

Klasyfikacja win na podstawie ich skadu chemicznego.



Adrianna Janik

Dane

UCI Machine Learning repository:

<http://archive.ics.uci.edu/ml/datasets/Wine>

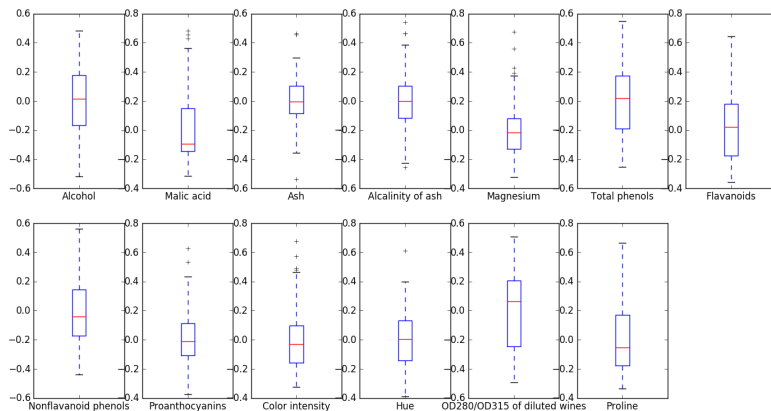
Ilo przykładów: 178

Ilo cech: 13

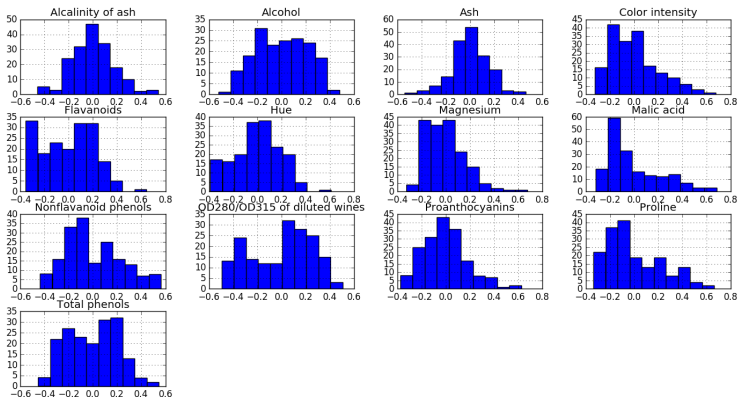
Cechy:

- ▶ Alcohol
- ▶ Malic acid
- ▶ Ash
- ▶ Alcalinity of ash
- ▶ Magnesium
- ▶ Total phenols
- ▶ Flavanoids
- ▶ Nonflavanoid phenols
- ▶ Proanthocyanins
- ▶ Color intensity
- ▶ Hue
- ▶ OD280/OD315 of diluted wines
- ▶ Proline

