학습함



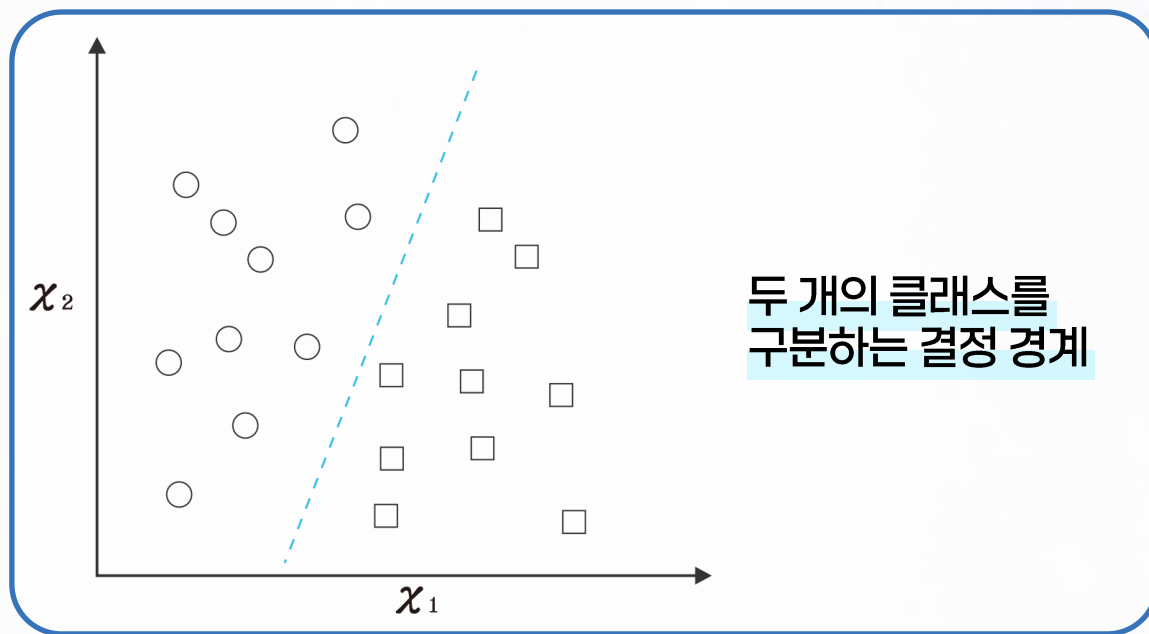


02 | 지도학습으로 미래 예측: 분류



지도학습으로 미래 예측

- ◆ 아래의 그림에서 지도학습 알고리즘이 두 클래스를 구분하는 규칙은 점선으로 나타난 결정 경계(decision boundary)임
 - 새로운 데이터의 x_1, x_2 값이 주어지면 두 개의 범주 중 하나로 분류함





02 | 지도학습으로 미래 예측: 분류



지도학습으로 미래 예측

⚙️ 두 개 이상의 클래스 레이블을 가진 경우를 생각해 보자.

◆ 지도학습 알고리즘으로 학습한 예측 모델은 훈련 데이터셋에 있는 여러 클래스 레이블을 새로운 샘플에 할당할 수 있음

➤ 이런 다중 분류(multiclass classification)의 전형적인 예는 손글씨 숫자 인식임





02 | 지도학습으로 미래 예측: 회귀



지도학습으로 미래 예측

△ 회귀: 연속적인 출력 값 예측

◆ 회귀는 연속적인 출력 값을 예측하는 것임

➤ 회귀는 예측 변수(또는 설명 변수)와 연속적인 반응 변수(또는 종속 변수)가 주어졌을 때 출력 값을 예측하기 위해 두 변수 사이의 관계를 찾음

─ 기계학습 분야에서는 예측 변수를 일반적으로 “특성(feature)”이라고 부름

➔ 반응 변수를 “타겟(target)”이라고 부름

➔ 예를 들어 대학생의 키로 몸무게를 예측한다고 가정해 보자.

