

2-5

심화로 향하는 인공지능 개념완성

: 머신러닝 구현 순서





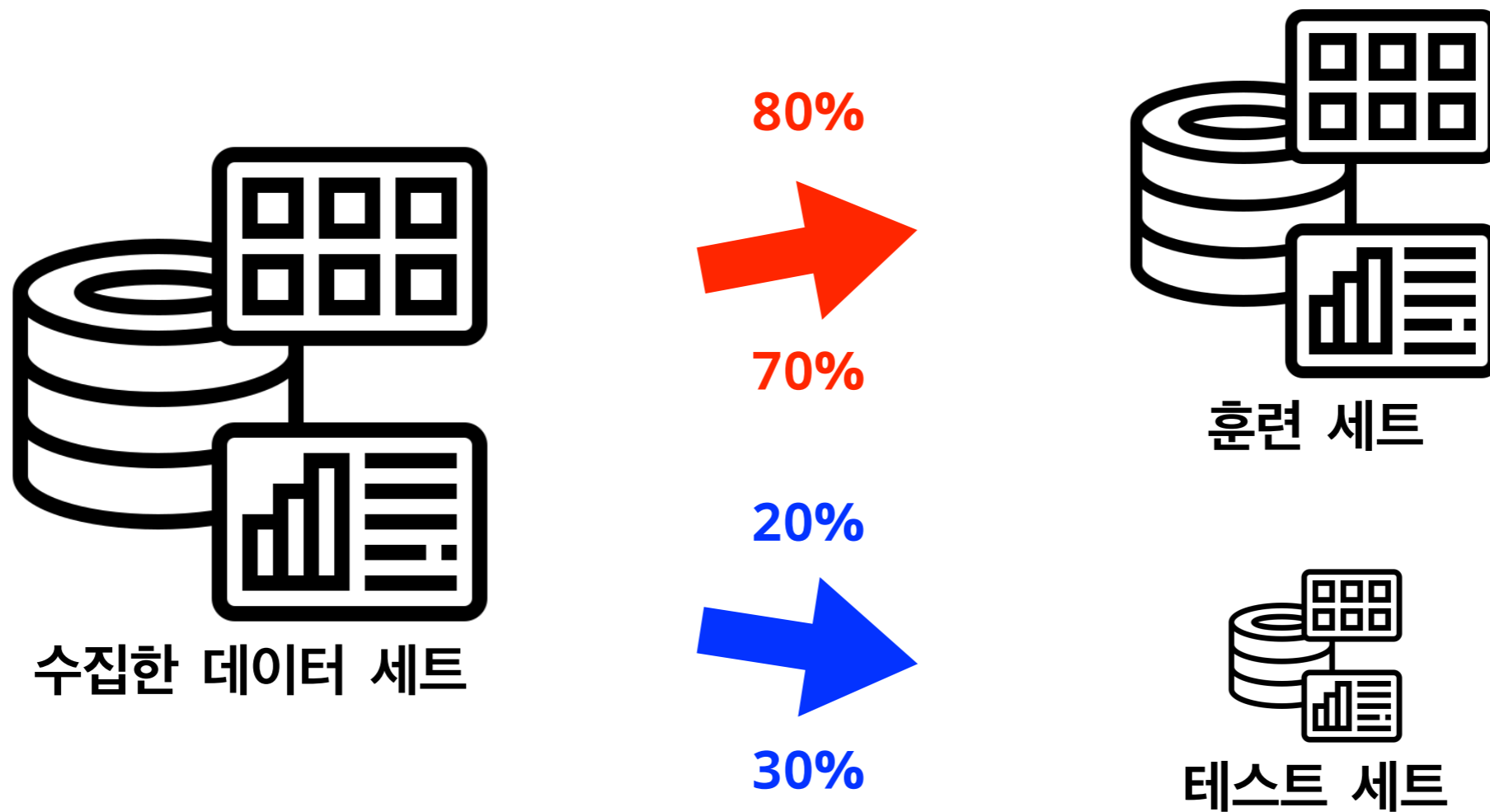
머신러닝 구현을 위한 순서

1. 문제 정의
2. 데이터 수집
3. 데이터 탐색
 - 훈련/테스트 세트 분리
 - 데이터 이해/시각화
 - 데이터 전처리
4. 모델링 및 성능 평가
5. 모델 활용