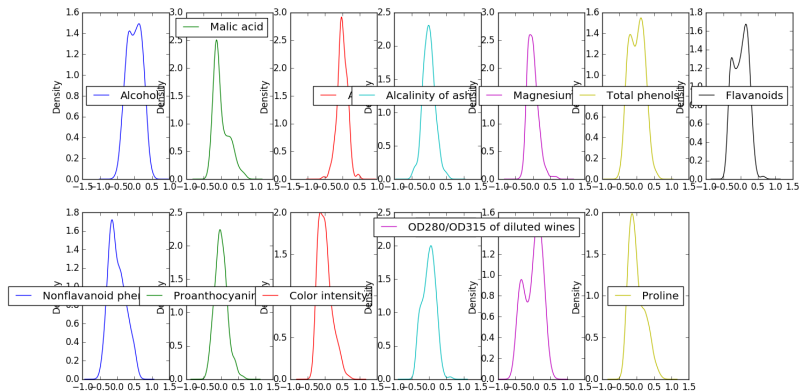
Wykresy pudekowe



Histogramy



Wykresy gstości



Normalizacja i podzia zbioru danych

$$X_{\text{norm}} = \frac{x - \mu}{x_{\text{max}} - x_{\text{min}}}$$

Klasa	Ilo przykadów	%
1	59	33%
2	71	39%
3	48	27%

zbiór treningowy: 142

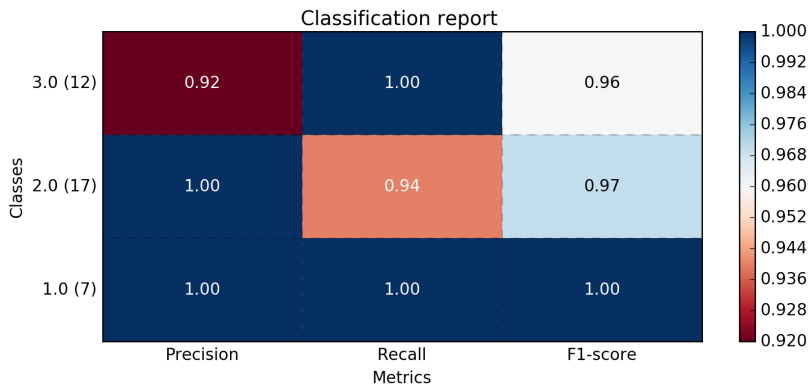
zbiór walidacyjny: 36

kroswalidacja: 10-fold

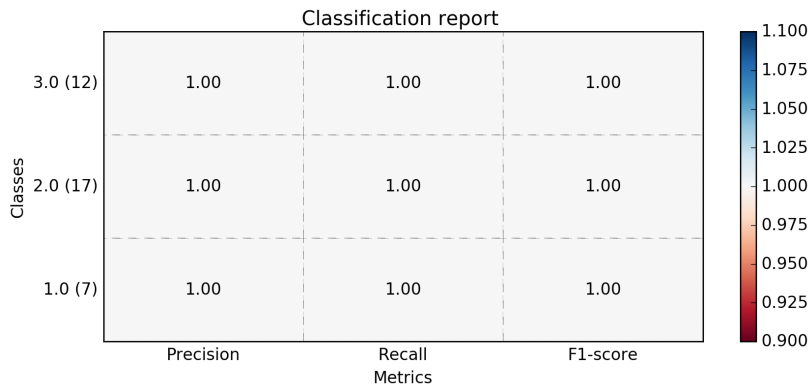
Rozwaane algorytmy

- ▶ Regresja logistyczna
- ▶ LDA
- ▶ K-Najbliszych ssiadów
- ▶ Drzewo decyzyjne
- ▶ Naiwny Bayes
- ▶ SVM
- ▶ Perceptron Wielowarstwowy
- ▶ Regresja logistyczna One vs all

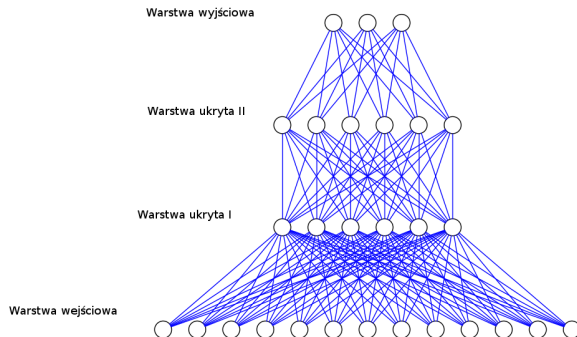
Liniowa analiza dyskryminacyjna LDA



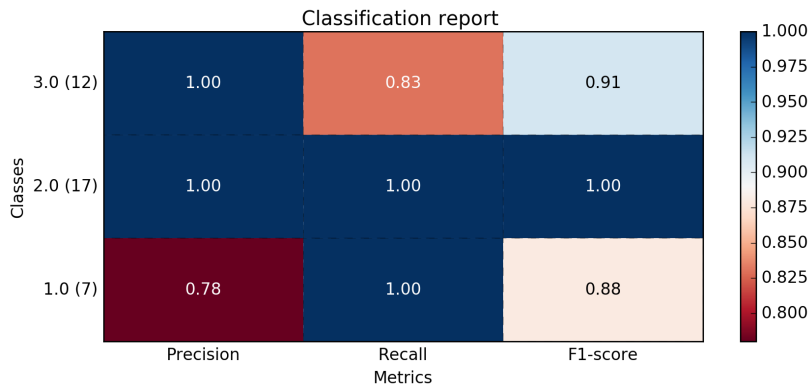
Wielowarstwowy perceptron MLP



Struktura sieci neuronowej



Regresja logistyczna - One vs all



Narzędzia

- ▶ Python 2.7.12
- ▶ biblioteki: numpy, matplotlib, scipy, scikit-learn



Wyniki

Algorytm	wynik
MLP	100%
LDA	98%
LR	95%
NB	94%
SVM	94%
KNN	92%
CART	91%
LR One vs all	85%