▪ 예측 변수를 키가 되고, 반응 변수는 몸무게가 됨



02 | 지도학습으로 미래 예측: 회귀



지도학습으로 미래 예측

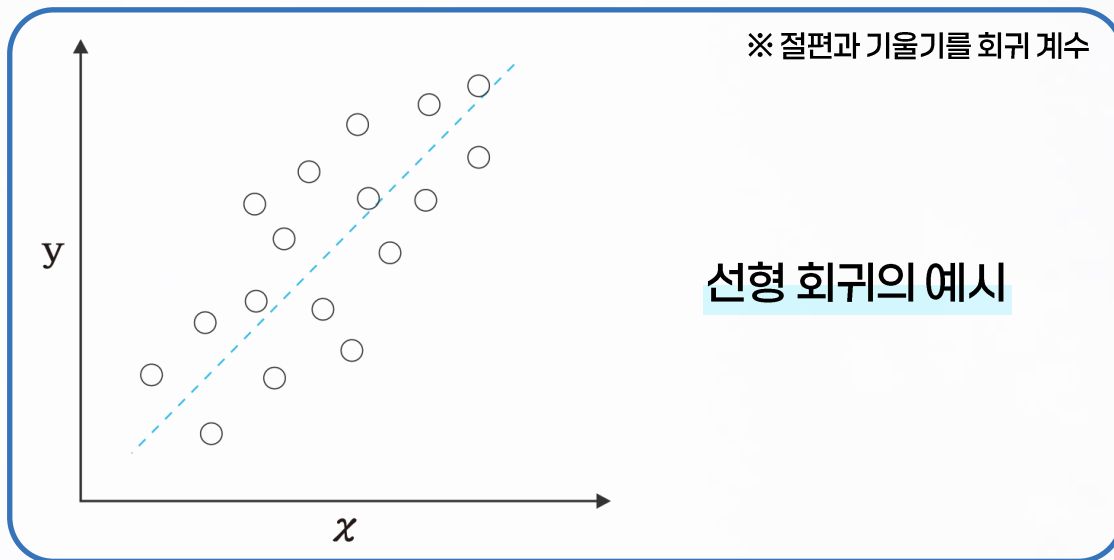
◆ 아래의 그림은 **선형 회귀**(linear regression)의 개념을 나타냄

➤ 특성 x 와 타겟 y 가 주어지면 데이터 포인트와 직선 사이 거리가 최소가 되는 직선을 그을 수 있음

— 이 거리는 일반적으로 **평균 제곱 거리**를 사용함

➔ 데이터에서 학습한 직선의 기울기와 절편(intercept)을 사용하여 새로운 데이터의 출력 값을 예측함

■ 학생의 몸무게(kg) = $\beta_0 + \beta_1 \times$ 학생의 키(cm)
절편 기울기





03 | 비지도 학습으로 숨겨진 구조 발견



비지도 학습으로 숨겨진 구조 발견

- ⚙ 비지도 학습(unsupervised learning)에서는 레이블 되지 않거나 구조를 알 수 없는 데이터를 다룸
 - ◆ 비지도 학습 기법을 사용하면 알려진 출력 값이나 보상 함수의 도움을 받지 않고 의미 있는 정보를 추출하기 위해 데이터 구조를 탐색할 수 있음



비지도 학습 기법으로 숨겨진 구조 발견



03 | 비지도 학습으로 숨겨진 구조 발견: 군집



비지도 학습으로 숨겨진 구조 발견

△ 군집: 서브그룹 찾기

- ◆ 군집(clustering)은 사전 정보 없이 쌓여 있는 그룹 정보를 의미 있는 서브그룹(subgroup) 또는 클러스터(cluster)로 조직하는 탐색적 데이터 분석 기법임
 - 분석 과정에서 만든 각 클러스터는 어느 정도 유사성을 공유하고 다른 클러스터와는 비슷하지 않은 샘플 그룹을 형성함
 - 군집은 정보를 조직화하고 데이터에서 의미 있는 관계를 유도하는 훌륭한 도구임
 - ➔ 예를 들어 마케터가 관심사를 기반으로 고객을 그룹으로 나누어 각각에 맞는 마케팅 프로그램을 개발할 수 있음

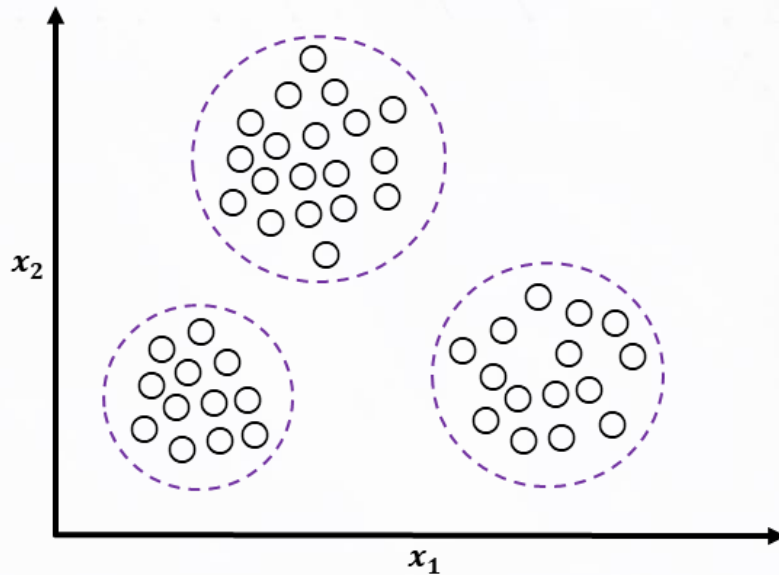


03 | 비지도 학습으로 숨겨진 구조 발견: 군집



비지도 학습으로 숨겨진 구조 발견

- ◆ 아래 그림은 군집이 어떻게 레이블 되지 않는 데이터를 특성 x_1 과 x_2 의 유사도를 기반으로 세 개의 개별적인 그룹으로 조직화하는지 보여 줌



