



인공신경망과딥러닝입문

Lecture 06. 머신러닝 프로세스

동덕여자대학교
데이터사이언스 전공
권 범

목차

❖ 01. 머신러닝 프로세스 이해하기

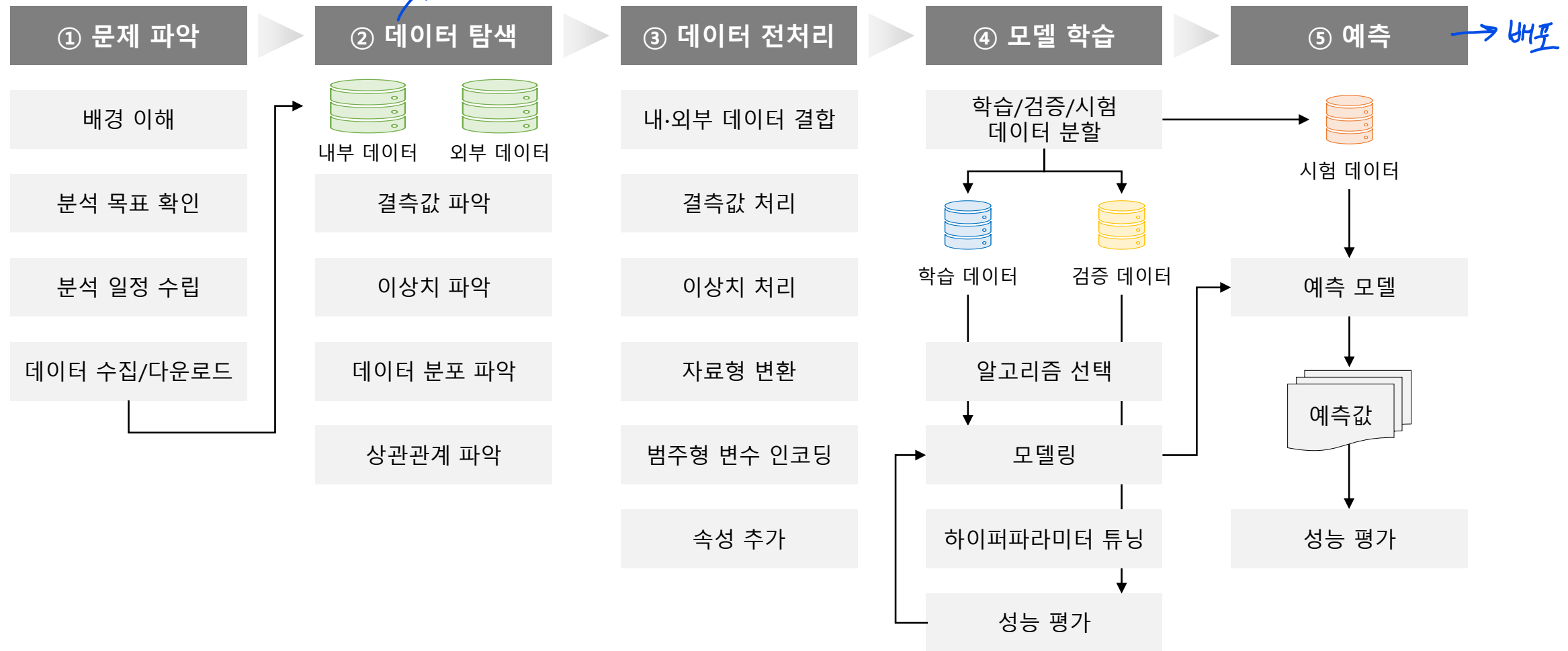
❖ 02. 특징 추출이란?

01. 머신러닝 프로세스 이해하기

02. 특징 추출이란?

01. 머신러닝 프로세스 이해하기

❖ 머신러닝 프로세스 (1/6) *시각화/통계적 분석*



01. 머신러닝 프로세스 이해하기

❖ 머신러닝 프로세스 (2/6)

① 문제 파악

- ◆ 주어진 문제를 파악함
- ◆ 해당 분야의 **도메인 지식**(Domain Knowledge)을 습득하는 것이 좋음
- ◆ 데이터 분석의 목표를 설정함
- ◆ 데이터 분석 방법과 일정을 수립함
- ◆ 필요한 (내부 또는 외부) 데이터를 **수집** 또는 **다운로드** 함
 - 내부 데이터: 조직/집단 내에서 자체적으로 보유하고 있는 데이터
 - 외부 데이터: **공공데이터포털** 등 외부에서 수집하는 데이터