데이터 탐색(훈련/테스트 세트 분리)

: 80-20 혹은 70-30

머신러닝의 성능 평가를 위해 훈련 데이터와 테스트 데이터 분리하기



데이터 탐색

: 교차 검증(Cross Validation)

하나의 훈련 데이터로만 학습시키면 발생하는 과적합을 제거할 때 사용

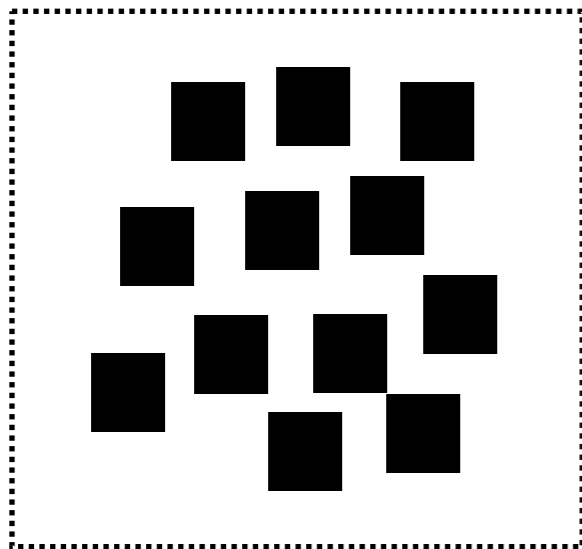


데이터 탐색(훈련/테스트 세트 분리)