각 클러스터는 어느 정도 유사성,
다른 클러스터는 비슷하지 않은 샘플 그룹을 형성



03 | 비지도 학습으로 숨겨진 구조 발견: 차원 축소



비지도 학습으로 숨겨진 구조 발견

△ 차원 축소: 데이터 압축

◆ 차원 축소(dimensionality reduction)는 고차원의 데이터를 다루어야 하는 경우 주로 사용함

▶ 즉, 하나의 관측 샘플에 많은 측정 지표가 있음

─ 이로 인해 기계학습 알고리즘의 계산 성능과 저장 공간의 한계에 맞닥뜨릴 수 있음

➔ 차원 축소는 잡음(noise) 데이터를 제거하기 위해
특성 전처리 단계에서 종종 적용하는 방법임

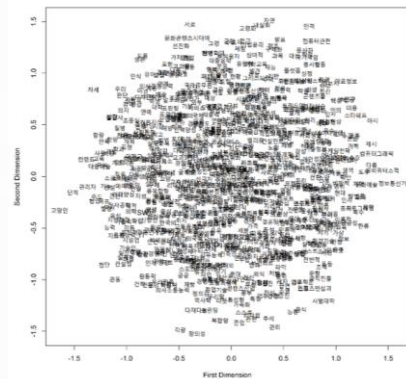


▶ 차원 축소는 관련 있는 정보를 대부분 유지하면서

더 작은 차원을 가진 부분 공간(subspace)으로 데이터를 압축함

1 차원 축소는 데이터 시각화에도 유용함

➔ 예를 들어 아래 그림과 같이 **100차원**(고차원) 특성을 **2차원 특성 공간**으로 **축소**하여 **2D 산점도**로 **시각화**할 수 있음



단어를 100차원으로 학습 후 2차원으로 축소 결과

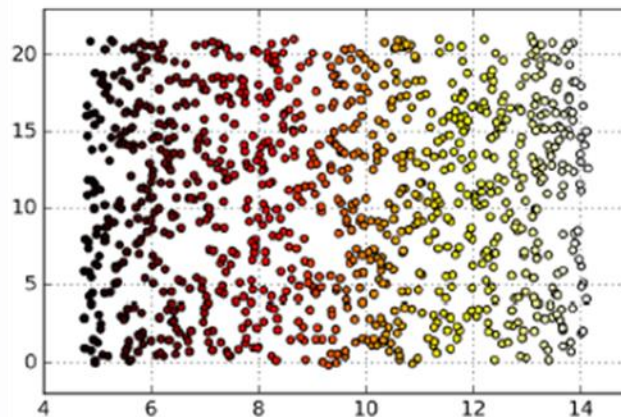
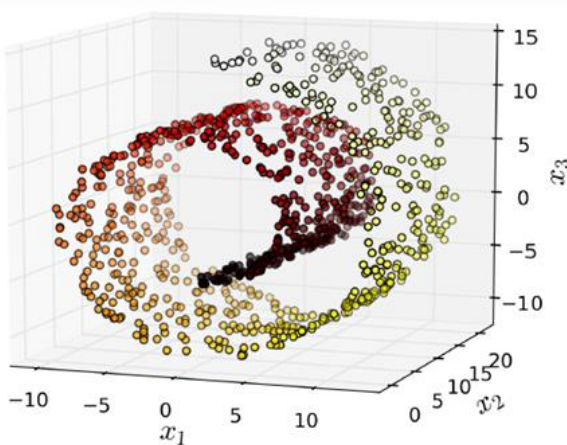


03 | 비지도 학습으로 숨겨진 구조 발견: 차원 축소



비지도 학습으로 숨겨진 구조 발견

- ◆ 아래 그림은 비선형(nonlinear) 차원 축소를 적용하여 **3D 스위스 롤(Swiss Roll) 모양의 데이터**를 **2D 특성의 부분 공간**으로 **압축**하는 예를 보여주고 있음
- 3D 스위스 롤 모양의 데이터 → 2D 특성의 부분 공간으로 압축



