📌 머신러닝 핵심 요약 (중간고사 범위 기준)1. 머신러닝의 정의와 동기

정의: 명시적 프로그래밍 없이 컴퓨터가 경험을 통해 작업 성능을 향상시키는 알고리즘.

대표 정의:

Arthur Samuel (1959): “컴퓨터가 명시적 프로그래밍 없이 학습할 수 있도록 하는 연구 분야”

Tom Mitchell (1997): “작업 T, 성능 측정 P, 경험 E를 통해 P가 향상되면, 학습이 일어난 것”

사용 이유:

규칙 기반 접근의 한계 → 유지 보수 어려움

복잡한 문제 자동 해결

변화에 적응 가능 (예: 스팸 필터)

데이터 마이닝, 자동화된 패턴 발견

2. 머신러닝 분류 방식✅ 지도 학습 (Supervised Learning)

입력 + 레이블 제공

주요 과제:

분류 (Classification): ex. 이메일 스팸 여부

회귀 (Regression): ex. 주택 가격 예측

✅ 비지도 학습 (Unsupervised Learning)

입력만 제공 (레이블 없음)

주요 과제:

군집화 (Clustering): KMeans, 계층 군집

차원 축소 (Dimensionality Reduction): PCA, t-SNE

이상치 탐지 (Outlier Detection)

✅ 강화 학습 (Reinforcement Learning)

에이전트가 환경과 상호작용하며 보상을 통해 정책(policy) 학습

3. 학습 방식에 따른 구분

배치 학습 (Batch Learning): 전체 데이터를 한 번에 학습, 오프라인 학습

온라인 학습 (Online Learning): 데이터가 도착할 때마다 순차 학습, 학습률 중요

외부 메모리 학습 (Out-of-Core Learning): 매우 큰 데이터셋에 대해 부분 학습

4. 사례 기반 vs 모델 기반 학습

사례 기반 (Instance-based): 데이터를 저장하고 유사도 기반 예측 (예: KNN)

모델 기반 (Model-based): 수학적 모델을 만들어 예측 (예: 선형 회귀)

5. 머신러닝 프로젝트 절차

문제 정의

데이터 수집 및 탐색

훈련/검증 데이터 분할

데이터 전처리 (결측치 처리, 인코딩 등)

모델 선택 및 훈련

평가 (검증 세트, 교차 검증)

하이퍼파라미터 튜닝 (GridSearch, RandomSearch)

배포 및 모니터링

6. 데이터 전처리

결측치 처리: 제거, 평균/중간값 대체

카테고리 처리:

OrdinalEncoder (순서형)

OneHotEncoder (명목형)

특성 스케일링: StandardScaler

파생 특성 생성: 사용자 정의 변환기 사용

7. 모델 학습 및 평가🔍 평가 지표

회귀: RMSE, MAE

분류: 정확도, 정밀도, 재현율, F1, ROC-AUC

🔍 과적합 vs 과소적합

과대적합 (Overfitting): 모델이 너무 복잡

규제 필요 (L1, L2), 데이터 확장

과소적합 (Underfitting): 모델이 너무 단순

더 복잡한 모델 사용, 특성 추가

8. 교차 검증과 하이퍼파라미터 튜닝

K-Fold Cross Validation: 성능 평균화

GridSearchCV: 모든 조합 탐색

RandomizedSearchCV: 일부 조합 무작위 탐색

9. 분류 시스템 실습 (MNIST 데이터셋)

숫자 이미지 (28x28) → 784 특성

SGDClassifier, KNN, SVM 등 실습

주요 개념:

오차 행렬 (Confusion Matrix)

정밀도/재현율 트레이드오프

ROC 곡선과 AUC

10. 다중 분류 & 다중 레이블 분류

다중 클래스 (Multiclass): OvR, OvO 전략

다중 레이블 (Multilabel): 여러 레이블 예측 (ex. 홀수 + 큰 숫자)

ClassifierChain으로 레이블 간 의존성 반영 가능

11. SVM (Support Vector Machine)

선형 SVM: 라지 마진 분류, LinearSVC

소프트 마진: 하드 마진의 이상치 민감성 해결

비선형 SVM:

다항식 커널

RBF (가우시안 커널)

문자열 커널 등

하이퍼파라미터: C, gamma, degree등

SVM 회귀 (SVR): 마진 안의 오차 허용