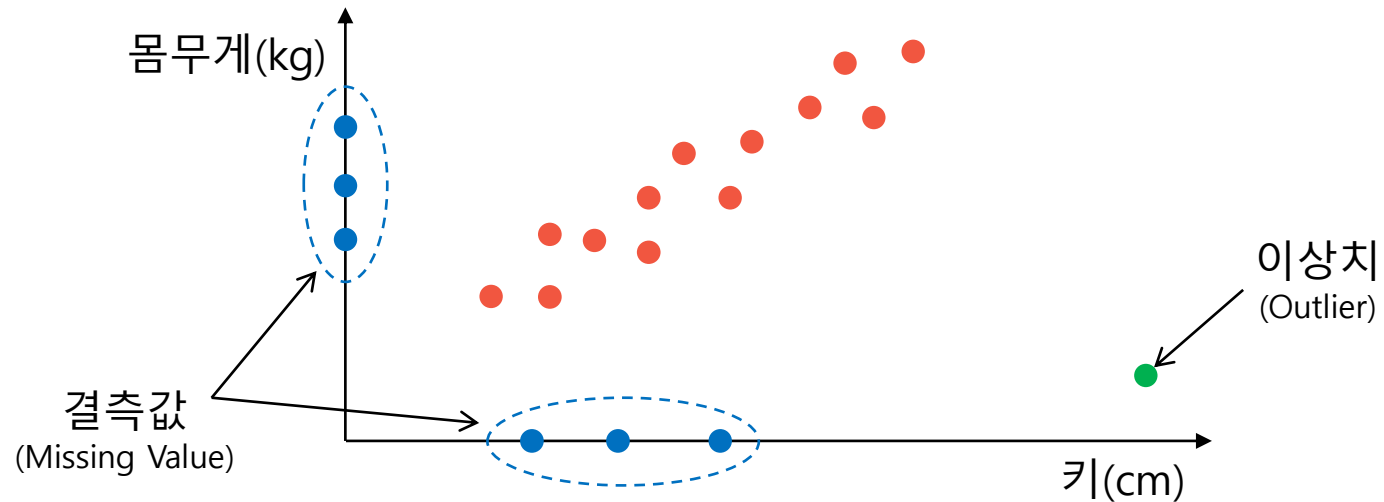
캐글 (kaggle)

01. 머신러닝 프로세스 이해하기

❖ 머신러닝 프로세스 (3/6)

② 데이터 탐색 → 시각화

- ◆ 데이터를 읽어 들이고, 데이터의 이상 유무를 확인함
- ◆ 데이터가 누락된 결측값(Missing Value)이 있는지 확인함
- ◆ 데이터가 정상 범위를 벗어난 이상치(Outlier)가 있는지 확인함
- ◆ 데이터 구조 및 특성을 파악하고 데이터의 분포와 상관관계를 파악함



01. 머신러닝 프로세스 이해하기

❖ 머신러닝 프로세스 (4/6)

③ 데이터 전처리 *pre-processing*

- ◆ 모델 학습이 가능한 형태로 데이터를 정리하는 단계
- ◆ 내부 데이터와 외부 데이터를 병합함
- ◆ 데이터 탐색 단계에서 확인 **결측값과 이상치를 처리**함
- ◆ 필요한 경우 **자료형을 변환**함 *주관식/단답형 = String → 숫자*
- ◆ 머신러닝 및 딥러닝 모델은 숫자형 데이터를 입력으로 받기 때문에
문자열(범주형) 데이터를 숫자형으로 변환하는 작업이 필요함
- ◆ 새로운 속성을 추가하거나 불필요한 속성을 제거하기도 함

강아지와 고양이 분류 문제



레이블: "강아지"

↓
0



레이블: "고양이"

↓
1

문자열

숫자형

01. 머신러닝 프로세스 이해하기

❖ 머신러닝 프로세스 (5/6)

④ 모델 학습

◆ 전처리가 끝난 데이터셋(Dataset)을 ① 학습, ② 검증, ③ 시험 데이터셋으로 나눔

- ✓ 학습 데이터셋(Train Dataset): 모델 학습에 사용되는 데이터
- ✓ 검증 데이터셋(Validation Dataset): 모델이 학습 데이터셋에 과적합(Overfitting)되는 것을 방지하기 위해서 학습의 조기 종료(Early Stopping)를 위해 사용됨
- ✓ 시험 데이터셋(Test Dataset): 선택된 최종 모델의 성능 평가에 사용되는 데이터

◆ 머신러닝 알고리즘을 선택하고 모델을 설계함

◆ 학습 데이터셋을 이용하여 모델을 학습시킴

◆ 검증 데이터셋을 이용하여 학습을 마친 모델의 성능을 평가함

◆ 모델의 성능을 높일 수 있도록 하이퍼파라미터(Hyperparameter)를 튜닝함

◆ 최종 모델을 선택함

아래와 같은 표현을 많이 사용합니다.
“학습 데이터를 이용하여 모델을 학습(Learning)시킨다.”
“모델을 학습 데이터에 피팅(Fitting)시킨다.”