스위스 롤 데이터셋을 3D 특성 → 2D 특성으로 압축



04 | 강화 학습으로 반응형 문제 해결



강화 학습으로 반응형 문제 해결

- △ 강화 학습은 환경과 상호 작용하여 시스템(에이전트) 성능을 향상하는 것이 목적임
 - ◆ 환경의 현재 상태 정보는 보상(reward) 신호를 포함하기 때문에 지도학습과 관련된 분야로 생각할 수 있음
 - › 보상 함수는 얼마나 행동이 좋은지를 측정한 값임
 - ─ 에이전트는 환경과 상호 작용하여 보상이 최대화되는 일련의 행동을 강화 학습으로 학습함
 - ➔ 탐험적인 시행착오(trial and error) 방식이나 신중하게 세운 계획을 사용함



04 | 강화 학습으로 반응형 문제 해결

강화 학습으로 반응형 문제 해결

△ 강화 학습에는 여러 하위 분류가 있음

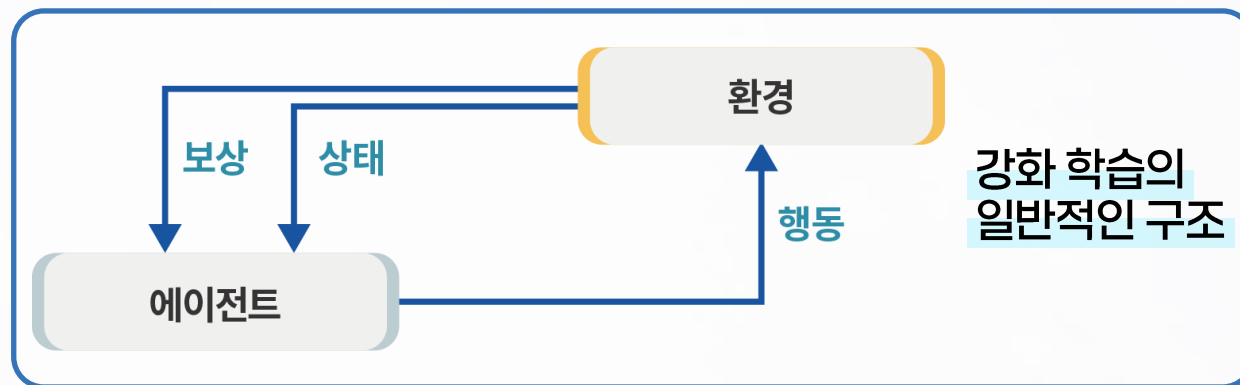
◆ 아래의 그림은 강화 학습의 일반적인 구조를 나타냄

