

Klassifikation mit Python und scikit-learn

Kategorien zuverlässig vorhersagen und Klassifikationsmodelle richtig bewerten

01 INTRO

Klassifikationsverfahren ordnen Beobachtungen vordefinierten Klassen zu und gehören zu den häufigsten Anwendungen des Machine Learning. Die Qualität eines Klassifikators lässt sich dabei selten sinnvoll mit einer einzigen Kennzahl beschreiben.

Der Kurs behandelt den vollständigen Klassifikationsprozess mit Python und scikit-learn. Neben unterschiedlichen Modellansätzen stehen Entscheidungsgrenzen, Wahrscheinlichkeiten, unausgeglichene Klassen und die Auswahl geeigneter Bewertungsmetriken im Mittelpunkt.

DAUER

2 Tage

FORMAT

Präsenz · Live Online · Inhouse

VORKENNTNISSE

Python-Grundkenntnisse und grundlegendes Verständnis von Machine Learning

02 TRAININGSPROGRAMM

Inhalte

Klassifikationsprobleme

- Klassen und Labels
- binäre Klassifikation
- Multiclass
- Multilabel im Überblick
- typische Anwendungsfälle
- Klassifikation versus Regression

Daten für Klassifikation

- Features
- Zielvariable
- Kodierung
- fehlende Werte
- Skalierung
- kategoriale Merkmale
- Datenqualität

Logistische Regression

- Grundidee
- Wahrscheinlichkeit
- Entscheidungsgrenze
- Koeffizienten
- Regularisierung
- Interpretation

Weitere Klassifikationsverfahren

- k-Nearest Neighbors
- Naive Bayes
- Support Vector Machines im Überblick
- Auswahl nach Problemstellung
- unterschiedliche Modellannahmen

Entscheidungsgrenzen und Wahrscheinlichkeiten

- Scores
- Wahrscheinlichkeiten
- Thresholds
- Klassenentscheidung
- Kalibrierung im Überblick
- fachliche Konsequenzen

Confusion Matrix

- True Positive
- False Positive
- True Negative
- False Negative
- Fehlerarten
- Kosten unterschiedlicher Fehlentscheidungen

Klassifikationsmetriken

- Accuracy
- Precision
- Recall
- Specificity
- F1
- Balanced Accuracy
- Auswahl nach Anwendungsfall

ROC und Precision-Recall

- ROC Curve
- AUC
- Precision-Recall Curve
- Threshold-Auswahl
- Grenzen einzelner Kennzahlen
- unausgeglichene Klassen

Imbalanced Data

- Klassenverteilung
- Auswirkungen auf Training und Bewertung
- Gewichtung
- Resampling
- geeignete Metriken
- Datenleckage vermeiden

Validierung und Modellvergleich

- Train-Test-Split
- stratifizierte Aufteilung
- Cross Validation
- Baselines
- Hyperparameter
- Generalisierung

Fehleranalyse

- falsch klassifizierte Beobachtungen
- Segmente mit hoher Fehlerrate
- Feature-Einflüsse
- Thresholds
- Datenprobleme
- Modellgrenzen

Praktisches Klassifikationsprojekt

- Problem formulieren
- Daten vorbereiten
- Modelle trainieren
- Metriken auswählen
- Threshold optimieren
- Fehler analysieren
- Ergebnisse bewerten

03 DURCHFÜHRUNG

Rahmen und Organisation

Dauer	2 Tage
Durchführung	Präsenz oder Live Online, vorzugsweise als Inhouse-Schulung
Schwerpunkt	Inhouse-Schulungen für Unternehmen und Teams
Sprache	Deutsch, englische Fachbegriffe und Dokumentation
Unterlagen	Kursunterlagen, Beispiele und Übungsaufgaben
Technik	Eigener Rechner mit Python, Jupyter und scikit-learn

Inhouse-Schulungen

Eigene Klassifikationsaufgaben und Datenbestände können in die Übungen einbezogen werden.

04 ZIELGRUPPE

Für wen ist der Kurs gedacht?

Data Analysts, Entwickler und angehende Data Scientists, die Klassifikationsmodelle entwickeln und deren Ergebnisse fundiert bewerten möchten.

Voraussetzungen

Grundkenntnisse in Python und Machine Learning.

05 LERNZIELE

Was Sie nach dem Kurs können

Die Teilnehmer können Klassifikationsprobleme modellieren, verschiedene Klassifikatoren einsetzen, geeignete Bewertungsmetriken auswählen und Fehlentscheidungen systematisch analysieren.

06 FAQ

Häufige Fragen

Warum reicht Accuracy nicht aus?

Weil unterschiedliche Fehlerarten und unausgeglichene Klassen zu irreführenden Ergebnissen führen können. Der Kurs behandelt deshalb mehrere Metriken und deren fachliche Bedeutung.

Werden Random Forests behandelt?

Sie können als Vergleichsmodell eingesetzt werden. Bäume und Ensemble-Verfahren sind Schwerpunkt eines eigenen Kurses.

Werden neuronale Netze behandelt?

Nein. Dafür ist der Grundlagenkurs zu neuronalen Netzen vorgesehen.

UC IT Service · Ulrich Cuber · kontakt@uc-it.de

<https://www.uc-it.de/coaching/klassifikation-scikit-learn/>