→ Fit()

01. 머신러닝 프로세스 이해하기

❖ 머신러닝 프로세스 (6/6)

⑤ 예측

- ◆ 예측해야 하는 시험 데이터셋을 모델에 입력하고, 모델이 예측한 값을 형식에 맞게 정리함
- ◆ 시험 데이터셋의 레이블(Label) 값과 모델이 예측한 값을 비교하여 모델의 성능을 평가함

↓
결과 좋으면 배포

FE

02. 특징 추출이란?

Feature Extraction

01. 머신러닝 프로세스 이해하기

02. 특징 추출이란?

❖ 학습 및 시험 단계 살펴보기 (x, y)

① 학습 단계(Training Phase) <지도 학습>

학습 데이터셋

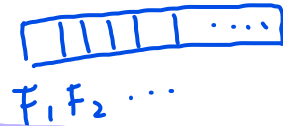


레이블(Label) y

Raw Data x
(숫자, 텍스트, 이미지 등)
동영상

특징 추출
(Feature Extraction)

Feature Vector

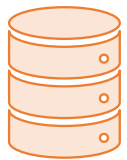


머신러닝
모델

학습(Training, Learning)
or
피팅(Fitting)
 y_1, y_2, \dots

② 시험 단계(Test Phase)

시험 데이터셋



Raw Data x
(숫자, 텍스트, 이미지 등)

특징 추출
(Feature Extraction)

Feature Vector

머신러닝
모델

예측값 \hat{y}

학습이 완료된
모델

Mean Squared Error

$$\frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|^2 = \text{오차 error}$$

Mean Absolute Error (=MAE)

02. 특징 추출이란?

❖ 특징 추출 (1/4)

상상해 봅시다!
아래 이미지들을 보고 연상되는 동물은?



꼬리



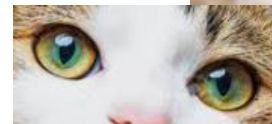
귀



발바닥



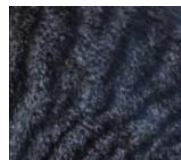
코



눈



수염



무늬

02. 특징 추출이란?

❖ 특징 추출 (2/4)

- 사람은 어떻게 고양이의 신체 일부분만 보고도, 고양이라는 것을 추론해 낼 수 있을까?



사람은 지금까지 살아오면서 다양한 고양이를 보왔고,
고양이가 갖고 있는 특징들을 이미 학습을 통해 알고 있음

고양이가 갖고 있는 몇가지 특징들

특징 #1: 뽕족한 귀



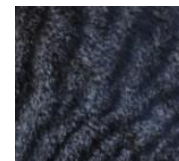
특징 #2: 긴 꼬리



특징 #3: 핑크색 발바닥



특징 #4: 무늬(삼색, 치즈색, 고등어색)

