

### Project Overview:

Goal: which factor among the 13 differences is the key or the most contributing factor to distinguish between different wines.

Dataset: Wine dataset, from UC Irvine Machine Learning Repository  
<https://archive.ics.uci.edu/dataset/109/wine>,

### Data Processing:

How you loaded: I downloaded the file, convert it to csv file manually and read the csv in RUST

Cleaning: no data cleaning, since there is no missing value

### Code Structure:

#### Key Function and Types:

Struct Wine: for basic wine type for data processing

fn mean\_features: average function, takes in a wine struct and a feature name, return the average value of that feature's value of that type of wine

fn std\_features: std function, takes in a wine struct and a feature name, return the standard deviation of that feature's value of that type of wine

#### Workflow:

Take the mean, std, and find the importance vector through (max mean - min mean)/mean std

### Test:

```
(base) zyc@crc-dot1x-nat-10-239-75-211 project1 % cargo run
```

```
Finished `dev` profile [unoptimized + debuginfo] target(s) in 0.11s
```

```
Running `target/debug/project1`
```

```
_____ alcohol average _____
```

```
Class 1: 13.745
```

```
Class 2: 12.279
```

```
Class 3: 13.154
```

```
_____ alcohol standard deviation _____
```

```
Class 1: 0.462
```

```
Class 2: 0.538
```

```
Class 3: 0.530
```

```
alcohol Score: 2.8739097
```

```
_____ malicacid average _____
```

```
Class 1: 2.011
```

```
Class 2: 1.933
```

```
Class 3: 3.334
```

```
_____ malicacid standard deviation _____
```

```
Class 1: 0.689
```

Class 2: 1.016  
Class 3: 1.088  
malicacid Score: 1.5054389  
\_\_\_\_\_ ash average \_\_\_\_\_  
Class 1: 2.456  
Class 2: 2.245  
Class 3: 2.437  
\_\_\_\_\_ ash standard deviation \_\_\_\_\_  
Class 1: 0.227  
Class 2: 0.315  
Class 3: 0.185  
ash Score: 0.8695098  
\_\_\_\_\_ alccalinity\_of\_ash average \_\_\_\_\_  
Class 1: 17.037  
Class 2: 20.238  
Class 3: 21.417  
\_\_\_\_\_ alccalinity\_of\_ash standard deviation \_\_\_\_\_  
Class 1: 2.546  
Class 2: 3.350  
Class 3: 2.258  
alccalinity\_of\_ash Score: 1.6111999  
\_\_\_\_\_ magnesium average \_\_\_\_\_  
Class 1: 106.339  
Class 2: 94.549  
Class 3: 99.312  
\_\_\_\_\_ magnesium standard deviation \_\_\_\_\_  
Class 1: 10.499  
Class 2: 16.753  
Class 3: 10.890  
magnesium Score: 0.92727745  
\_\_\_\_\_ total\_phenol average \_\_\_\_\_  
Class 1: 2.840  
Class 2: 2.259  
Class 3: 1.679  
\_\_\_\_\_ total\_phenol standard deviation \_\_\_\_\_  
Class 1: 0.339  
Class 2: 0.545  
Class 3: 0.357  
total\_phenol Score: 2.806958  
\_\_\_\_\_ flavanoids average \_\_\_\_\_  
Class 1: 2.982  
Class 2: 2.081  
Class 3: 0.781  
\_\_\_\_\_ flavanoids standard deviation \_\_\_\_\_

Class 1: 0.397  
Class 2: 0.706  
Class 3: 0.294  
flavanoids Score: 4.7273936  
\_\_\_\_\_ nonflavanoid\_phenols average \_\_\_\_\_  
Class 1: 0.290  
Class 2: 0.364  
Class 3: 0.447  
\_\_\_\_\_ nonflavanoid\_phenols standard deviation \_\_\_\_\_  
Class 1: 0.070  
Class 2: 0.124  
Class 3: 0.124  
nonflavanoid\_phenols Score: 1.485147  
\_\_\_\_\_ proanthocyanins average \_\_\_\_\_  
Class 1: 1.899  
Class 2: 1.630  
Class 3: 1.154  
\_\_\_\_\_ proanthocyanins standard deviation \_\_\_\_\_  
Class 1: 0.412  
Class 2: 0.602  
Class 3: 0.409  
proanthocyanins Score: 1.5722569  
\_\_\_\_\_ color\_intensity average \_\_\_\_\_  
Class 1: 5.528  
Class 2: 3.087  
Class 3: 7.396  
\_\_\_\_\_ color\_intensity standard deviation \_\_\_\_\_  
Class 1: 1.239  
Class 2: 0.925  
Class 3: 2.311  
color\_intensity Score: 2.8894968  
\_\_\_\_\_ hue average \_\_\_\_\_  
Class 1: 1.062  
Class 2: 1.056  
Class 3: 0.683  
\_\_\_\_\_ hue standard deviation \_\_\_\_\_  
Class 1: 0.116  
Class 2: 0.203  
Class 3: 0.114  
hue Score: 2.6229079  
\_\_\_\_\_ OD280\_OD315\_of\_diluted\_wines average \_\_\_\_\_  
Class 1: 3.158  
Class 2: 2.785  
Class 3: 1.684

