

Dokumentation Machine Learning und Intelligente Datenanalyse

Projekt 7 - Raphael Kluge

Vorbereitung:

Daten

- Datei laser.mat wird eingelesen
- Konvertieren von y zu einem Vektor
- Plotten von Beispielen zum Näheren Verständnis der Daten
- Aufteilen der Daten in Trainings- und Testdaten

Methoden:

Kernel-Funktionen

- Lineare Klassifikation mit regressivem ERM
- Regularisierer: R2-Regularisierer
- Verlustfunktion: Hinge loss
- Prediction mit Gram-Matrix
- Plotten der Training- und Test-Scores

Polynomischer Kernel:

- Gram-Matrix wird mit polynomischen Kernel berechnet ($K = \gamma X_1 X_2^T$)^p
- diese wird mit einer Funktion umwickelt, in der Hyperparameter übergeben werden

DTW:

- Abstandsfunktionen (boolean, absolut, quadratisch)
- Kernel-Funktionen werden mit einer Funktion umwickelt, in der Hyperparameter übergeben werden
- in dieser Funktion werden für alle Hyperparameter und Distanzfunktionen eine Liste von Kernel-Funktionen generiert
- einzelne Werte der K-Matrix werden extra mit Kernel-Funktion berechnet

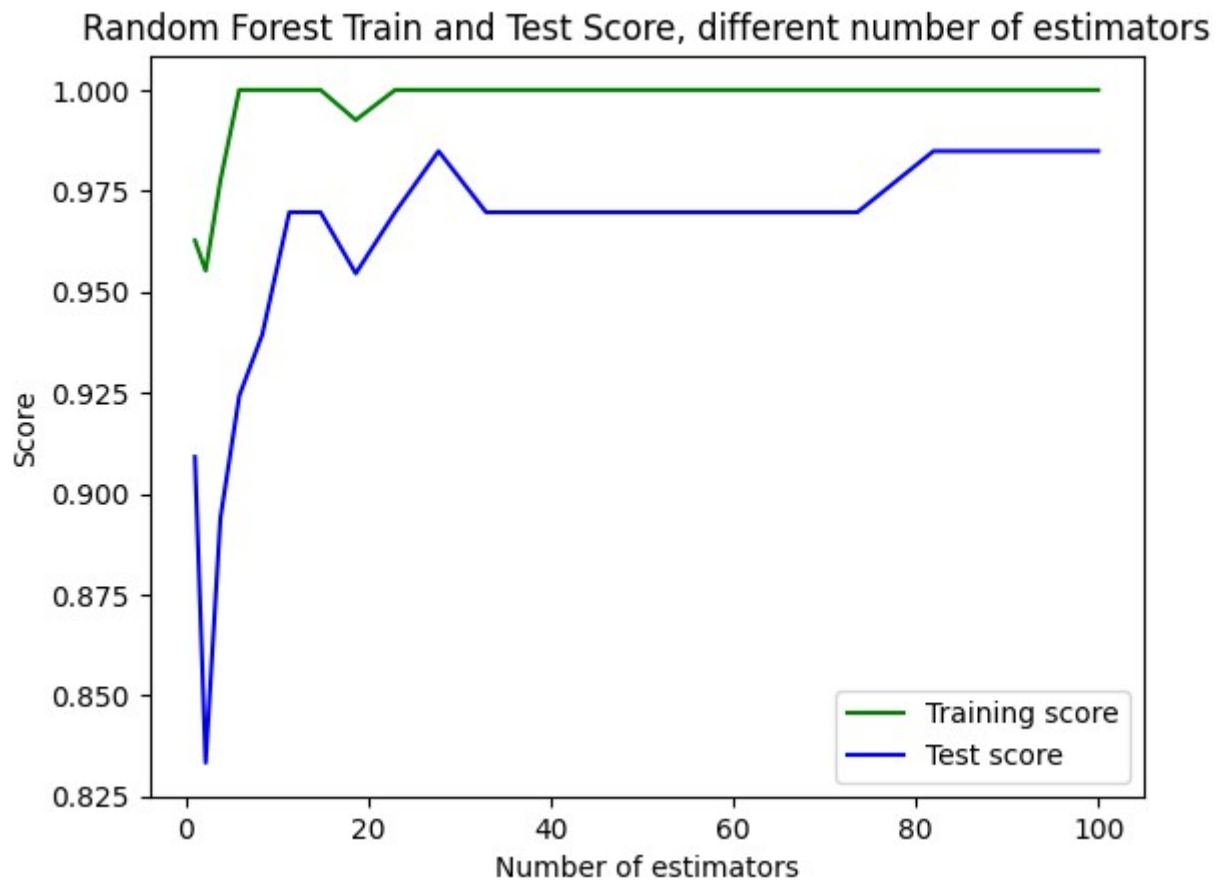
Test von weiteren Kernel-Funktionen...

