

1. Grundbegriffe

Feature (Merkmal): Messbare Eigenschaft eines Objekts. Beispiel: Pixel, Farbe, Gewicht.
Datenpunkt: Gesamtheit aller Merkmale eines Objekts. Beispiel: Ein einzelnes Bild mit allen Pixelwerten. Label: Richtiger Zielwert beim Training. Wird vom Modell als Feedback genutzt, um die Vorhersage zu verbessern. Klasse: Menge aller möglichen Labels in einem Klassifikationsproblem. Beispiel: Hund, Katze, Pferd. Klassifikation: Zuordnung eines Datenpunkts zu einer Klasse. Beispiel: Bild wird als „Katze“ klassifiziert. Regression: Vorhersage eines kontinuierlichen Zahlenwerts (z. B. Preis, Temperatur).

2. 4 Schritte eines guten Datensatzes

1. Ausreichende Daten pro Klasse 2. Hohe Variabilität innerhalb einer Klasse 3. Konsistenz bei Helligkeit, Farbe, Hintergrund 4. Fehlerbereinigung

3. KI-Zyklus (6 Stufen)

1. Problemstellung 2. Datensammlung 3. Daten sichten und bereinigen (4 Schritte) 4. Modellauswahl und Training 5. Testen und Evaluieren 6. Ausrollen

4. KI-Familien (Lernarten)

Überwachtes Lernen: Mit Labels trainieren, Vorhersagen treffen. Unüberwachtes Lernen: Ohne Labels, Muster finden, Cluster bilden. Verstärkendes Lernen: Lernen durch Belohnung und Bestrafung, Agent handelt optimal.

5. KI-Domänen

Computer Vision: - Typische Daten: Bilder, Videos - Aufgaben: Objekterkennung, Segmentierung, Klassifikation - Beispiele: Gesichtserkennung, Verkehrszeichenerkennung Natural Language Processing (NLP): - Typische Daten: Text, Sprache - Aufgaben: Textklassifikation, Sentimentanalyse, Übersetzung - Beispiele: Chatbots, Übersetzer Tabellendaten (Data): - Typische Daten: Strukturierte CSV/SQL - Aufgaben: Prognosen, Klassifikationen, Mustererkennung - Beispiele: Preisprognosen, Kundenanalyse

6. Evaluationsmethoden und Metriken

Wahrheitsmatrix: TP (True Positive), FP (False Positive), FN (False Negative), TN (True Negative)
Formeln: $\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$ $\text{Precision} = \frac{TP}{TP + FP}$ $\text{Recall} = \frac{TP}{TP + FN}$ $F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
Recall: wenn möglichst viele echte Positive erkannt werden müssen. Precision: wenn Vorhersagen zuverlässig sein müssen.

7. Overfitting und Underfitting

Underfitting: Modell ist zu einfach, erkennt Muster nicht → hoher Trainings- und Testfehler.
Overfitting: Modell lernt Trainingsdaten auswendig → niedriger Trainingsfehler, hoher Testfehler.
Gegenmaßnahmen bei Overfitting: - Dropout - Regularisierung (L2) - Early Stopping - Mehr Daten / Dataaugmentation - Weniger Epochen
Gegenmaßnahmen bei Underfitting: - Komplexeres Modell

- Mehr Features - Längeres Training

8. Training – zentrale Begriffe

Epoche: Ein kompletter Durchlauf des gesamten Trainingsdatensatzes. Batchgröße: Anzahl der Datenpunkte pro Trainingsschritt. Lernrate: Stärke der Gewichtsänderung im Training.