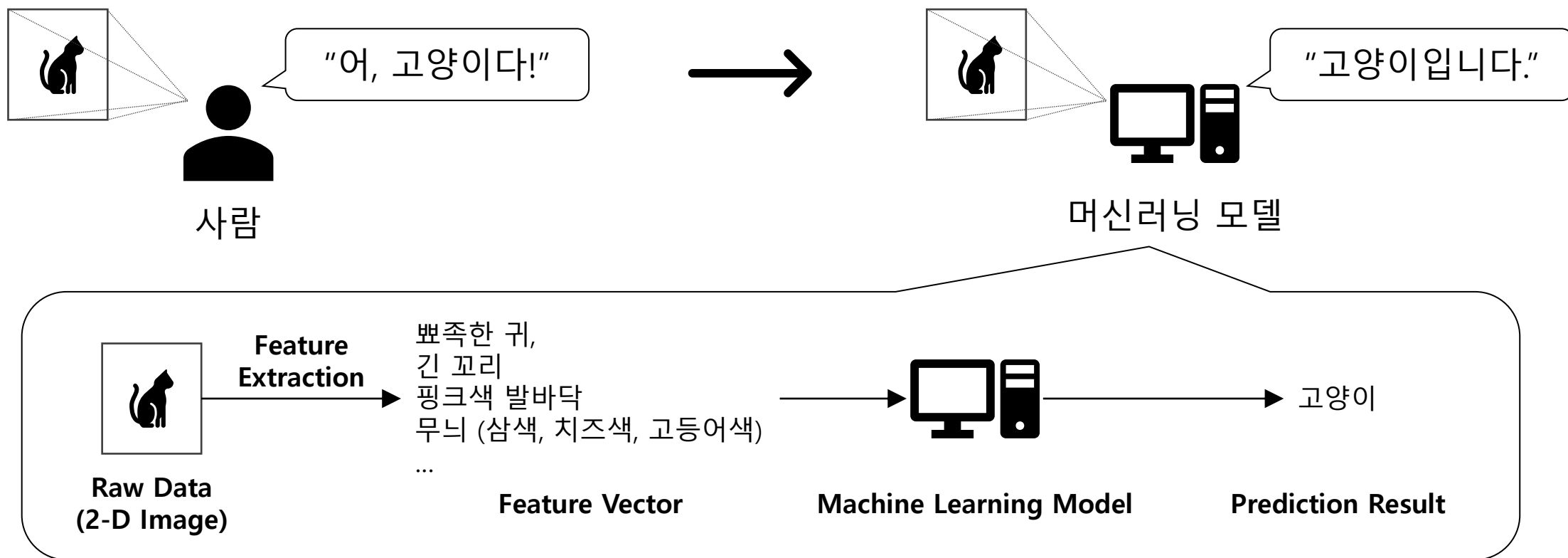


02. 특징 추출이란?

❖ 특징 추출 (3/4)

- 그렇다면, 머신러닝 모델이 주어진 이미지를 보고 고양이로 인지 할 수 있게 하려면 어떻게 해야 할까?

◆ 특징 추출: 머신러닝 모델이 잘 예측할 수 있도록 Raw Data로부터 특징이 되는 정보를 뽑아내는 작업

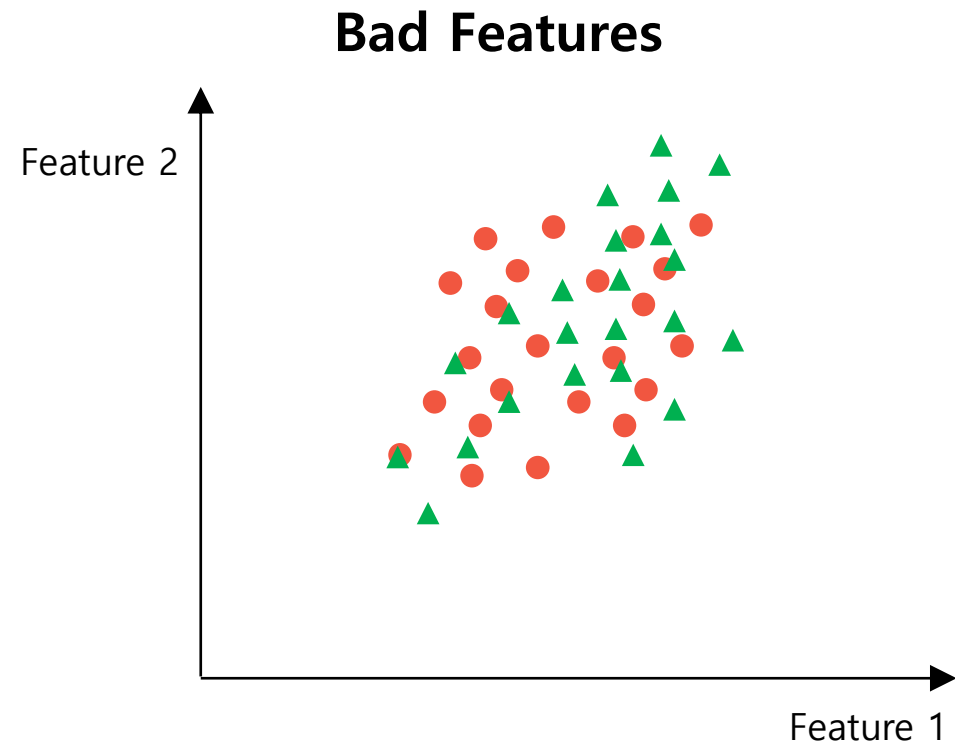


02. 특징 추출이란?

❖ 특징 추출 (4/4)

- Good and Bad Features

좋은 Features를 설계(Design)하는 것이 매우 중요함!



❖ 01. 머신러닝 프로세스 이해하기

❖ 02. 특징 추출이란?

THANK YOU!

Q & A

- Name: 권범
- Office: 동덕여자대학교 인문관 B821호
- Phone: 02-940-4752
- E-mail: bkwon@dongduk.ac.kr