



부산대학교

PUSAN NATIONAL UNIVERSITY

Course Goals and Structure

What a Model Learns

Models, Evaluation, and a Semester of Practice

BX3600087-140 / DS3600087-140 / AB3600087-140

기계학습개론

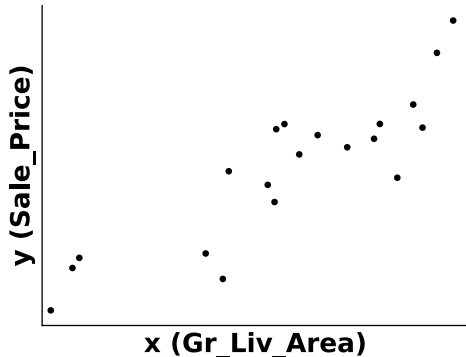
2026학년도 2학기

오용경

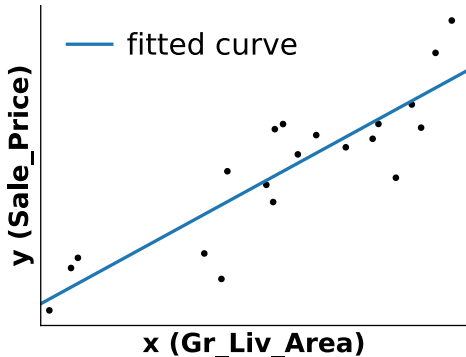
데이터사이언스전공

yongkyung.oh@pusan.ac.kr

자료에서 모델을 학습하는 과정



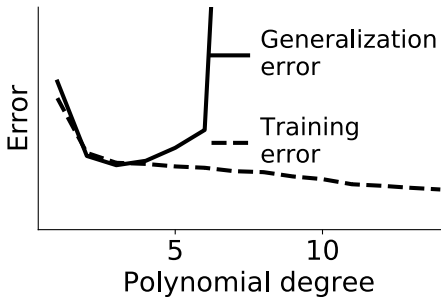
주택의 거주 면적 x 와 판매 가격 y 를 함께 관측한 학습 자료



학습 자료에서 추정된 직선 관계

과적합과 일반화

과적합(overfitting)은 학습 자료에서만 오차가 낮은 상태, 일반화(generalization)는 새 자료에서도 오차가 낮은 정도

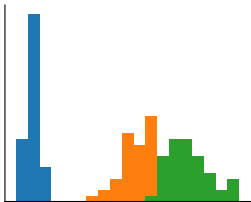


모델 복잡도에 따른 두 오차의 변화

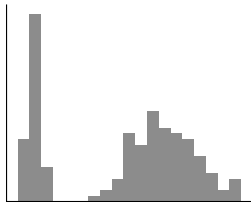
- 모델이 복잡해질수록 학습 자료의 오차는 계속 낮아짐
- 그러나 새 자료에서 잔 오차는 어느 지점부터 다시 커짐
- 복잡도는 학습에 쓰지 않은 자료에서 잔 오차로 고름

학습의 유형

품종 정보를 함께 주고 학습하면 지도학습(supervised learning), 품종 정보 없이 측정값만으로 구조를 찾으면 비지도학습(unsupervised learning)이다



품종을 색으로 구분한 꽃잎 길이. setosa를 가르는 기준이 보인다



같은 꽃잎 길이에서 품종 정보를 뺀 분포. 두 무리로 나뉜 구조가 보인다

■ setosa ■ versicolor ■ virginica

Adapted from scikit-learn MOOC (2022).

기계학습이란 무엇인가

DEFINITION

기계학습(machine learning)

자료에서 규칙을 학습해 모델로 나타내고, 새 입력의 결과를 예측하거나 관측할 수 없는 값을 추정하는 방법

ESTIMATION

잡음이 섞인 관측 자료에서 실제 값을 추정하는 문제

GENERALIZATION

한 번도 보지 못한 자료에서 결과를 예측하는 문제

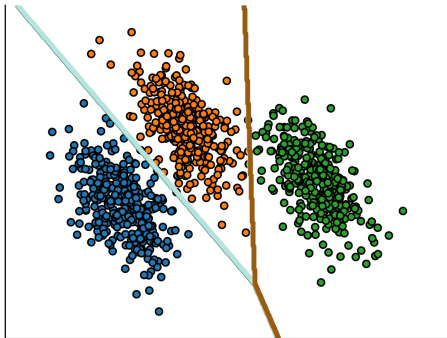
ASSUMPTION

새 자료도 학습 자료와 같은 분포에서 온다는 가정

관측한 자료를 설명하면서 새 자료에도 적용되는 모델이 필요하다

Source: MIT 6.390 (2026), Introduction.

