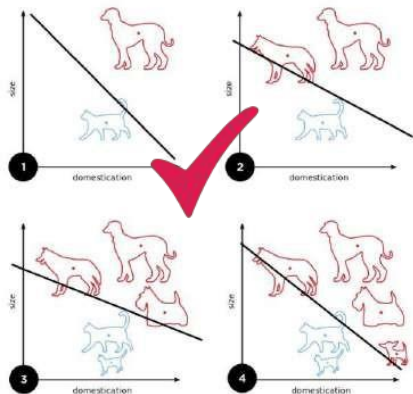
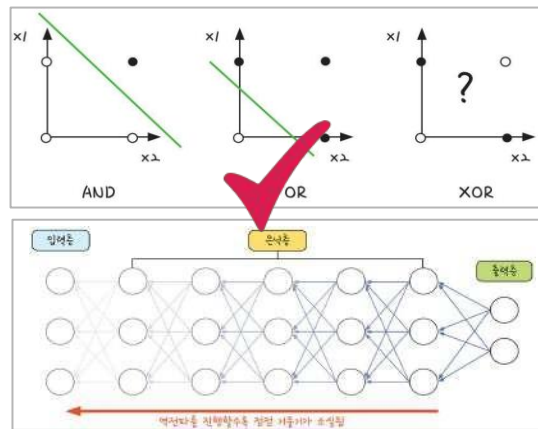

3교시: 모델 개발의 실제

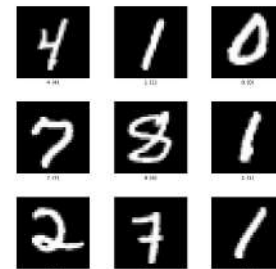
머신러닝 개념 이해



다층 퍼셉트론, 신경망



CNN / 전이학습

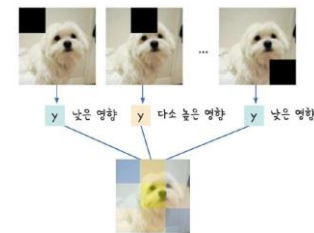


딥러닝

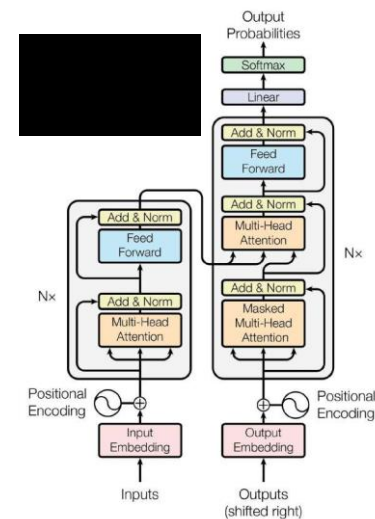
GAN/AE



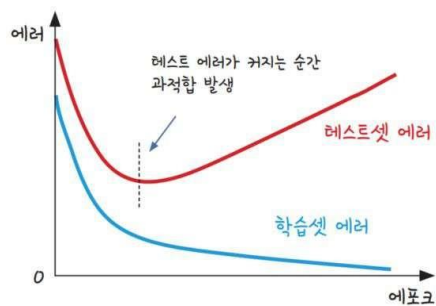
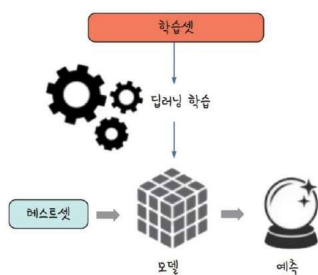
XAI



