

Wine Quality Assessment: Neural Networks on Chemical Properties

Josuke MINAMIGUCHI

1 Introduction

1.1 Motivation

赤または白ワインの化学的特性 (特徴量) に基づいて, そのワインが品質基準—品質カテゴリの値が 6 以上—を満たすかどうかを予測したい. また, 同じデータセットに対してロジスティック回帰分析 (LR) を用いたときの精度と, ニューラルネット (NN) を用いた場合の精度とを比較したい—LR によるワインの品質予測については, ”Wine Quality Assessment: Logistic Regression on Chemical Properties”を参照.

1.2 Information of the data

データセットとして, UC Irvine Machine Learning Repository から取得したワインの品質データを利用した. そのデータセットは, 赤ワインおよび白ワインの化学的特性—アルコール含量, pH, 密度など—からなる特徴量の列と品質 (1 ~ 10) を表す列とから構成される. 赤ワインと白ワインのデータ数はそれぞれ, 1599, 4898 であった.

2 Method

特徴量データを標準化したのち, 主成分分析を行ってデータ分布の確認を行った.

Keras ライブラリを用いて フィードフォワード型ニューラルネット—入力層 11 ノード; 隠れ層 8 ノード \times 3 層; 出力層 2 ノード—でモデル化した. データを 訓練データ : テストデータ = 8 : 2 に分割して学習を行った. (エポック数とバッチ数) は, 赤ワインでは (80, 16), 白ワインでは (245, 16) とした. Stochastic gradient descent(SGD) により最適化を行った.

その後, Accuracy, Precision, Recall, F1 score, AUC から予測を評価した

以上の分析を赤ワインと白ワインのそれぞれについて行った.