주차	주제	내용
1	기계학습 기초	정의와 학습 유형
2	수학 기초	선형대수, 미분, 확률
3	학습과 일반화	과적합, 데이터 전처리
4	선형 모델	선형 회귀, 로지스틱 회귀, 전처리 파이프라인
5	최적화	경사하강법, 확률적 경사하강법
6	퍼셉트론	학습 규칙, 단층 구조의 한계
7	신경망	다층 퍼셉트론, 역전파
8	중간고사	1~7주
9	모델 평가	교차검증, 평가지표
10	정규화	편향-분산 분해, 불균형 자료 처리
11	의사결정나무	분할 기준, 가지치기
12	확률 모델	나이브 베이즈, 최대우도 추정
13	거리와 표현	k-최근접 이웃, 서포트 벡터 머신, 군집화, 주성분분석
14	앙상블	랜덤 포레스트, 그래디언트 부스팅
15	파이프라인과 실무	실무 이슈, 종합 정리
16	기말고사	9~15주 중심



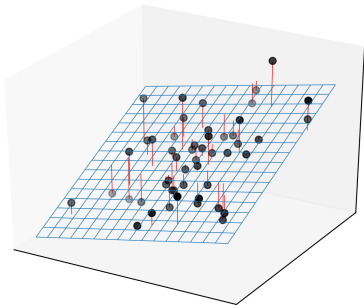
DEFINITION AND TYPES OF LEARNING

정의와 학습 유형

- 기계학습의 정의
- 지도학습과 비지도학습

예측 문제의 종류

- 회귀
- 분류



두 특성의 선형 결합이 예측값을 이루는 평면

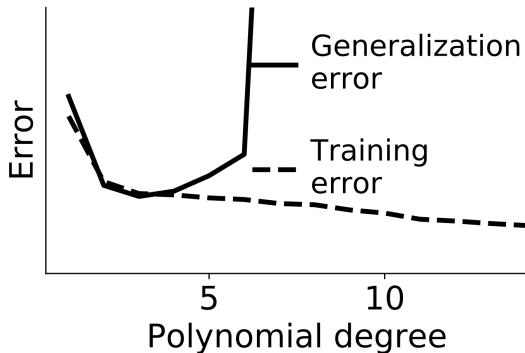
MATHEMATICAL FOUNDATIONS

선형대수

- 벡터와 행렬
- 행렬 연산

미분과 확률

- 도함수와 그래디언트
- 확률 분포



LEARNING AND GENERALIZATION

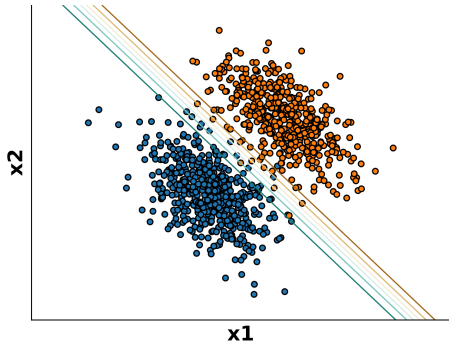
학습과 일반화

- 과적합
- 일반화 오차

데이터 전처리

- 척도 조정
- 결측과 이상치

선형 회귀와 로지스틱 회귀



로지스틱 회귀가 예측한 확률의 등고선. 선마다 확률 수준이 다르다

LINEAR MODELS

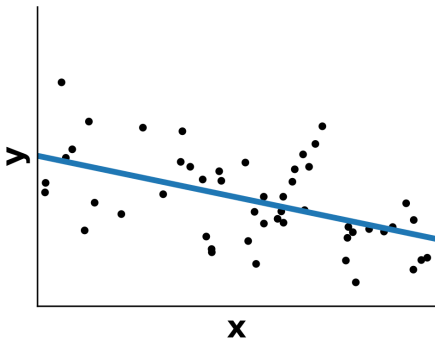
선형 모델

- 선형 회귀
- 로지스틱 회귀

전처리 파이프라인

- 변환과 학습의 연결

경사하강법과 확률적 경사하강법



GRADIENT DESCENT

경사하강법

- 그래디언트와 갱신
- 학습률

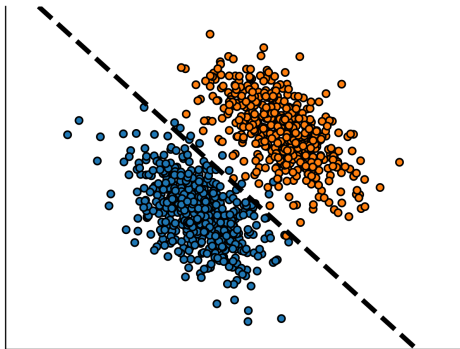
확률적 경사하강법

- 미니배치

제곱오차가 가장 작은 직선. 경사하강법은 이 직선의 계수를 반복 갱신으로 찾음

Source: scikit-learn MOOC (2022).