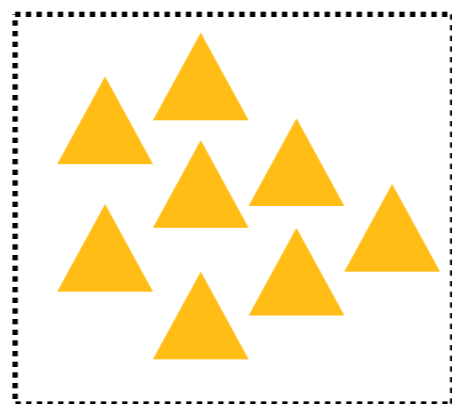
: 샘플링 편향 제거

네모와 세모를 분류하는 데이터 세트 만들기

A방법

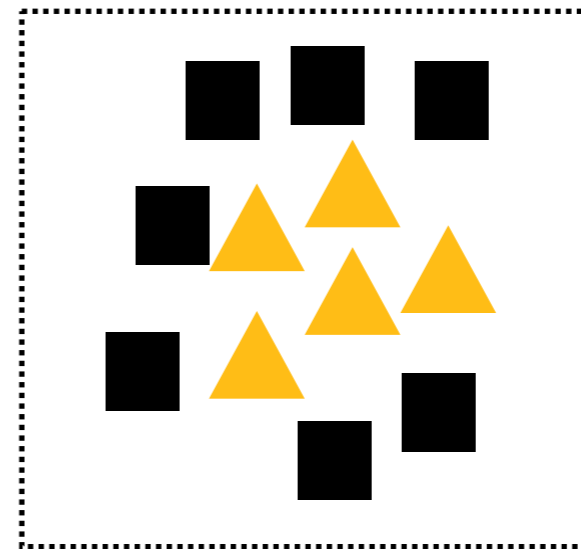


훈련 세트