3 Results

3.1 Distributions of the data

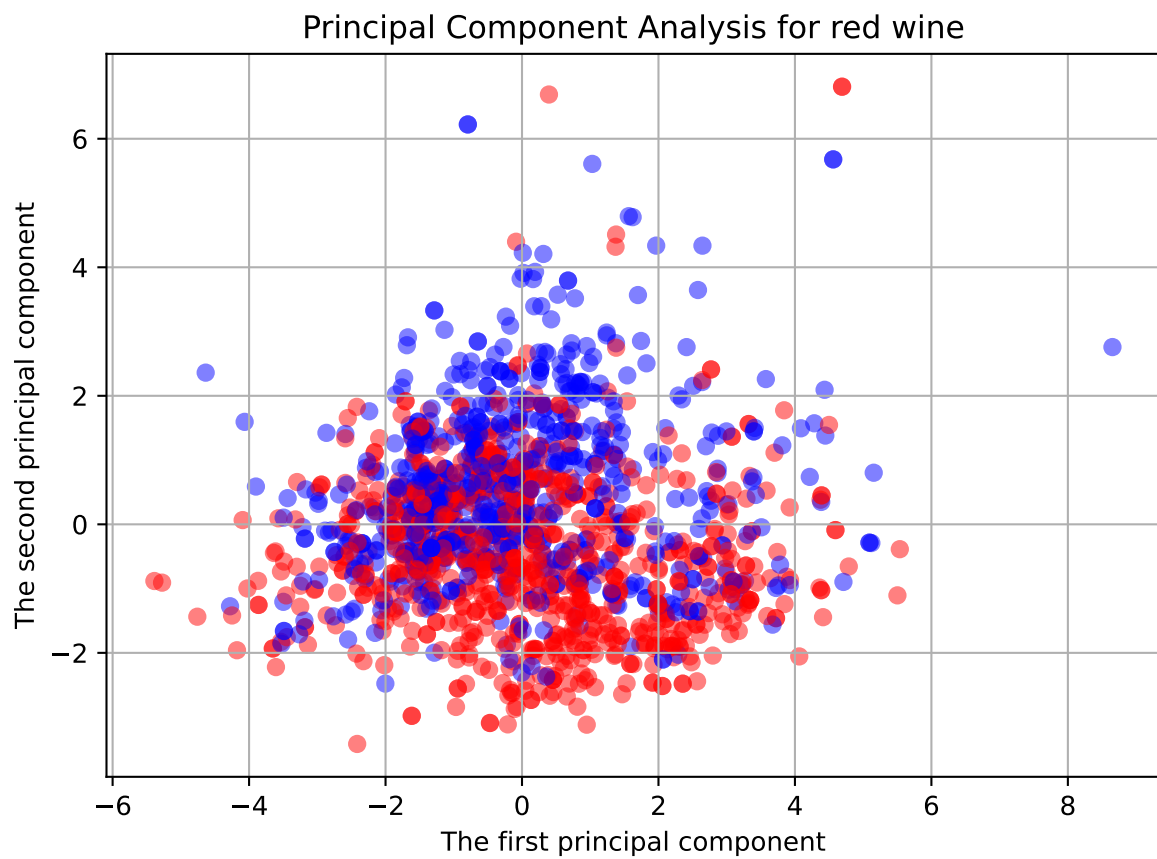


図 1: 標準化した赤ワインのデータの主成分分析

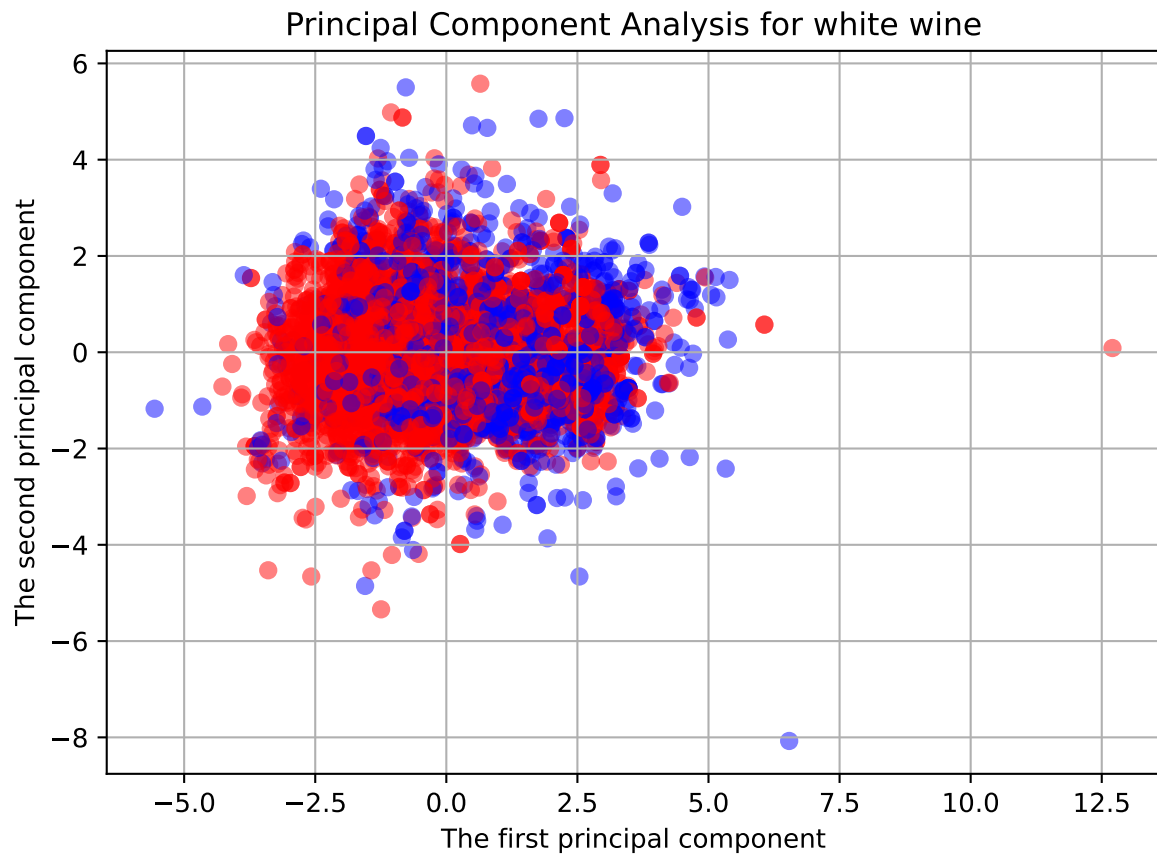


図 2: 標準化した白ワインのデータの主成分分析

3.2 Scores

表 1: 赤および白ワインのスコア

	Accuracy	Precision	Recall	F1	AUC
Red with NN	0.752 ± 0.003	0.782	0.758	0.770	0.823
White with NN	0.758 ± 0.004	0.796	0.848	0.821	0.822
Red with LR	0.741	0.786	0.737	0.761	0.819
White with LR	0.749	0.778	0.877	0.825	0.791

4 Discussion

赤および白ワインについてフィードフォワード型ニューラルネットワーク—入力層 11 ノード; 隠れ層 8 ノード × 3 層; 出力層 2 ノード—により, その品質を予測した。

ニューラルネットワークを用いた予測においては, 白ワインの方が赤ワインよりも—少なくとも Accuracy について—高精度であることがわかった。より厳密な評価には Accuracy 以外の指標についても誤差が必要で

ある。これは、白ワインの方がデータ数が多かった—赤ワインは 1599 個、白ワインは 4898 個のデータ—ためであると考えられた。(エポック数, バッチ数) を赤ワインでは (80, 16), 白ワインでは (245, 16) としたことも要因だと考えられた。例えば, エポック数を 80 に揃えた場合に結果がどのように変わるのかも調べたい。

LR による品質予測—”Wine Quality Assessment: Logistic Regression on Chemical Properties”—の結果と今回の NN による予測とを Accuracy と AUC とで比べると NN のほうが高精度だと推測された。より厳密には指標の誤差が必要。Precision や Recall, F1 によって NN と LR の予測の良し悪しを推測することは誤差なしでは難しいと考えた。

”Wine Quality Assessment: Logistic Regression on Chemical Properties”と同様に, より厳密な比較のためには, 評価指標の誤差をつけるべきである。