퍼셉트론의 학습 규칙과 한계



대부분 분리되지만 몇 점이 경계를 넘음

PERCEPTRON

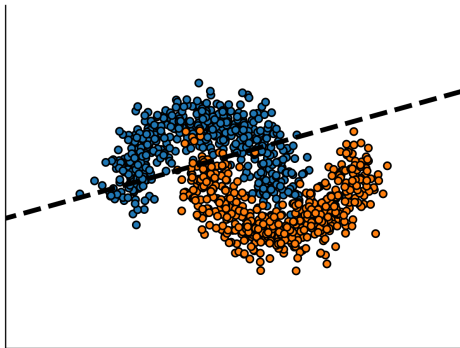
학습 규칙

- 가중치 갱신
- 종료 조건

단층 구조의 한계

- 선형 분리 가능성

Source: scikit-learn MOOC (2022).



직선 경계로는 두 범주를 나눌 수 없는 자료

MULTILAYER PERCEPTRON AND BACKPROPAGATION

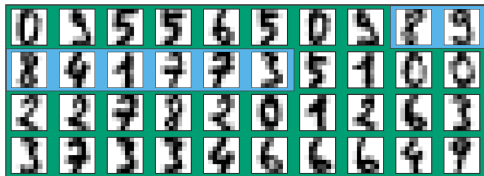
다층 퍼셉트론

- 은닉층
- 활성화함수

역전파

- 연쇄 법칙
- 그래디언트 전달

교차검증과 평가지표



Training data

Testing data

한 번의 분할에서 학습(초록)과 평가(파랑)에 쓰이는 부분

CROSS-VALIDATION AND METRICS

교차검증

- 자료 분할
- 반복 평가

평가지표

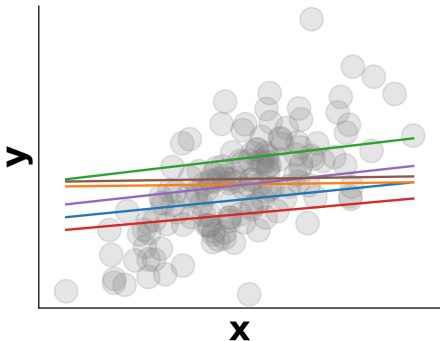
- 정확도
- 정밀도와 재현율

팀 프로젝트 안내

- 팀 구성

Source: scikit-learn MOOC (2022).

정규화와 편향-분산 분해



회색 점에서 5개씩 뽑은 학습 표본 6개에 강한 정규화로 적합한 회귀선. 기울기가 모두 작고 서로 비슷하다(큰 편향, 작은 분산)

REGULARIZATION AND BIAS-VARIANCE

정규화

- 벌점 항
- 계수 축소

편향-분산 분해

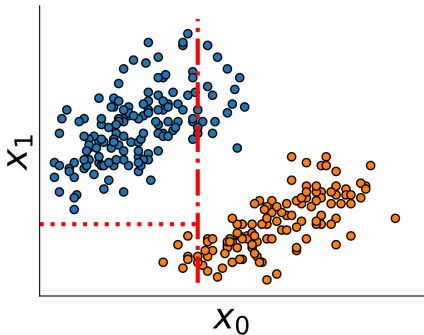
- 과소적합과 과적합

불균형 자료 처리

- 표본 조정

Source: scikit-learn MOOC (2022).

의사결정나무의 분할 기준과 가지치기



DECISION TREES

분할 기준

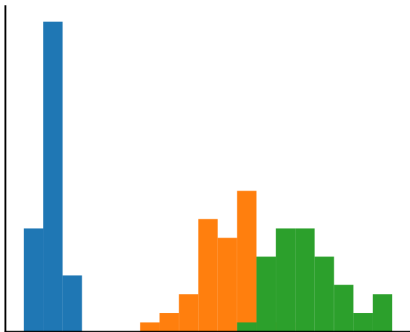
- 불순도
- 정보 이득

가지치기

- 복잡도 매개변수
- 과적합 억제

Source: scikit-learn MOOC (2022).