\_\_\_\_\_ 0D280\_0D315\_of\_diluted\_wines standard deviation \_\_\_\_\_

Class 1: 0.357

Class 2: 0.497

Class 3: 0.272

0D280\_0D315\_of\_diluted\_wines Score: 3.9286873

\_\_\_\_\_ proline average \_\_\_\_\_

Class 1: 1115.712

Class 2: 519.507

Class 3: 629.896

\_\_\_\_\_ proline standard deviation \_\_\_\_\_

Class 1: 221.521

Class 2: 157.211

Class 3: 115.097

proline Score: 3.6219308

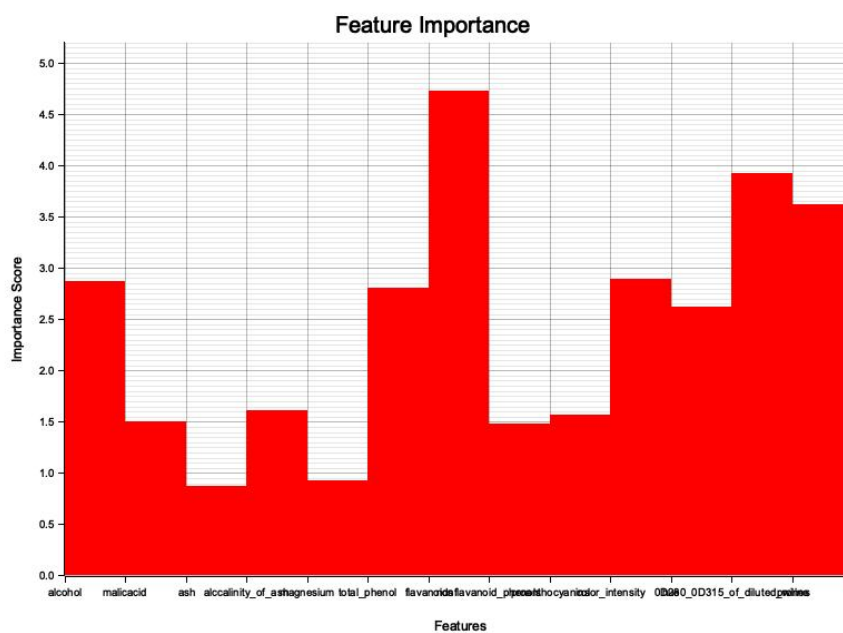
```
[
  (
    "alcohol",
    2.8739097,
  ),
  (
    "malicacid",
    1.5054389,
  ),
  (
    "ash",
    0.8695098,
  ),
  (
    "alccalinity_of_ash",
    1.6111999,
  ),
  (
    "magnesium",
    0.92727745,
  ),
  (
    "total_phenol",
    2.806958,
  ),
  (
    "flavanoids",
    4.7273936,
  ),
  (
```

```

        "nonflavanoid_phenols",
        1.485147,
    ),
    (
        "proanthocyanins",
        1.5722569,
    ),
    (
        "color_intensity",
        2.8894968,
    ),
    (
        "hue",
        2.6229079,
    ),
    (
        "OD280_OD315_of_diluted_wines",
        3.9286873,
    ),
    (
        "proline",
        3.6219308,
    ),
]
(base) zyc@crc-dot1x-nat-10-239-75-211 project1 %

```

Result:



We found out that, according to the importance score, Flavanoids, OD280\_OD315\_of\_diluted\_wines, Proline are the first 3 contributor of classifying the three different type of wines accordingly.