➤ 강화 학습 에이전트는 환경과 상호 작용하여 보상을 최대화하는 것임

─ 각 상태는 양의 보상이나 음의 보상과 연관됨

➔ 보상은 체스 게임의 승리나 패배처럼 전체 목표를 달성하는 것으로 정의할 수 있음

➔ 강화 학습은 행동을 수행하고 즉시 얻거나 지연된 피드백을 통해 얻은 전체 보상을 최대화하는 일련의 행동을 학습함



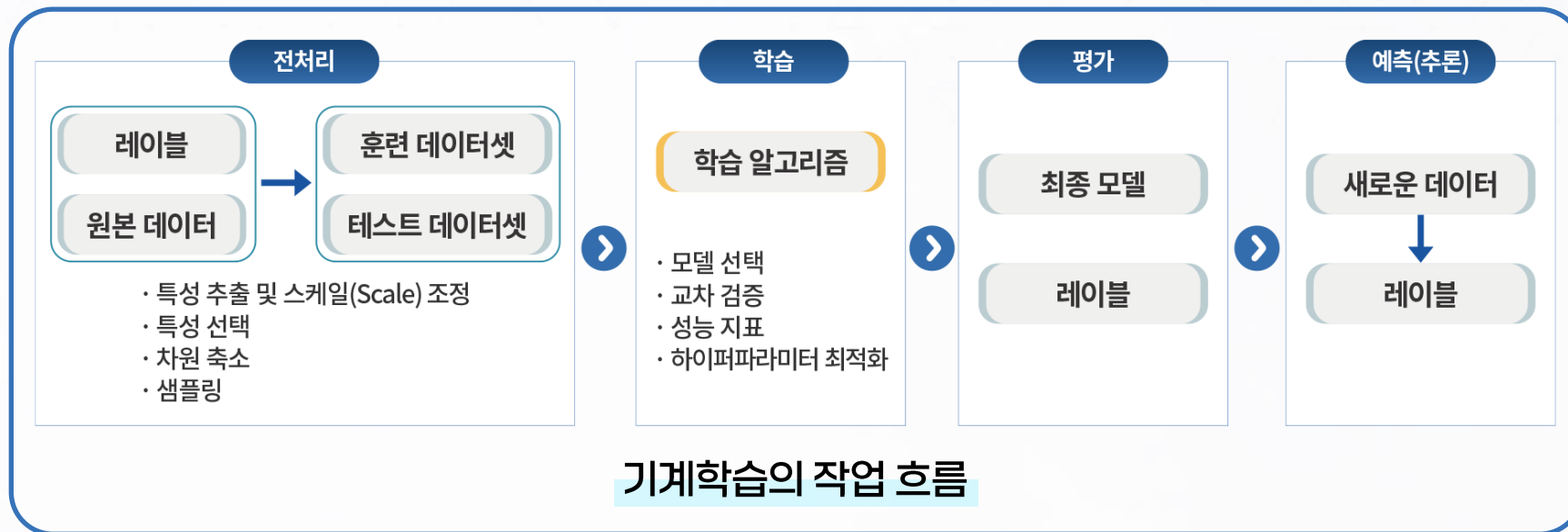


05 | 기계학습 시스템 구축 로드맵



기계학습 시스템 구축 로드맵

- ◆ 아래의 그림은 예측 모델링에 기계학습을 사용하는 전형적인 작업 흐름을 나타냄
 - 아래의 작업 흐름에서 학습 알고리즘과 함께 기계학습 시스템의 중요한 부분을 살펴봄





06 | 전처리: 데이터 형태 갖추기

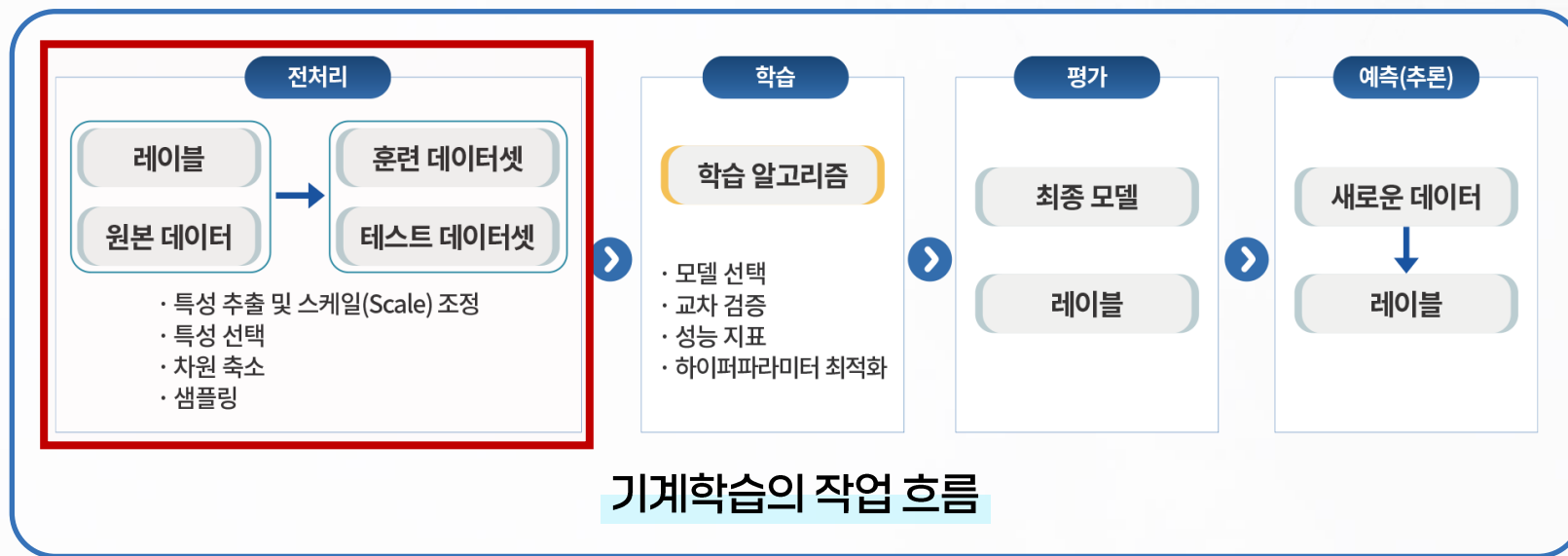


전처리: 데이터 형태 갖추기

△ 기계학습 시스템을 구축할 수 있는 로드맵(roadmap)을 생각해 보자.