Transformer



모델링의 기초



3교시 : 모델 개발의 실제

01

분류를 위한 기초



<실습> 다중 분류 방법

02

과적합의 이해 및 모델 검증의 방법



<실습> 모델 검증 전략

03

모델의 성능을 향상시키는 법



<실습> 성능 향상 기법

04

실제 데이터로 모델 만들어 보기



<실습> 실데이터 적용 사례

분류를 위한 기초

다중 분류 문제

- 세 개 이상의 클래스(범주)로 데이터를 분류하는 문제

아이리스 꽃 분류



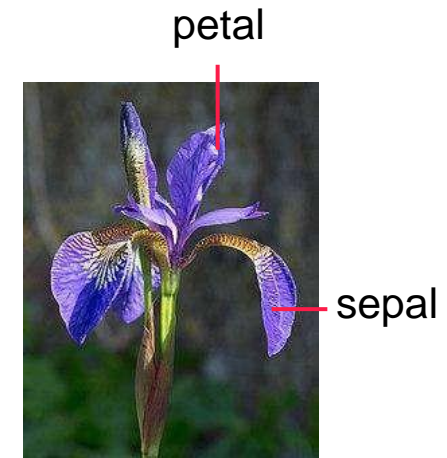
Iris-virginica



Iris-setosa



Iris-versicolor



- 아이리스 꽃 세 종류의 특징을 기록한 데이터셋
- 꽃받침 길이 (sepal_length), 꽃받침 너비 (sepal_width), 꽃잎 길이 (petal_length), 꽃잎 너비 (petal_width) 등 네 가지 특징 포함

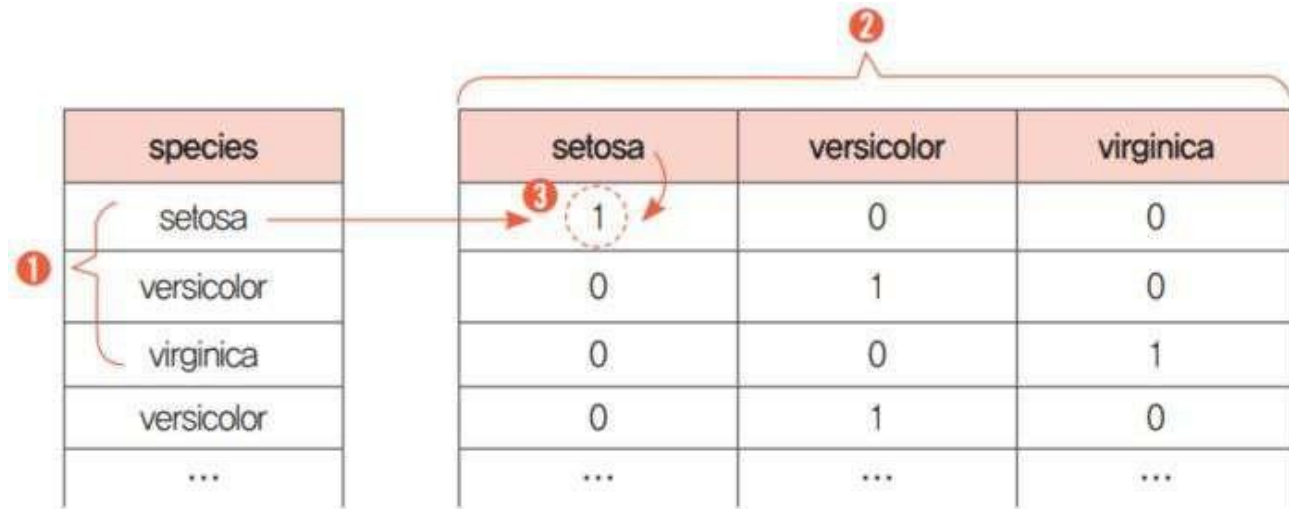
원-핫 인코더

컴퓨터가 카테고리를 숫자로 바꿔서 이해하도록 만드는 방법

원리


각 카테고리를 숫자의 리스트로 바꿈

해당 카테고리의 위치에 1을 넣고, 나머지 위치에는 0을 넣음



소프트맥스

- 여러 숫자를 0과 1 사이의 값으로 변환하고, 모든 값의 합이 1이 되도록 만드는 함수
- 여러 클래스 중 하나를 선택하는 문제에서 확률을 계산하기 위해 사용
- 결과가 확률로 표현되므로 어떤 클래스가 더 가능성이 높은지 명확히 알 수 있음

샘플		setosa일 확률	versicolor일 확률	virginica일 확률
1번 샘플	예측 실행 	0.2	0.7	0.1
2번 샘플		0.8	0.1	0.1
3번 샘플		0.2	0.2	0.6



