

Machine Learning

Project proposal

1 Team members:

- Luis Sierra Muntané
- Àlex Batlle Casellas
- Aleix Torres i Camps

2 Problem: **Wine type and quality determination**

Searching for a interesting data set we found [this one](#) that quickly intrigue us. Because the title says that it is a data set about wine quality, but only wine quality? Which are the features that can determine wine quality?

So looking at the information given in the same webpage, we found that the input variables are all physico-chemical components and properties of the wine. So the wine itself, without commercial information. Therefore it is a pure scientific analysis of the wine, doesn't matter the brand or the price.

That is what we find interesting, which objective chemical attributes of the wine affect, or can help us predict the subjective punctuation (in a 0-10 scale) that the oenologists gave them. This is the problem that the web page suggest, but we decided to add extra challenge. We saw that there were two types of wines: red and white, so how about to predict wine type too? Because, a priori, we don't know which physicochemical traits of the wine make them white or red.

So, finally, this is our (double) final problem: determine the quality of the wine (which can be thought as a regression problem) and also determine which is its type (which is a pure classification problem).

3 References:

1. Web page where the data set can be found:
<http://archive.ics.uci.edu/ml/datasets/Wine+Quality>
2. Relevant paper (a previous approach):
<https://www.sciencedirect.com/science/article/pii/S0167923609001377?via%3Dihub>
3. Source. Paulo Cortez, University of Minho, Guimarães, Portugal:
<http://www3.dsi.uminho.pt/pcortez/Home.html>
4. *Vinho verde*: the winery where the wine came from:
<https://www.vinhoverde.pt/en/>