◆ 학습 알고리즘이 주어진 원본 데이터로 최적의 성능을 내기에 적합한 경우는 매우 드물

➢ 데이터 전처리는 모든 기계학습 애플리케이션에서 가장 중요한 단계 중 하나임



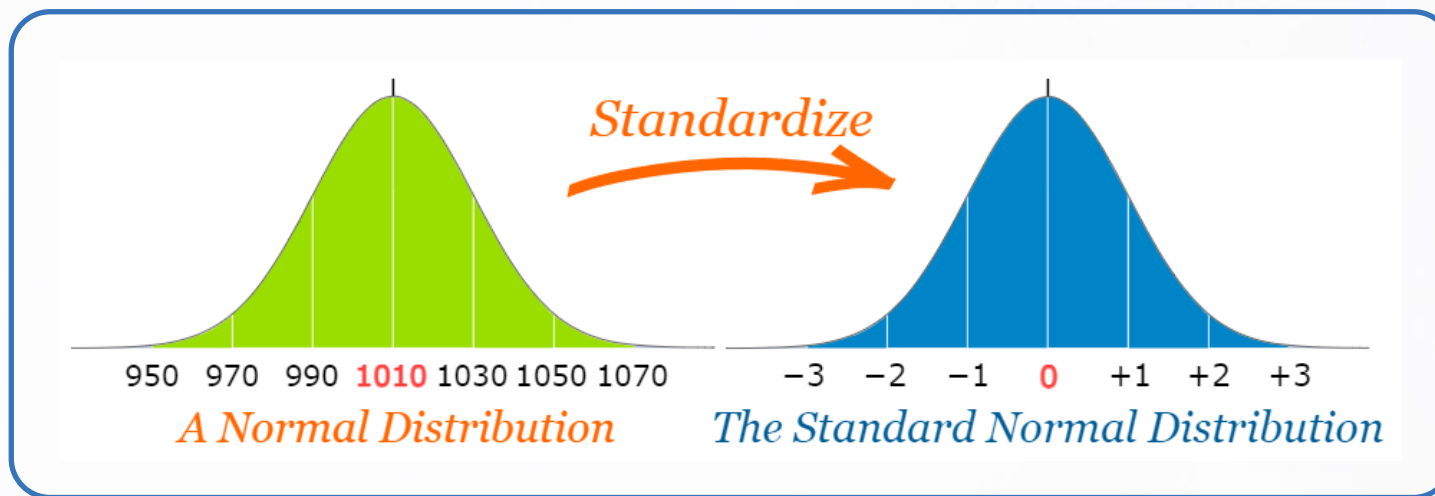


06 | 전처리: 데이터 형태 맞추기



전처리: 데이터 형태 맞추기

- △ 많은 기계학습 알고리즘에서 최적의 성능을 내려면 선택된 특성이 같은 스케일을 가져야 함
 - ◆ 특성을 $[0, 1]$ 범위로 변환하는 정규화로 변환
 - ◆ 평균이 0이고 단위 분산이 1을 가진 표준 정규 분포(standard normal distribution)로 변환



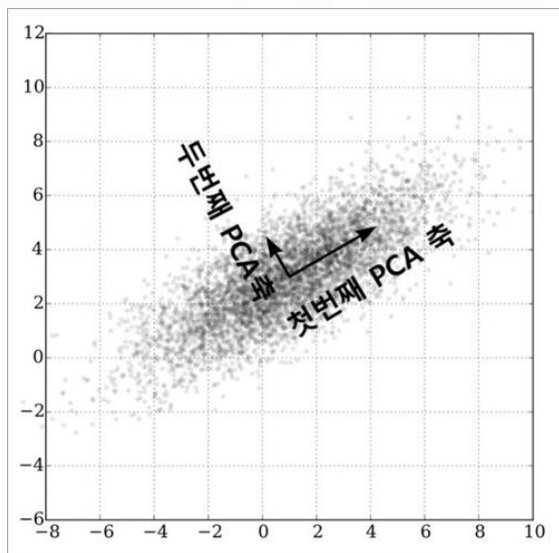


06 | 전처리: 데이터 형태 갖추기



전처리: 데이터 형태 갖추기

- ⚠ 일부 선택된 특성은 매우 상관관계가 높아 어느 정도 중복된 정보를 가질 수 있음
 - ◆ 이런 경우 차원 축소 기법을 사용하여 특성을 저차원 부분 공간으로 압축함
 - 특성 공간의 차원을 축소하면 저장 공간이 덜 필요하고 학습 알고리즘을 더 빨리 실행할 수 있음
 - ─ 어떤 경우에는 차원 축소가 모델의 예측 성능을 높이기도 함



