Machine Learning

Classification exercise report - Group 27

Borna Feldsar
Martin Matak
Soeren Nickel

November 23, 2017

1. Datasets

1.1 Dataset *KDD Cup 1998*

Source: https://www.kaggle.com/c/184702-tu-ml-ws-17-kdd-cup-1998

1.1.1 Characteristics of the dataset

number of samples	number of dimensions	number of classes	preprocessing needed
??	??	2	??

1.1.2 Preprocessing

...

1.2 Dataset *Breast cancer*

Source: https://www.kaggle.com/c/184702-tu-ml-ws-17-breast-cancer

1.2.1 Characteristics of the dataset

number of samples	number of dimensions	number of classes	preprocessing needed
??	??	2	??

1.2.2 Preprocessing

...

1.3 Dataset *Car evaluation*

Source: https://archive.ics.uci.edu/ml/datasets/car+evaluation

1.3.1 Characteristics of the dataset

number of samples	number of dimensions	number of classes	preprocessing needed
1728	6	4	yes

1.3.2 Preprocessing

...

1.4 Dataset *Detect Malicious Executable(AntiVirus)*

Source: https://archive.ics.uci.edu/ml/datasets/Detect+Malacious+Executable(AntiVirus)

1.4.1 Characteristics of the dataset

number of samples	number of dimensions	number of classes	preprocessing needed
373	513	2	yes

1.4.2 Preprocessing

...

2. Classifiers

2.1 Naive Bayes

2.2 Support Vector Machine - SVM

2.3 Decision Tree

2.4 Random Forest

2.5 Performance measures

3. Dataset *KDD Cup 1998*

3.1 Naive Bayes

3.2 Support Vector Machine - SVM

3.3 Decision Tree

3.4 Random Forest

3.5 Comparison of classifiers

4. Dataset *Breast cancer*

4.1 Naive Bayes

4.2 Support Vector Machine - SVM

4.3 Decision Tree

4.4 Random Forest

4.5 Comparison of classifiers

5. Dataset *Car evaluation*

5.1 Naive Bayes