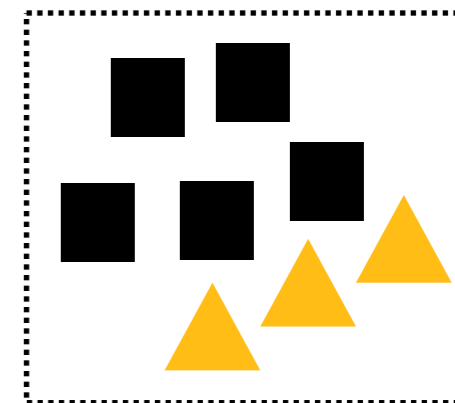


테스트 세트

B 방법



훈련 세트

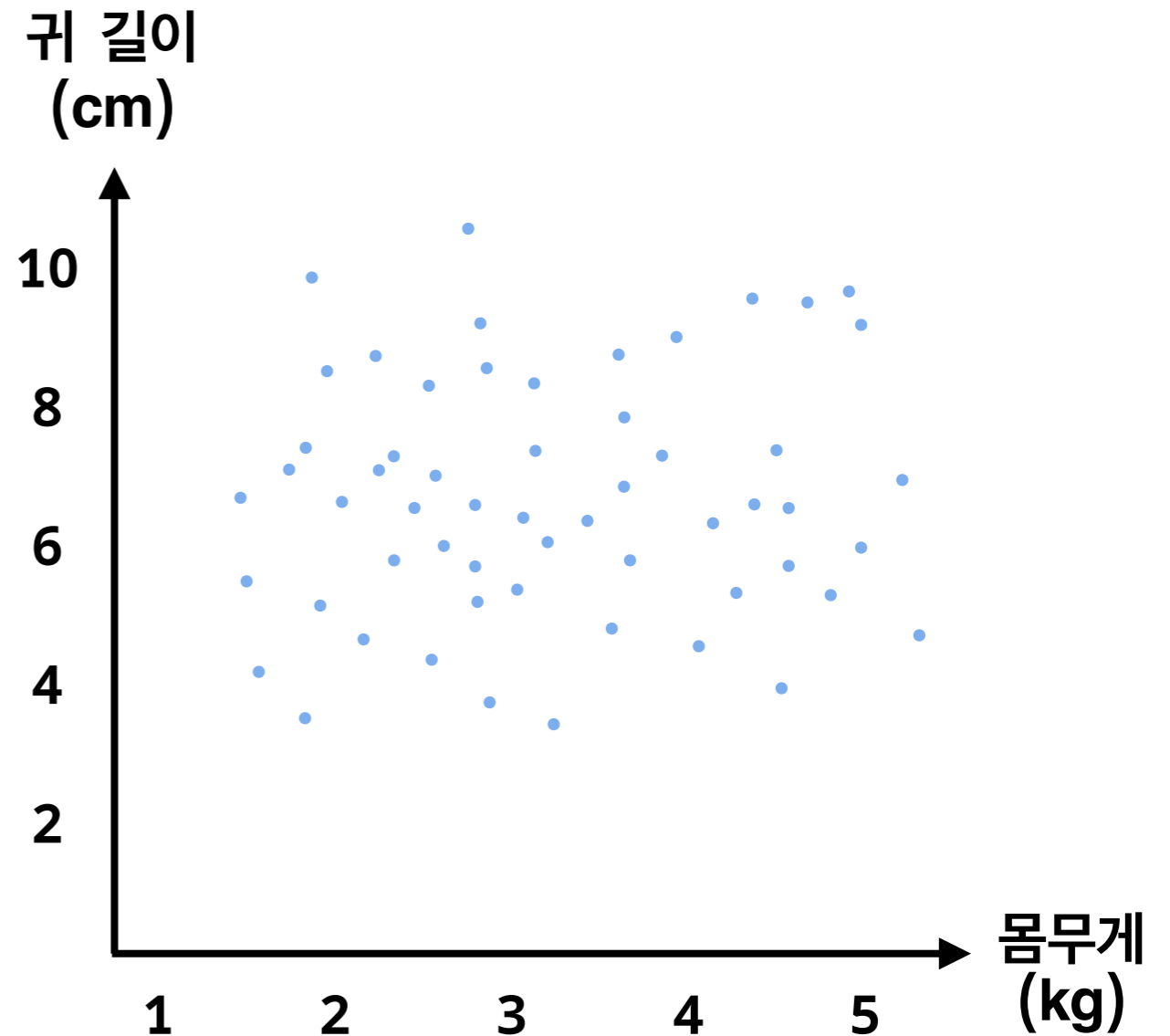
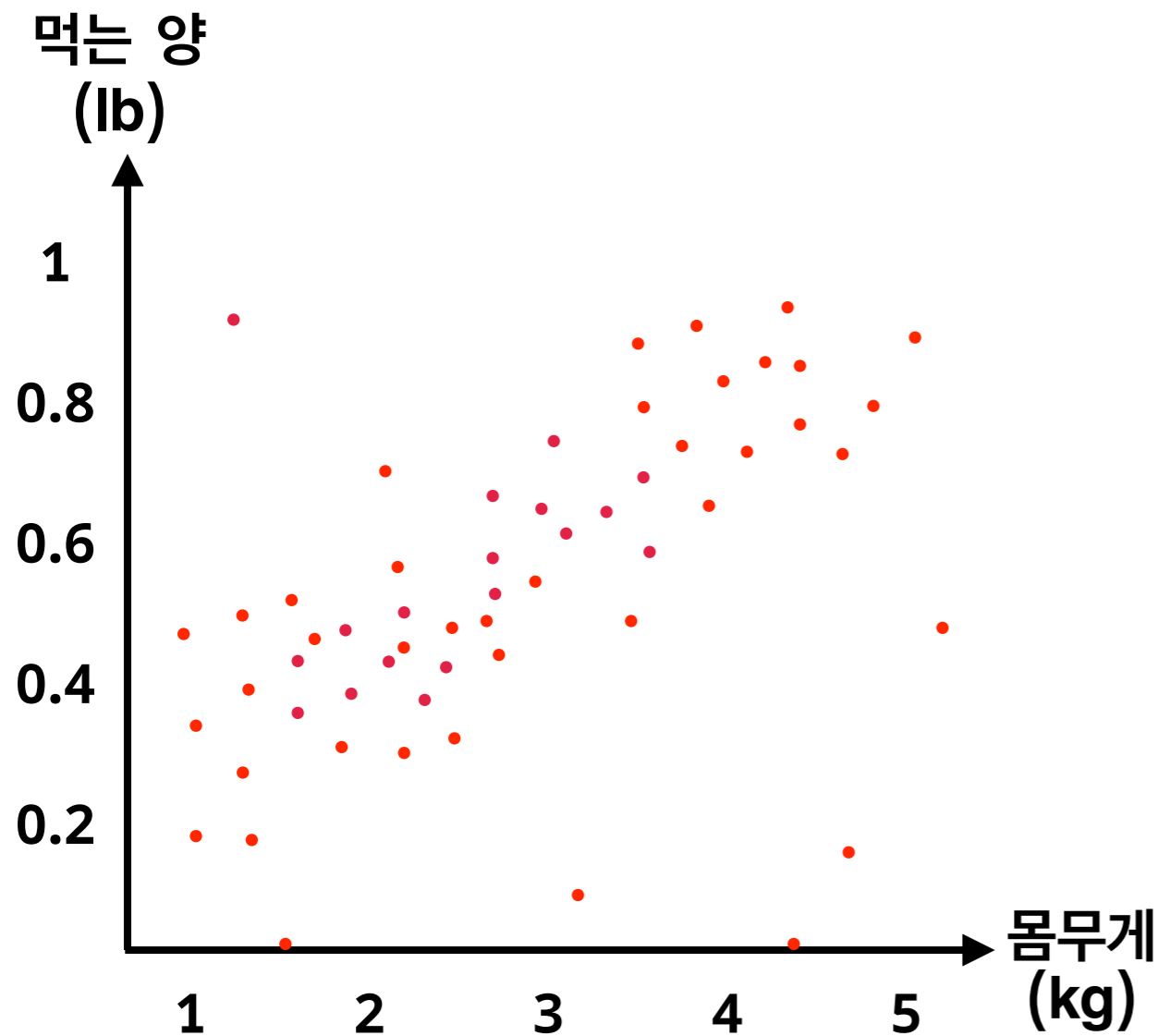


