

Content

The data was downloaded from UCI Machine Learning Repository:

<https://archive.ics.uci.edu/ml/datasets/wine+quality>

The two datasets are related to red and white variants of the Portuguese "Vinho Verde" wine. For more details, consult: [\[Web Link\]](#) or the reference [Cortez et al., 2009].

In the above reference, two datasets were created, using red and white wine samples. The inputs include objective tests (e.g. PH values) and the output is based on sensory data (median of at least 3 evaluations made by wine experts). Each expert graded the wine quality between 0 (very bad) and 10 (very excellent). 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.).

These datasets can be viewed as classification or regression tasks. The classes are ordered and not balanced (e.g. there are many more normal wines than excellent or poor ones). Outlier detection algorithms could be used to detect the few excellent or poor wines. Also, we are not sure if all input variables are relevant. So it could be interesting to test feature selection