



Berner Fachhochschule  
Haute école spécialisée bernoise  
Bern University of Applied Sciences

# FS24 CAS PML - Python

Niklaus Johner

[niklausbernhard.johner@bfh.ch](mailto:niklausbernhard.johner@bfh.ch)

# FS24 CAS PML - Python

## 18. Machine learning with scikit-learn

# scikit-learn

- ▶ scikit-learn enthält Funktionen zum:
  - ▶ Supervised-learning:
    - ▶ classification
    - ▶ regression
  - ▶ Unsupervised-learning
    - ▶ clustering
    - ▶ density-estimation
  - ▶ Daten Vorbereitung (preprocessing, feature extraction, feature selection)
  - ▶ Estimator scoring

# Supervised-learning

- ▶ Estimators sind Objekte die die folgenden Methoden implementieren:
  - ▶ *fit(features, target)*: von features und target Daten lernen
  - ▶ *predict(features)*: von features, target Werte voraussagen
  - ▶ *score(features, target)*: den Estimator evaluieren

# Supervised-learning

```
from sklearn.datasets import load_diabetes
diabetes = load_diabetes()
data = diabetes.data
target = diabetes.target

from sklearn.linear_model import LinearRegression

lin_reg = LinearRegression()
lin_reg.fit(data, target)

prediction = lin_reg.predict(data)
lin_reg.score(data, target)
```

# Supervised-learning

```
# Linear regressor
from sklearn import linear_model
linear_model.LinearRegression

from sklearn import svm
svm.SVC # SVM classifier
svm.SVR #SVM regressor

from sklearn import tree
tree.DecisionTreeClassifier
tree.DecisionTreeRegressor

from sklearn import neural_network
neural_network.MLPRegressor
neural_network.MLPClassifier
```

# Daten pre-processing

- ▶ Man kann die Daten für das learning vorbereiten mit dem *sklearn.preprocessing* Paket
- ▶ Processor Objekten implementieren:
  - ▶ *fit(data)*: bereitet den Processor vor
  - ▶ *transform(data)*: transformiert die Daten

```
from sklearn import preprocessing
proc = preprocessing.data.StandardScaler()
proc.fit(data)
data_normed = proc.transform(data)
```

# Estimator scoring

- ▶ Die Qualität eines Estimators kann mit Funktionen von *sklearn.metrics* analysiert werden.
- ▶ Die Scoring-Funktionen nehmen die Target-Werte und eine Voraussage als Argumente.

```
# Score an estimator
from sklearn import metrics
# regression
metrics.mean_squared_error(target, prediction)
# classification
metrics.roc_auc_score(target, prediction)
```