테스트 세트

데이터를 이해하는 가장 쉬운 방법

: 데이터 시각화

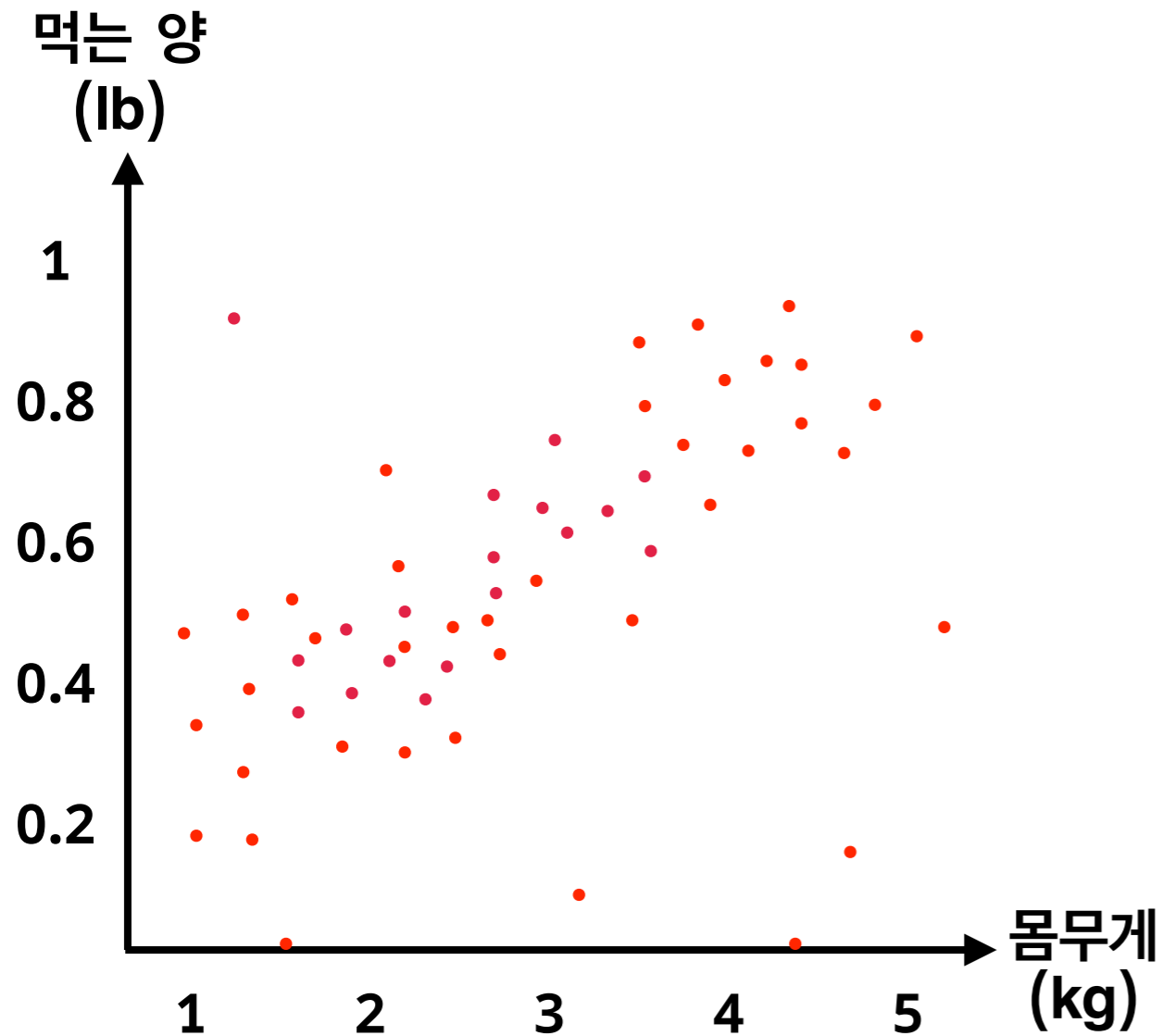
데이터 특성간의 관계를 눈으로 확인하기



학습을 효과적으로 하기 위한 방법

: 데이터 전처리(이상치/결측치 처리)

