

Wine classification based on their chemical analysis



Adrianna Janik

Data

UCI Machine Learning repository:

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

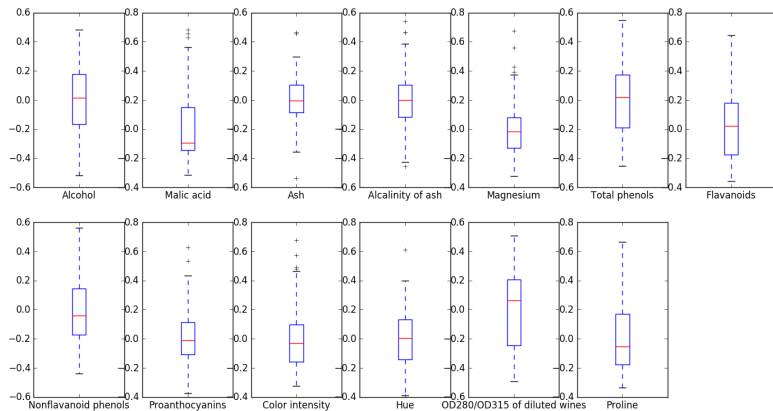
Number of instances: 178

Number of attributes: 13

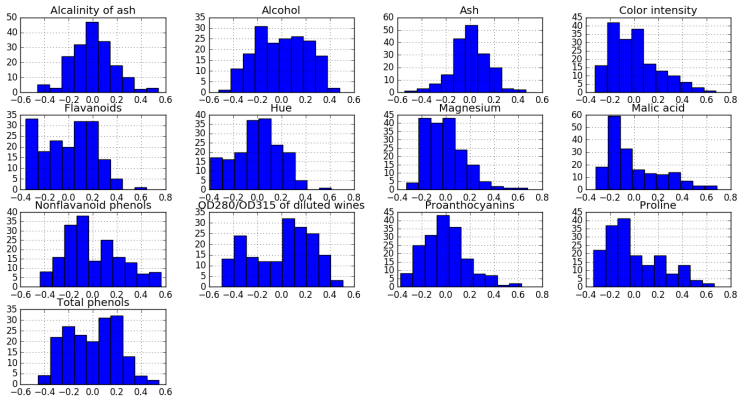
Attributes:

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

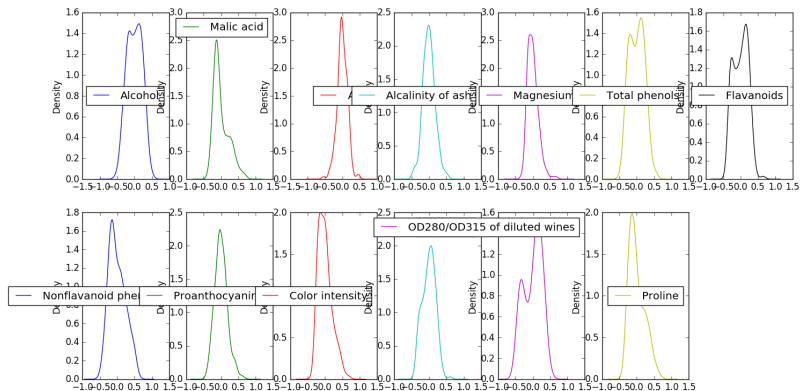
Box plots



Histograms



Density charts



Normalization and data set division

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

Class	Instances	%
1	59	33%
2	71	39%
3	48	27%

training data set: 142

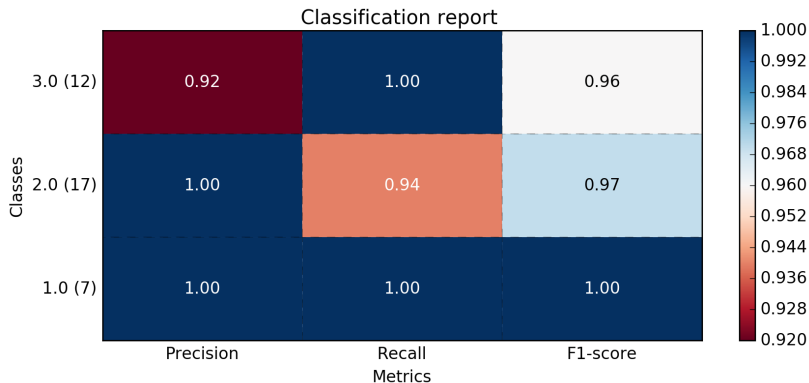
validation data set: 36

cross-validation: 10-fold

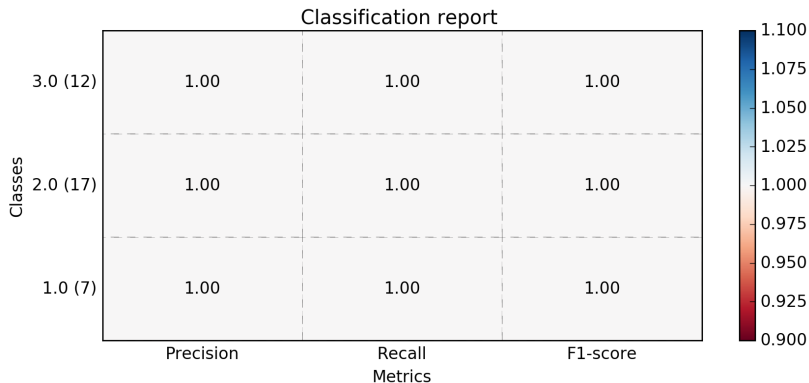
Analyzed algorithms

- ▶ Logistic regression
- ▶ LDA
- ▶ K-nearest neighbors
- ▶ Decision tree
- ▶ Naive Bayes
- ▶ SVM
- ▶ Multilayer perceptron
- ▶ Logistic regression One vs all

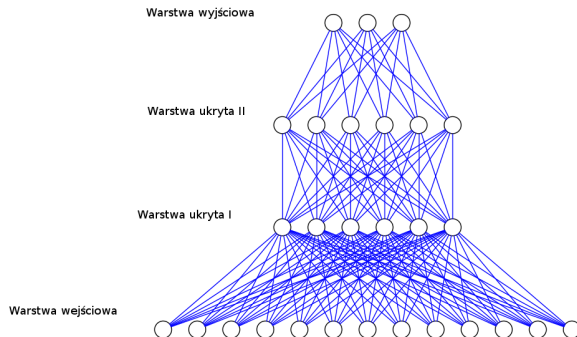
Linear discriminant analysis LDA



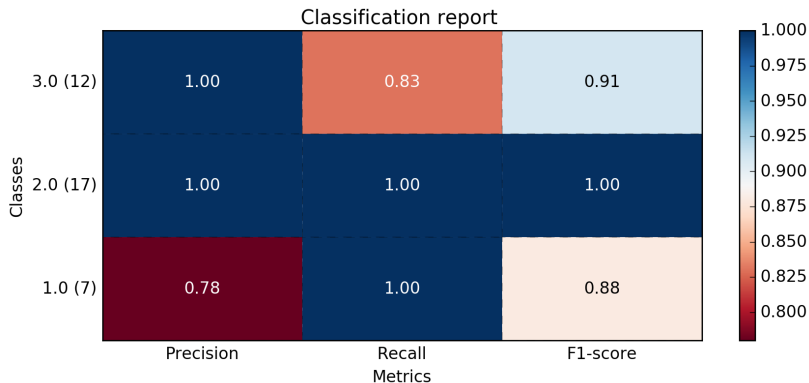
Multilayer perceptron MLP



Neural network structure



Logistic regression - One vs all



Tools

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



Results

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