차원 축소

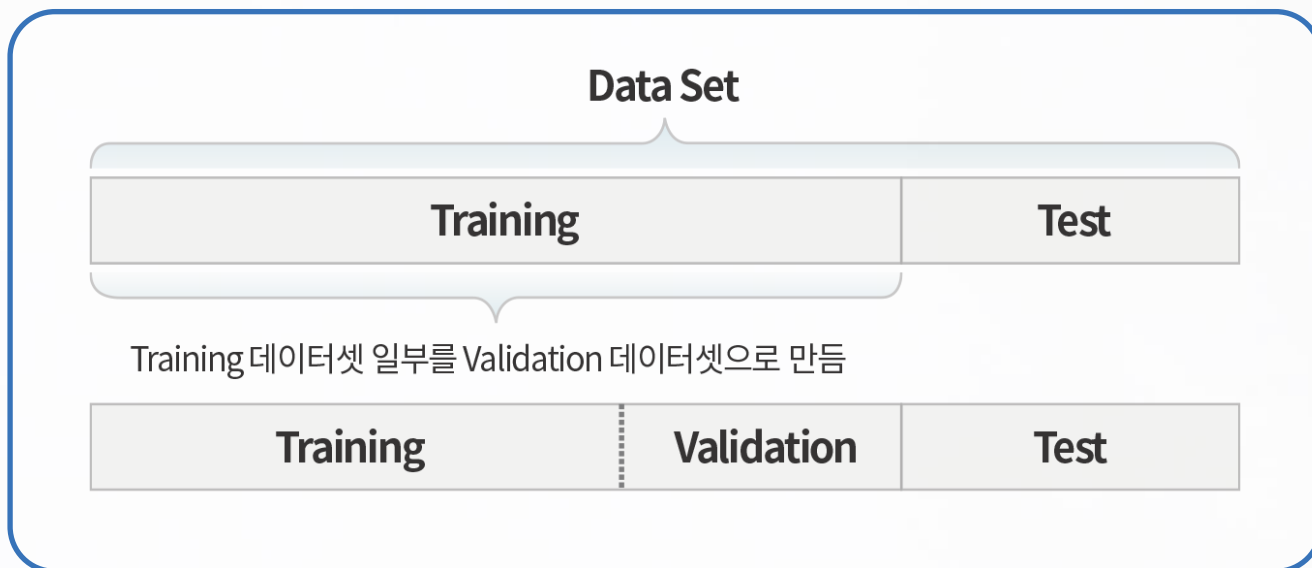


06 | 전처리: 데이터 형태 갖추기



전처리: 데이터 형태 갖추기

- ⚠ 기계학습 알고리즘이 **훈련 데이터셋**에서 잘 작동하고 **새로운 데이터셋**에서도 잘 **일반화되는지 확인**이 필요함
 - ◆ 이렇게 하려면 **랜덤하게 훈련 데이터셋**과 **테스트 데이터셋**으로 **나누어야 함**
 - **훈련 데이터셋**에서 **기계학습 모델**을 **훈련**하고 **최적화**함
 - **테스트 데이터셋**은 최종 **모델 평가 용도**로만 **사용**함





07 | 예측 모델 훈련과 선택



예측 모델 훈련과 선택

- ⚙ 기계학습 알고리즘은 각기 다른 문제를 해결하기 위해 개발되었음
- ◆ 데이비드 윌퍼트의 공짜 점심 없음 이론의 중요한 핵심 포인트는 아무런 대가도 치르지 않고 학습할 수는 없다는 것임

“공짜 점심 없음 (No free lunch)”

데이비드 윌퍼트 (David Wolpert)



07 | 예측 모델 훈련과 선택



예측 모델 훈련과 선택

- △ 예를 들어 **분류 알고리즘**은 저마다 **태생적인 편향**이 있음
 - ◆ 작업에서 **아무런 가정도 하지 않는다**면 **어떤 하나의 분류 모델**이 더 우월하다고 말할 수 없음
 - 현실에서는 **가장 좋은 모델**을 **훈련**하고 **선택**하기 위해 **최소한 몇 가지 알고리즘**을 **비교**해야 함

“가진 도구가 망치밖에 없다면 모든 문제가 못으로 보일 것입니다.”

에어브러햄 매슬로 (Abraham Maslow)



07 | 예측 모델 훈련과 선택



예측 모델 훈련과 선택

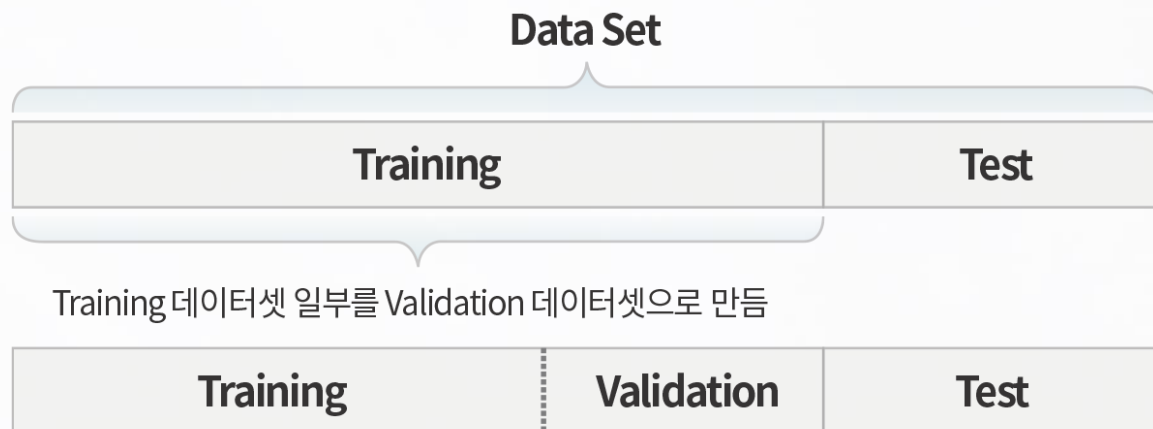
△ 모델 선택에 테스트 데이터셋을 사용하지 않고 최종 모델을 평가하는 방법을 생각해 보자.