이상치와 결측치를 제거하거나 채우기

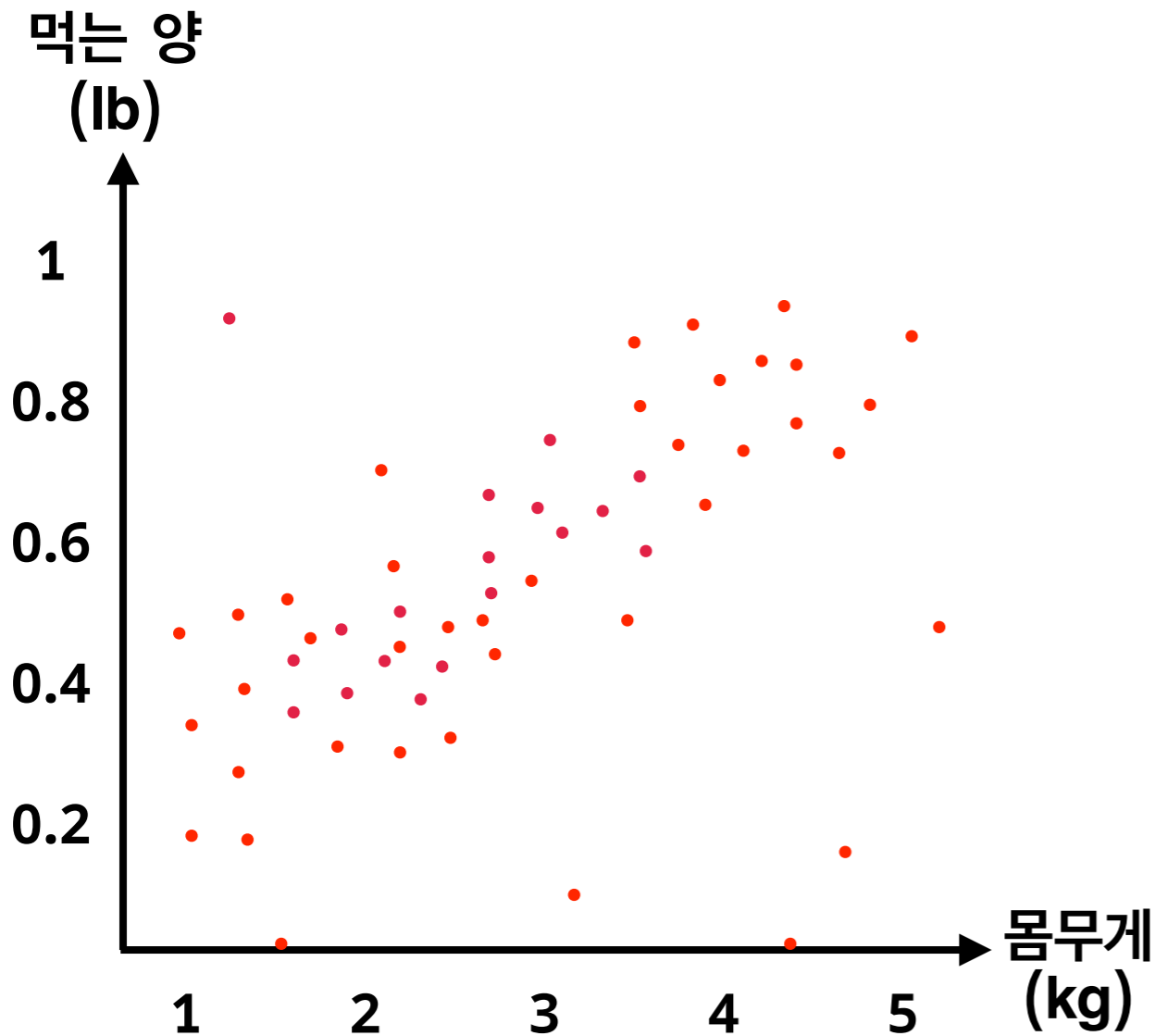


번호	무게	길이	성별	먹는양	색깔
1	3.6	39	-	0.5	흰
2	5	48	-	-	검
3	2.1	-	-	0.97	갈
4	4.4	43	남	0.2	갈
5	2.9	28	여	0.1	흰
6	-	50	여	-	흰
7	-	-	남	0.4	흰
8	-	34	여	0.5	검
...					
10000	1.6	20	여	0.33	검

학습을 효과적으로 하기 위한 방법

: 데이터 전처리(정규화 하기)

데이터 간의 기준을 정해주기

