



Alchemist Train Session

CHAP 06 차원 축소

차원 축소 실습 목차

01. 시작하기 전에

01-1. Guide-Line

01-2. Dataset 자료 다운 & 소개

02. Red Wine Quality 워크북



01.

시작하기 전에



Guide-Line

- Github에서 Notebook file 다운 & Kaggle 사이트 참고하기
- 조원과 워크북 질문 답하기 & Jupyter Notebook 코드 빈칸 채우기
- 팀별 발표🔥

[Session](#) / [Practice](#) / 7주차 /

Add file ▾



yellowsubmarine372 Add files via upload

3d30ccd · now History

Name	Last commit message	Last commit date
..		
RedWineQuality	Add files via upload	now
README.md	Create README.md	2 months ago

Red Wine Quality 자료 다운



UCI MACHINE LEARNING AND 1 COLLABORATOR · UPDATED 6 YEARS AGO

▲ 2771

New Notebook

Download (26 kB)



Red Wine Quality

Simple and clean practice dataset for regression or classification modelling



Data Card Code (1749) Discussion (29) Suggestions (0)

About Dataset

Context

The two datasets are related to red and white variants of the Portuguese "Vinho Verde" reference [Cortez et al., 2009]. Due to privacy and logistic issues, only physicochemical (inputs) and sensory (the output) variables are available (e.g. there is no data about grape types, wine brand, wine selling price,

These datasets can be viewed as classification or regression tasks. The classes are more normal wines than excellent or poor ones).

This dataset is also available from the UCI machine learning repository, <https://archive.ics.uci.edu/ml/datasets/wine+quality>, I just shared it to kaggle for convenience. (If I am mistaken and the public license type disallows me from doing so, I will take this down if requested.)

winequality-red.csv (100.95 kB)



Detail Compact Column

10 of 12 columns ▾

About this file

Context

This dataset is related to red variants of the Portuguese "Vinho Verde" wine. For more details, consult the reference [Cortez et al., 2009]. Due to privacy and logistic issues, only physicochemical (inputs) and sensory (the output) variables are available (e.g. there is no data about grape types, wine brand, wine selling price, etc.).

The datasets can be viewed as classification or regression tasks. The classes are ordered and not balanced (e.g. there are much more normal wines than excellent or poor ones).

This dataset is also available from the UCI machine learning repository, <https://archive.ics.uci.edu/ml/datasets/wine+quality>, I just shared it to kaggle for convenience. (If I am mistaken and the public license type disallows me from doing so, I will take this down if requested.)

Content

For more information, read [Cortez et al., 2009].

# fixed acidity	# volatile acidity	# citric acid	# residual sugar	# chlorides
most acids involved with wine or fixed or	the amount of acetic acid in wine, which at too high	found in small quantities, citric acid can add	the amount of sugar remaining after	the amount of salt in the wine

Data Explorer

Version 2 (100.95 kB)

winequality-red.csv

Red Wine Quality Dataset 소개

Data Context

- Kaggle의 레드 와인 퀄리티 데이터셋은 포르투갈산 "Vinho Verde" 레드와인과 화이트와인에 대한 두 가지 데이터셋으로 구성. 데이터셋은 Cortez et al., 2009년 연구에 기반한 것으로, 포도 종류, 와인 브랜드, 판매 가격 등과 같은 정보는 개인정보 보호 및 물류 문제로 인해 포함되어 있지 않음.
- 물리화학적(입력 변수) 및 감각적(출력 변수) 특성에 초점을 맞추고 있음.
- 이 데이터셋은 분류 혹은 회귀 작업으로 활용될 수 있으며, 클래스는 정렬되어 있지만 균형이 맞지 않다. 즉, 우수하거나 불량한 와인보다는 보통의 와인이 훨씬 더 많다.
- 이 데이터셋을 사용하여 다양한 예측 모델을 개발하고, 와인의 품질을 결정하는데 중요한 요소가 무엇인지를 탐색할 수 있다.

Red Wine Quality Dataset 소개

Data Description

- 1 - fixed acidity
- 2 - volatile acidity
- 3 - citric acid
- 4 - residual sugar
- 5 - chlorides
- 6 - free sulfur dioxide
- 7 - total sulfur dioxide

- 8 - density
- 9 - pH
- 10 - sulphates
- 11 - alcohol

Output variable (based on sensory data):

- 12 - quality (score between 0 and 10)

02.

Red Wine Quality 워크북

RedWine Quality 워크북 GuideLine

1. 빈 셀 채우기

2. Loading & Reading the Libraries

3. Data inspection

4. Basis Statics

2. 시각화 진행 및 이상치 처리

5. Feature Analysis & Finding Outliers

3. Quiz 1 답하기

6. Feature Transformation

4. Quiz 2 & Quiz 3 답하기

9. PCA Implemenation

5. Conclusion 읽고 이해해보기



수고하셨습니다