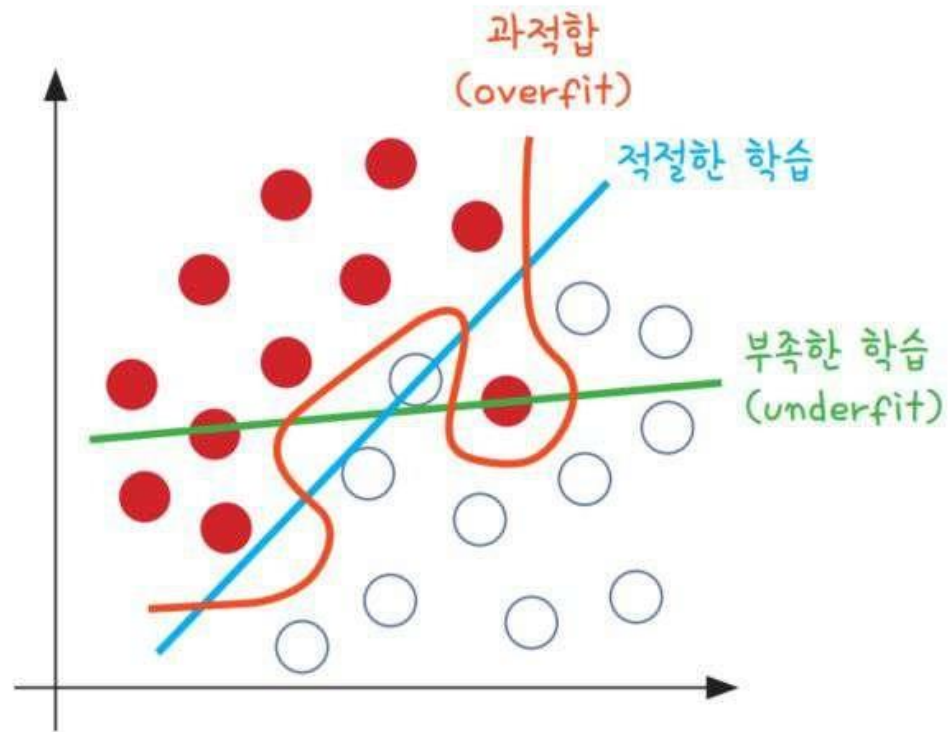
실습: 다중 분류 방법

https://github.com/taehojo/fastcampus_ai

과적합의 이해 및 모델 검증의 방법

과적합

- 모델이 훈련 데이터에 너무 잘 맞아서 새로운 데이터에는 잘 작동하지 않는 상태

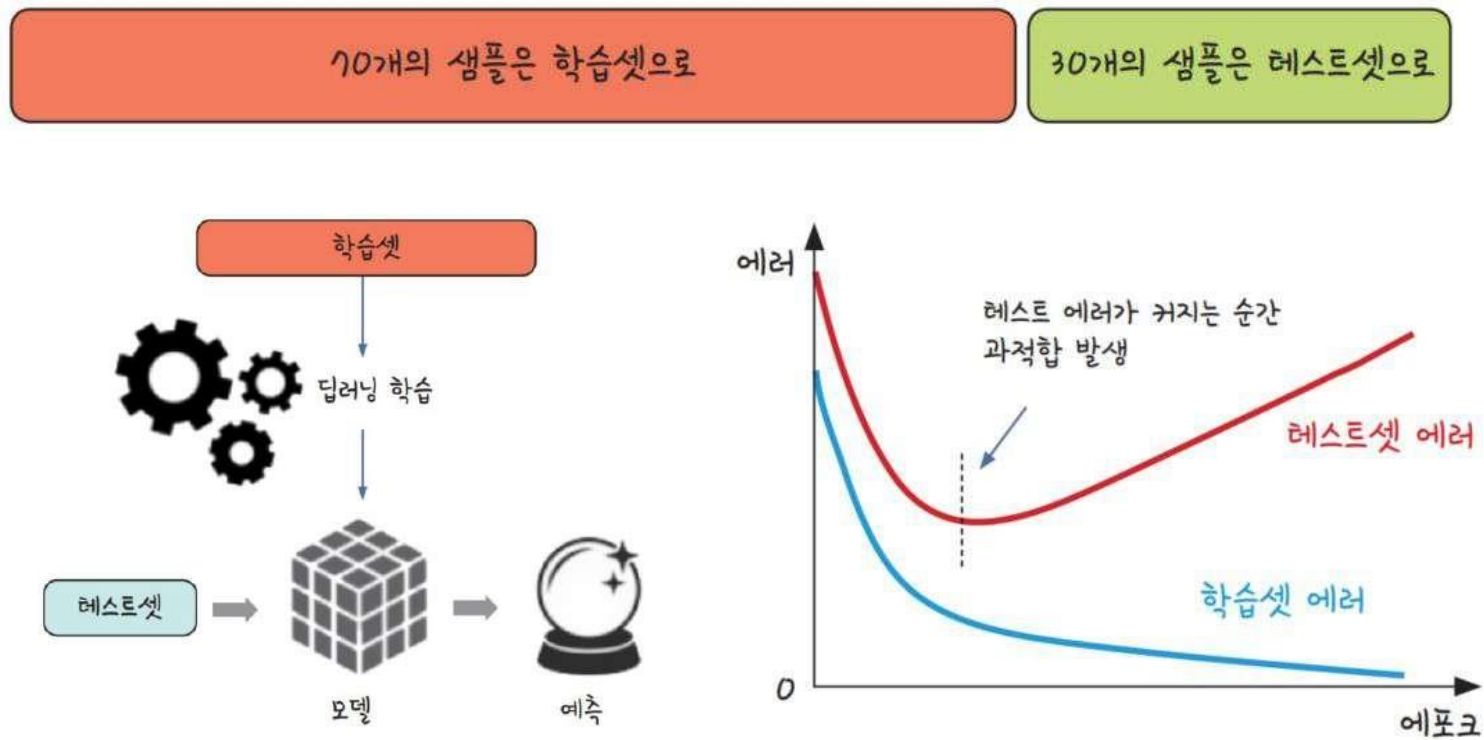


이유:

- 훈련 데이터가 너무 적을 때:
 - 데이터를 외워버림
- 훈련 시간이 너무 길 때:
 - 데이터를 지나치게 학습함

학습셋, 테스트셋

- 모델의 학습과 테스트를 따로 준비



교차 검증

- 데이터를 여러 부분으로 나누어 모델을 학습하고 평가하는 방법





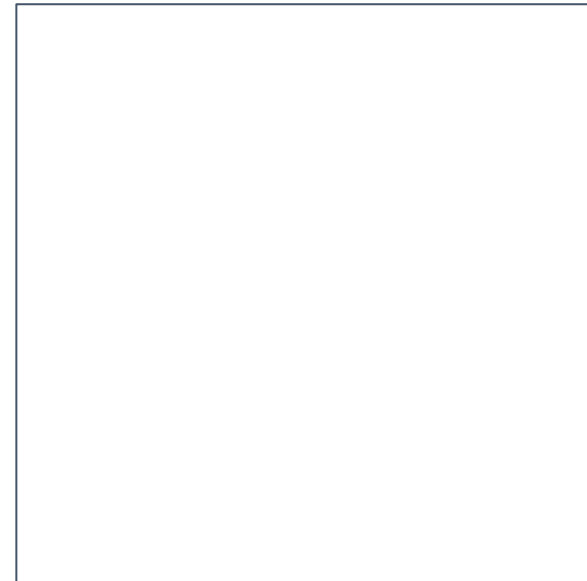
실습: 모델 검증 전략

https://github.com/taehojo/fastcampus_ai

모델의 성능을 향상시키는 법

모델의 성능을 향상시키기 위해 필요한 것들

- 검증: 모델이 잘 작동하는지 확인하는 과정
- 기존 모델의 업데이트: 새로운 데이터나 기술로 모델을 개선
- 눈으로 확인 (그래프로 확인): 그래프를 사용해 모델의 성능을 시각적으로 평가