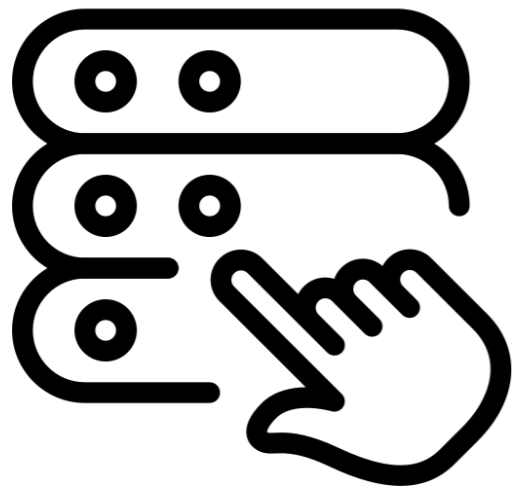


몸무게와 먹는 양을 측정한 스케일이 다름

=> 몸무게는 kg, 먹는 양은 lb(파운드)
=> 일정한 기준으로 만들어야 함

본격적으로 머신을 학습시켜볼까?

: 머신러닝 훈련 및 성능 평가



머신러닝 모델 선택



훈련 세트로 학습



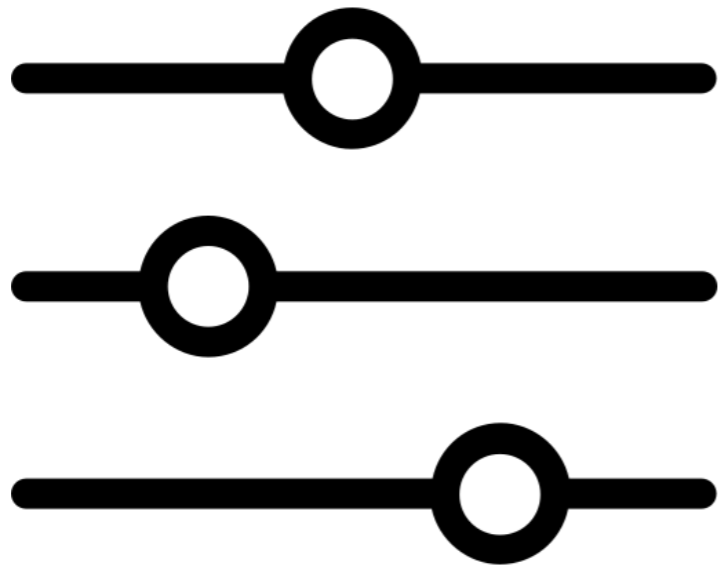
테스트 세트로 검증



머신의 정확도를 높여보자!

: 머신러닝 조정

파라미터 조정 및 모델 재선택



각 머신러닝 모델에 사용되는 파라미터
학습을 몇번 시키는가
학습률