Random-Forest

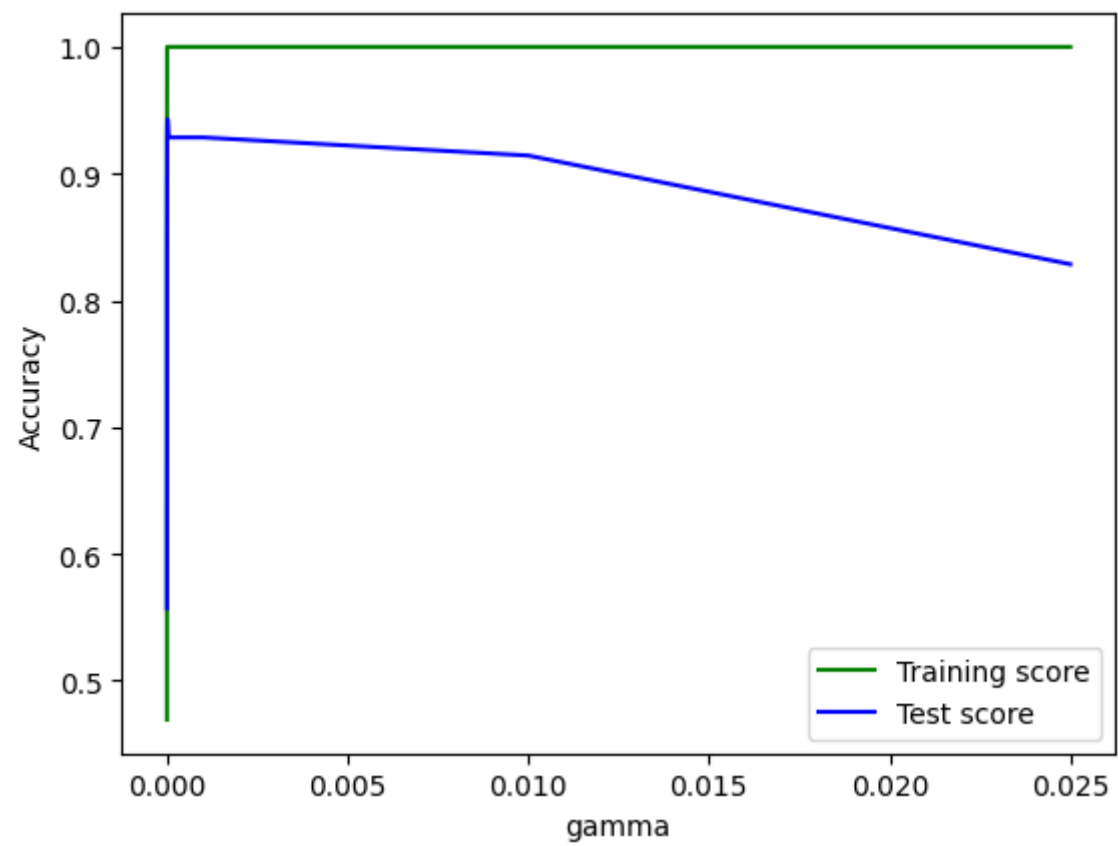
- RandomForestClassifier aus sklearn
- Funktion, die unterschiedliche Anzahl an Estimatoren nimmt
- gibt Classifier sowie Training Accuracy und Test Accuracy zurück
- Plotten der Ergebnisse mit unterschiedlicher Anzahl an Estimatoren

Bewertung

Random Forest hat sich als beste Methode herausgestellt mit einer Testgenauigkeit von ca. 98% und einer Cross-Validation-Score von ca. 94%. Die Kompilierzeit ist sehr schnell und ab ungefähr 20 Estimatoren wird eine sehr hohe Genauigkeit erreicht.



Danach kommen die polynomischen Kernels. Dafür hat die Suche nach geeigneten Hyperparametern sehr lange gedauert. Schließlich wurden im Bereich von ca. 0.000001 bis 0.001 geeignete Ergebnisse gefunden.



mit Kernel-Funktionen

Entscheidungen begründen!!