12181998 김태훈 – 수업 외 방법으로 분류

분류 방식 – 머신러닝 알고리즘

1. 서포트 벡터 머신 (Support Vector Machine)

2. 로지스틱 회귀 (Logistic Regression)

3. K-이웃 알고리즘 (K-Nearest Neighbors)

4. 랜덤 포레스트 (Random Forest)

5. 가우시안 나이브 베이즈 (Gaussian Naïve Bayes)

각 알고리즘 설명

1. 서포트 벡터 머신 (SVM)

데이터를 선형적으로 구분할 수 있는 최적의 결정 경계를 만드는 두 개의 서포트 벡터를 찾아 분류하는 알고리즘

Python scikit-learn package의 SVC (support vector machine classifier) 모델로 구현

2. 로지스틱 회귀 (Logistic Regression)

선형 회귀 모델을 시그모이드 함수에 대입하여 도출된 0과 1 사이의 확률값을 통한 분류 알고리즘

Python scikit-learn package의 LogisticRegression 모델로 구현

3. K-이웃 알고리즘 (K-Nearest Neighbors)

각 데이터 포인트와 가장 가까운 K개의 이웃 데이터 중 가장 많은 클래스로 예측하는 알고리즘

Python scikit-learn package의 KNeighborsClassifier 모델로 구현

4. 랜덤 포레스트 (Random Forest)

여러 개의 Decision Tree를 생성하여 각 데이터셋을 분류하고, 분류 결과를 조합하여 최종 분류를 하는 알고리즘

Python scikit-learn package의 RandomForestClassifier 모델로 구현

5. 가우시안 나이브 베이즈 (Gaussian Naïve Bayes)

각 클래스에 대한 특성들의 확률 분포를 가우시안 분포를 통하여 추정한 후, 베이즈 정리의 사후 확률을 통해 데이터를 분류하는 알고리즘

Python scikit-learn package의 GaussianNB 모델로 구현

머신러닝 코드 구성

1. SearchCV.py

각 모델의 하이퍼 파라미터 튜닝 및 Cross-Validation을 통한 분류 성능 검증하는 파이썬 클래스 구현 코드

2. Classification\_by\_ML.py

이상치 제거된 데이터셋을 기준으로 정규화된 데이터셋으로 각 모델을 훈련 및 평가(분류)하는 실행 코드

분류 결과 : 모든 모델에서 정확도 100%의 분류 성능을 나타냄

학습 결과 및 평가 결과는 Classification\_by\_ML.log 파일에 기록되어 있음