

프로젝트 제안서

5조 이유정, 김나경, 김정하, 김수민, 양승준

1. 주제

적색 와인 품질 예측

2. 프로젝트 목표

본 프로젝트의 목표는 적색 와인의 화학적 특성(물리, 화학적 변수)을 바탕으로 품질 점수를 예측할 수 있는 심층 신경망(Deep Neural Network, DNN) 모델을 구축하고, 이를 단계적으로 발전시켜 최적화하고, 높은 일반화 성능을 보이는 모델을 개발하는 것입니다.

타깃 변수인 품질(quality)은 평가를 통해 부여된 3~8 사이의 정수형 점수로, 와인의 품질 수준을 나타냅니다. 이러한 특성에 따라 본 프로젝트는 품질 점수를 예측하는 회귀(regression) 문제로 정의됩니다.

모델은 총 11개의 물리·화학적 특성(feature)을 입력 변수로 사용하여 품질 점수를 예측하도록, DNN 모델의 성능 향상 과정 자체에 초점을 맞춰 진행될 예정입니다.

우선, 단순한 구조의 DNN 모델을 구축하고 Training 데이터를 통한 학습을 통해 의도적으로 오버피팅을 발생시켜 모델이 훈련 데이터에 과도하게 적응하는 양상을 분석하고, 이후 단계에서 일반화 성능을 향상시키기 위한 다양한 방법을 적용할 예정입니다.

이 과정에는 Dropout Layer 추가, L2 정규화(Regularization), 모델 파라미터 조정 등이 포함될 예정입니다.

최종적으로는 이러한 반복적 실험을 통해 모델의 일반화 성능을 점진적으로 개선하고, 안정적인 품질 예측 성능을 갖춘 DNN 모델을 완성하는 것을 목표로 합니다.

3. 데이터셋 개요

출처: Kaggle “Red Wine Quality” 데이터셋

<https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009/data>

- 표본 수: 1,599개의 적색 와인 샘플

입력 변수(특징):

- fixed acidity (고정 산도)
- volatile acidity (휘발성 산도)
- citric acidity (구연산 함량)
- residual sugar (잔류 당분)
- chlorides (염화물 함량)
- free sulfur dioxide (유리 이산화황)
- total sulfur dioxide (총 이산화황)
- density (밀도)
- pH (산도)
- sulphates (황산염 함량)
- alcohol (알코올 도수)
- quality(품질, target^값)

4. 예상 모델 및 알고리즘

회귀 문제 해결을 위해 비선형 모델인 DNN(Deep Neural Network)을 사용할 예정입니다.

- DNN

먼저 단순한 DNN 모델을 구축하고 훈련하여 기준 성능을 설정하고, 테스트 데이터의 성능을 평가하고 과적합 여부를 분석합니다.

- 최적화를 통한 성능 개선:

핵심 단계에서는 과적합을 방지하고 예측 성능을 향상시키기 위해 다양한 기법을 적용하고 실험합니다. 다음과 같은 기법을 사용할 예정입니다.

- L2 Regularization
- Dropout
- Early Stopping
- Batch Normalization
- Data Augmentation
- Learning Rate Scheduling
- k-Fold Cross-Validation

5. 팀원 역할 분담

본 프로젝트는 팀원 간의 협업을 통해 진행되며, 역할은 다음과 같이 나뉩니다.

- 공통 역할:

모든 팀원은 다음의 공통 업무를 수행합니다.

- 전체 모델 설계에 대한 협업 및 피드백 제공
- 데이터셋 조사 및 분석 참여

- 개별 주요 역할:

모델링 및 보고서 작성 (김정하, 이유정, 김수민, 양승준)

- 데이터 전처리, 모델 학습 및 최적화 등 모델링 전 과정을 담당
- 주차별로 돌아가며 보고서 작성

PPT 제작 (김나경)

- 최종 발표 자료 PPT 제작 담당