◆ 테스트 데이터셋과 실제 데이터에서 어떤 모델이 잘 동작할지 어떻게 알 수 있을까요?

➤ 이 질문에 대한 답변을 위해 다음과 같은 방법을 생각할 수 있음

1 다양한 교차 검증 기법을 사용함

➔ 교차 검증에서는 모델의 일반화 성능을 예측하기 위해 훈련 데이터를
훈련 데이터 셋과 검증 데이터셋으로 나눔





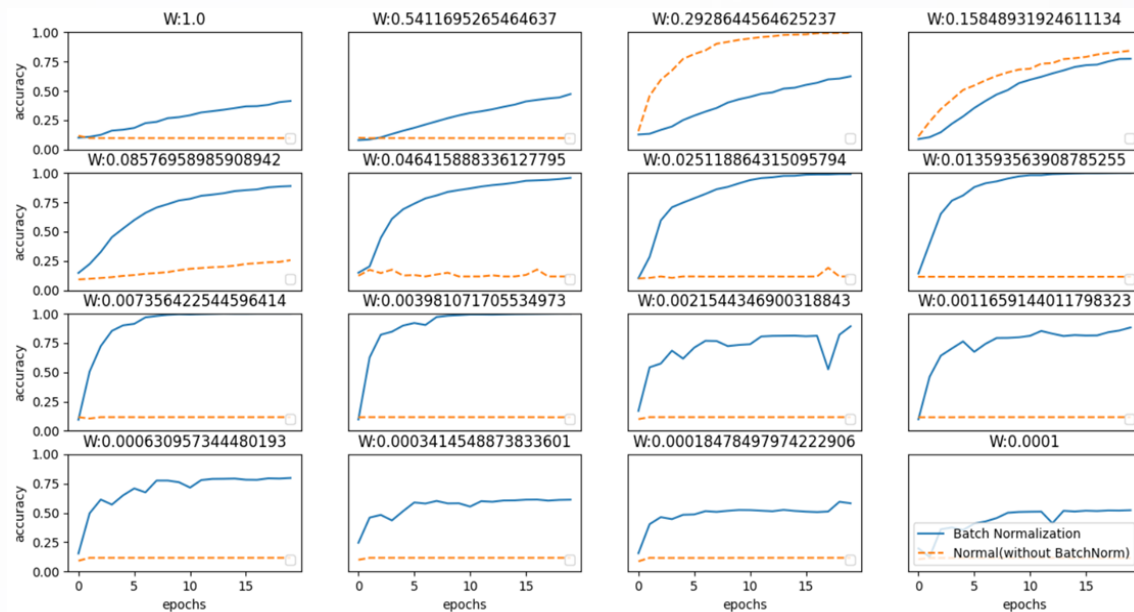
07 | 예측 모델 훈련과 선택



예측 모델 훈련과 선택

2 기계학습 라이브러리에서 제공하는 알고리즘의 기본 하이퍼파라미터(hyper parameter)가 현재 작업에 최적이라고 기대할 수 없음

→ 하이퍼파라미터는 데이터에서 학습하는 파라미터가 아닌 모델 성능을 향상하기 위해 사용자가 직접 설정해줌



실선이배치 정규화를 사용한 경우, 점선이 사용하지 않은 경우: 표준편차는 각 그래프 위에 표기



08 | 모델을 평가하고 새로운 샘플로 예측(추론)



모델을 평가하고 새로운 샘플로 예측(추론)

⌘ 훈련 데이터셋에서 최적의 모델을 선택함

◆ 테스트 데이터셋을 사용하여 이전에 본 적이 없는 데이터에서 얼마나 성능을 내는지 예측하여 일반화 오차를 예상함

◆ 이 성능에 만족한다면 이 모델을 사용하여 미래의 새로운 데이터를 예측할 수 있음

➢ 이전에 언급한 특성 스케일 조정과 차원 축소 같은 단계에서 사용한 파라미터는 훈련 데이터셋만 사용하여 얻은 것임을 주목해야 함

➢ 동일한 파라미터를 테스트 데이터셋은 물론 새로운 모든 샘플을 변환하는 데 사용함