

Feature Engineering und Dimensionsreduktion

Aus Rohdaten geeignete Merkmale für Machine-Learning-Modelle entwickeln

01 INTRO

Die Qualität eines Machine-Learning-Modells hängt wesentlich davon ab, welche Informationen dem Modell in welcher Form zur Verfügung stehen. Feature Engineering übersetzt Rohdaten und Fachwissen in geeignete Merkmale und kann dadurch einen größeren Einfluss auf die Modellqualität haben als die Wahl eines komplexeren Algorithmus.

Der Kurs behandelt die systematische Entwicklung, Transformation, Auswahl und Reduktion von Features. Dabei stehen sowohl die Verbesserung von Modellen als auch die Vermeidung von Datenleckage und unnötiger Komplexität im Mittelpunkt.

DAUER

2 Tage

FORMAT

Präsenz · Live Online · Inhouse

VORKENNTNISSE

Grundkenntnisse in Python, Pandas und Machine Learning

02 TRAININGSPROGRAMM

Inhalte

Rolle von Features im Machine Learning

- Rohdaten und Modellmerkmale
- Informationsgehalt
- Signal und Rauschen
- Fachwissen
- Feature Space
- Modellabhängigkeit

Numerische Features

- Skalierung
- Standardisierung
- Normalisierung
- Transformationen
- Binning
- robuste Skalierung
- Ausreißer

Kategoriale Features

- Ordinal Encoding
- One-Hot Encoding
- seltene Kategorien
- hohe Kardinalität
- unbekannte Kategorien
- Leakage-Risiken

Fehlende Werte als Feature-Problem

- Imputation
- einfache Strategien
- gruppenabhängige Ansätze
- Missing Indicators
- Auswirkungen auf Modelle
- Training versus Produktion

Features aus vorhandenen Daten entwickeln

- Interaktionen
- Verhältnisse
- Aggregationen
- Datums- und Zeitmerkmale
- Gruppenmerkmale
- Domänenwissen

Feature Engineering für unterschiedliche Datentypen

- numerische Daten
- Kategorien
- Zeitinformationen
- Text im Überblick
- zyklische Daten
- unterschiedliche Verarbeitungspfade

Feature Selection

- Filterverfahren
- modellbasierte Auswahl
- Regularisierung
- Permutation Importance
- redundante Features
- Auswahl und Validierung

Curse of Dimensionality

- hohe Dimensionen
- Distanz
- Datenbedarf
- Overfitting
- Rechenaufwand
- Motivation für Reduktion

Principal Component Analysis

- Grundidee
- Varianz
- Hauptkomponenten
- Projektion
- erklärte Varianz
- Skalierung
- Interpretation

Weitere Verfahren zur Dimensionsreduktion

- Truncated SVD
- manifold-basierte Verfahren im Überblick
- t-SNE
- UMAP im Überblick
- Visualisierung versus Modellierung
- Grenzen der Interpretierbarkeit

Feature Engineering ohne Leakage

- Fit nur auf Trainingsdaten
- Preprocessing und Cross Validation
- Zielinformationen
- zeitliche Leakage
- scikit-learn Transformer
- ColumnTransformer
- Pipelines

Praktisches Feature-Engineering-Projekt

- Ausgangsfeatures untersuchen
- Transformationen entwickeln
- Features auswählen
- Dimensionsreduktion testen
- Modellqualität vergleichen
- Leakage prüfen

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 Datenbestände und vorhandene Modelle können als Ausgangspunkt für die Übungen dienen.

04 ZIELGRUPPE

Für wen ist der Kurs gedacht?

Data Analysts, Entwickler und Data Scientists, die ML-Modelle durch systematische Merkmalsentwicklung und -auswahl verbessern möchten.

Voraussetzungen

Grundkenntnisse in Python, Pandas und Machine Learning.

05 LERNZIELE

Was Sie nach dem Kurs können

Die Teilnehmer können geeignete Features entwickeln und transformieren, Merkmale auswählen, hochdimensionale Daten reduzieren und Feature Engineering ohne Datenleckage in ML-Workflows integrieren.

06 FAQ

Häufige Fragen

Ist PCA ein Feature-Selection-Verfahren?

Nicht im engeren Sinne. PCA erzeugt neue Komponenten aus vorhandenen Merkmalen, während Feature Selection bestehende Merkmale auswählt.

Werden Text-Embeddings behandelt?

Sie können eingeordnet werden. Der Schwerpunkt liegt auf allgemeinem Feature Engineering für klassische Machine-Learning-Verfahren.

Was unterscheidet den Kurs von den ML-Pipelines?

Hier steht die fachliche und methodische Entwicklung der Features im Mittelpunkt. Der Pipeline-Kurs behandelt die technische Konstruktion reproduzierbarer Verarbeitungsketten.

UC IT Service · Ulrich Cuber · kontakt@uc-it.de
<https://www.uc-it.de/coaching/feature-engineering/>