나이브 베이즈와 최대우도 추정



NAIVE BAYES AND MAXIMUM LIKELIHOOD

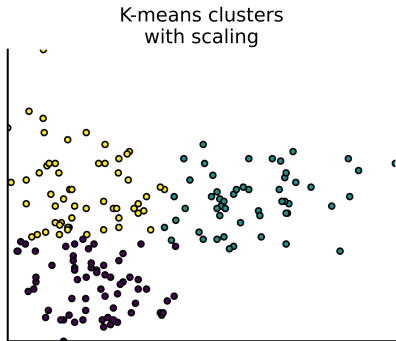
나이브 베이즈

- 조건부 독립
- 사후 확률

최대우도 추정

- 우도 함수

Source: scikit-learn MOOC (2022).



KNN, SVM, CLUSTERING, AND PCA

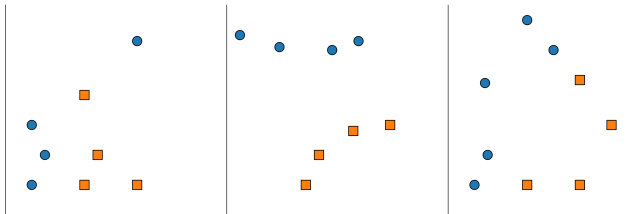
거리와 경계 기반 모델

- k-최근접 이웃
- 서포트 벡터 머신

군집화와 차원 축소

- 군집화
- 주성분분석

랜덤 포레스트와 그래디언트 부스팅



같은 자료에서 다시 뽑은 세 학습 표본

RANDOM FOREST AND GRADIENT BOOSTING

랜덤 포레스트

- 배깅
- 무작위 특성 선택

그래디언트 부스팅

- 순차적 보정

기계학습 파이프라인과 실무 이슈

전처리 학습 자료로 기준을 정하고 학습
자료와 새 자료를 같은 방식으로 변환



모델 학습 변환한 학습 자료로 적합



평가 같은 변환을 거친 새 자료로 예측하고
점수를 계산

PIPELINE AND PRACTICAL ISSUES

기계학습 파이프라인

- 전처리부터 평가까지

실무 이슈

- 자료 수집
- 모델 유지

종합 정리

- 한 학기 되짚기

평가 및 과제

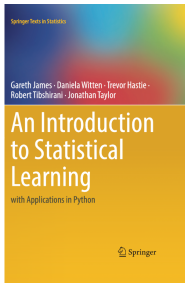
항목	비중	내용
중간고사	30%	기초와 일반화, 선형 모델, 경사하강법, 신경망 등
기말고사	30%	평가·정규화, 트리, 확률, 거리와 표현, 앙상블 등
과제	20%	Python 프로그래밍 과제
팀 프로젝트	20%	발표와 최종 산출물
수업 참여도	+10%	별도 반영

EXAMS

- 중간고사: 1~7주 중심
- 기말고사: 9~15주 중심

ASSIGNMENTS AND PROJECT

- 과제: 2~3주에 1회 정도
- 팀 프로젝트: 학기 후반에 구성하고 발표



MAIN TEXTBOOK

『기초부터 다지는 통계학 교과서 with 파이썬』

개러스 제임스 외 지음, 송영숙 · 유현조 옮김, 인사이트, 2024
product.kyobobook.co.kr

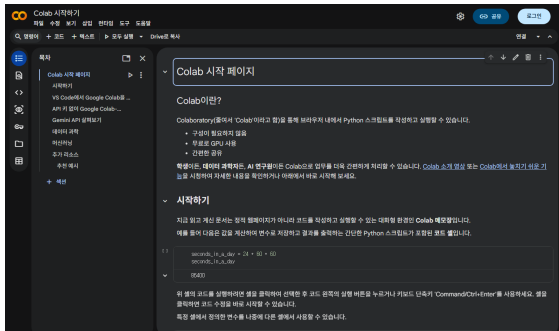
SUPPLEMENTARY TEXTBOOKS

An Introduction to Statistical Learning with Applications in Python

주교재의 원서. statlearning.com에서 무료 공개

Dive into Deep Learning

d2l.ai에서 무료 공개



제1주에 실습 환경(Google Colab)을 준비한다

Source: Google Colab (n.d.), Welcome to Colaboratory.

What Does It Mean to Estimate a Function from Data?

Next Lecture: Learning from Data

- 입력과 출력으로 보는 학습 문제
- 지도학습과 비지도학습
- 회귀 문제와 분류 문제