- 자동 중단: 모델이 더 이상 성능이 개선되지 않을 때 학습을 중단하는 기능





실습: 성능 향상 기법

https://github.com/taehojo/fastcampus_ai

실제 데이터로 모델 만들어 보기

실제 데이터로 프로젝트를 진행할 때의 주요 과제

- 결측치 처리하기: 데이터에서 누락된 값을 찾아 적절히 대체하거나 제거하는 과정
- 관련된 속성 파악하기: 데이터의 중요한 특징과 속성을 이해하고 분석하는 과정
- 검증 가능한 모델 만들기: 모델의 성능을 신뢰할 수 있게 평가하고 검증하는 과정



실제 데이터로 프로젝트를 진행할 때의 주요 과제

1. 데이터 준비 및 전처리	데이터 불러오기	데이터 분석 준비
	결측치 처리	데이터 완전성 확보
	상관관계 분석 및 중요한 속성 선택	영향을 크게 미치는 변수 파악
2. 모델 설계, 학습 및 평가	모델 정의, 컴파일, 학습, 조기 종료 설정	모델 성능 최적화 및 과적합 방지
	성능 평가, 예측 값 비교	모델 성능 확인 및 예측 능력 검토



실습: 실데이터 적용 사례

https://github.com/taehojo/fastcampus_ai