



# 인공지능 수업 4주차 핵심 요약

## 1. 데이터셋이 제일 중요!

- 인공지능의 성능은 얼마나 다양하고 충분한 데이터를 갖고 있느냐에 달려 있어.
- 예: "맥도날드"를 전 세계 사람들 발음으로 인식하려면 방언, 억양, 언어 다 포함된 방대한 데이터 필요!
- 데이터 편향 조심!
  - 예: 10만 개 중 9만 개가 A라면, 모델은 무조건 A만 예측하는 편향 생김.

## 2. 기계학습(Machine Learning) 기본 구조

- 데이터 구성:
  - 특징 벡터: 꽃잎 길이, 무게, 색깔 같은 수치들 → `x`
  - 레이블: "상" 사과, "중" 사과, "하" 사과 같은 정답 → `y`
- 예시 데이터: 아이리스(Iris) 꽃 데이터셋
  - 4개의 특징 (꽃잎/꽃받침의 길이 & 너비)
  - 3개의 꽃 종류 (Setosa, Versicolor, Virginica)

## 3. Supervised vs Unsupervised vs 딥러닝

종류	특징
규칙 기반 학습	사람이 직접 규칙 코딩
기계학습 (ML)	사람이 데이터에 레이블 붙여줌, 특징도 사람이 정의
딥러닝 (DL)	레이블만 주고, 나머지는 기계가 알아서 학습 (특징 추출도 자동)

## 4. 원-핫 인코딩 (One-hot Encoding)

- 클래스(A, B, C)를 0과 1로 표현
  - A → `[1, 0, 0]`, B → `[0, 1, 0]`, C → `[0, 0, 1]`
- 장점: 곱셈 연산으로 간단하게 클래스 비교 가능!

## 5. SVM (Support Vector Machine)

- 분류(Classification)에 강력한 알고리즘
- `fit()` 으로 학습, `predict()` 로 예측
- 하이퍼파라미터 `gamma`, `C` 등을 튜닝하면 성능 달라짐

## 6. 데이터 전처리 & 입력 방식

- 2차원 이미지도 벡터(1차원)로 변환해야 사용 가능

- 예: 8x8 이미지 → 64 차원 벡터
  - 읽는 순서도 중요 (랜덤 vs 순차적 → 편향 유발 가능)
- 

## 7. 일반화 능력 & 전이학습 (Transfer Learning)

- **일반화**: 훈련 데이터가 아니더라도 새 데이터도 잘 분류
  - **전이학습**: A 분야에서 배운 걸 B에 활용 가능  
예: 강아지 분류 모델을 고양이 분류에 일부 활용
- 

## 8. 정확도 측정 & 성능 평가

- \*정확도(Accuracy)\*\*가 기본
  - 공학 논문은 기존보다 조금이라도 성능이 좋아야 인정
  - **일반화 성능**도 중요: 새로운 환경에서도 잘 작동해야 함
- 

## 실생활 예시로 이해하기

- 계란 크기 자동 분류기
  - 무게 or 이미지로 특징 추출
  - 상/중/하로 레이블링
  - 기계학습 → 실시간 분류기로 작동
- 사과 분류
  - 색깔, 크기, 상처 여부 등을 기준으로 분류기 훈련