

5주차 과제

1. 서포트 벡터 머신의 근본적인 아이디어는 무엇인가?

마진을 최대화하는 초평면을 찾아 데이터 분류 시 일반화 성능을 높이는 것.

2. 서포트 벡터가 무엇인가?

결정 경계에 가장 가까워 마진 정의에 핵심 역할을 하는 학습 샘플.

3. SVM을 사용할 때 입력값의 스케일이 왜 중요한가?

특징별 스케일 차이가 거리 계산에 영향을 주어 모델 성능을 저하시킬 수 있으므로 동일 스케일로 변환해야 한다.

4. SVM 분류기가 샘플을 분류할 때 신뢰도 점수와 확률을 출력할 수 있는가?

`decision_function`으로 신뢰도 점수를 얻을 수 있고, `probability=True` 설정 시 확률 예측이 가능하다.

5. RBF 커널을 사용해 SVM 분류기를 훈련시켰더니 훈련 세트에 과소적합되었다. 이때, 어떤 값을 어떻게 수정해야 하는가?

규제 파라미터 `C`를 증가시키거나 `gamma` 값을 증가시켜 모델 복잡도를 높인다.

6. (심화) MNIST 데이터셋에 SVM 분류기를 훈련시켜라. (SVM 분류기는 이진 분류기이므로 OvR 전략을 사용하여 10개의 숫자를 분류해야 한다.)

`fetch_openml("mnist_784")`로 데이터 로드 후 `StandardScaler` 적용, `SVC(kernel="rbf", C=..., gamma=...)`로 OvR 학습, `accuracy_score`로 평가.

7. (심화) 캘리포니아 주택 가격 데이터셋에 SVM 회귀를 훈련시켜라.

`fetch_california_housing()`으로 데이터 로드 후 `StandardScaler` 적용, `SVR(kernel="rbf", C=..., gamma=..., epsilon=...)`로 학습, `mean_squared_error` 및 `r2_score`로 평가.