

シミュレーション技法特論最終課題

小林世奈 (239x025x)

1. 序論

本課題では、ワインの品質に関するデータ[1]を分析対象に、システムを開発した。データは、変数が12個あり、fixed acidity (酒石酸濃度)、volatile acidity (酢酸濃度)、citric acid (クエン酸濃度)、residual sugar (残留糖分濃度)、chlorides (塩化ナトリウム濃度)、free sulfur dioxide (遊離亜硫酸濃度)、total sulfur dioxide (総亜硫酸濃度)、density (密度)、pH (ph)、sulphates (硫酸カリウム濃度)、alcohol (アルコール度数)、quality (品質 (0 から 10 の間のスコア)) となっている。これらの変数から、品質の優劣に対する知見が得られるのではないかと考え、システムを開発した。またデータのサンプル数は140である。

2. 手法

本システムは、3つのビューで構成されている。1つ目は、変数同士の相関係数をヒートマップで表示したものである。2つ目は、データをMDSにより次元削減し、散布図で表示したもの(次元削減プロットと呼ぶ。)である。3つ目は、データを品質ごとにバーチャートで表示したものである。

3. 結果

Fig1は可視化結果である。(a)はfixed acidity (酒石酸濃度)、volatile acidity (酢酸濃度)、citric acid (クエン酸濃度)、residual sugar (残留糖分濃度)の4変数同士の相関係数をヒートマップで表したものである。マウスをヒートマップ上に置くと、相関係数が下に表示される。(b)はquality 以外の変数を次元削減し、散布図で表示した次元削減プロットである。プロットする際に、品質の数値により色付けがされている。(c)は品質ごとにバーチャートで表示したものである。バーチャート上を選択すると、その品質に対応するデータが、次元削減プロット上で表示される。

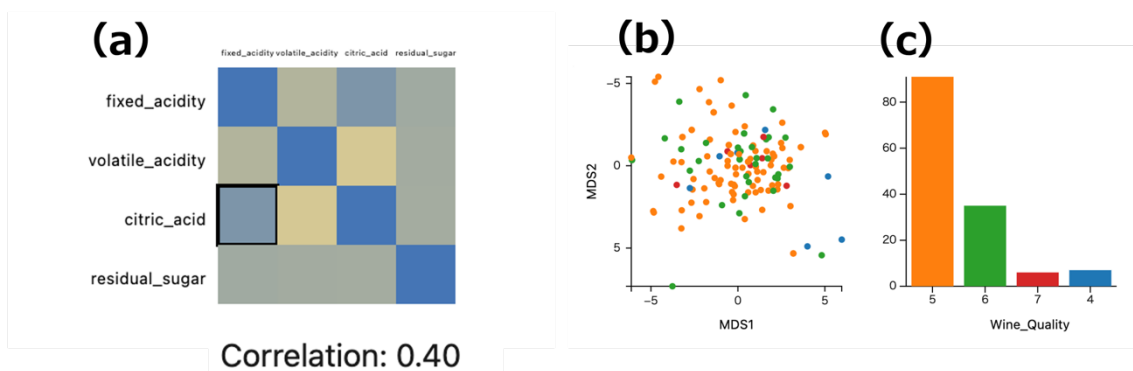


Fig1.可視化結果：(a) 変数間の相関係数ヒートマップ、(b) 次元削減プロット、
(c) バーチャート

4. 考察

(a)に関しては、変数同士の相関関係を概観することは、可能だが、品質の優劣に関する知見は、直接得られなかった。したがって、品質値ごとに変数同士の相関を見る必要がある

のではないかと考えられる。(b)に関しては、MDS を用いて次元削減を行い、品質ごとにクラスタを形成されるのではないかと考えたが、結果として、クラスタを形成しなかった。そこで、t-SNE や UMAP などのその他次元削減手法を用いるとクラスタがより形成され、品質の優劣につながる知見が得られるのではないかと考えられる。

5. 結論

本課題では、ワインに関する変数から、品質の優劣を決める要因を分析するために、可視化システムを開発した。結果として、知見は得られなかった。したがって、考察でも述べたように、システムの改善が必要だと考える。

6. 参考文献

[1]wine quality, the UC Irvine Machine Learning Repository ,
(<https://archive.ics.uci.edu/dataset/186/wine+quality>)