

Block 1: Multiple-Choice-Fragen

Grundlagen & Empirisches Fitting

January 11, 2026

Welche drei Bereiche bilden die Schnittmenge "Data Science"?

- A) Marketing, Vertrieb, Finanzen
- B) Statistik, Informatik, Domänenwissen
- C) Machine Learning, Deep Learning, AI
- D) Datenbanken, Cloud, APIs

Welche drei Bereiche bilden die Schnittmenge "Data Science"?

- A) Marketing, Vertrieb, Finanzen
- B) Statistik, Informatik, Domänenwissen
- C) Machine Learning, Deep Learning, AI
- D) Datenbanken, Cloud, APIs

Lösung: B)

Data Science kombiniert statistische Methoden, informatische Werkzeuge und fachliches Domänenwissen.

Welche Phase des CRISP-DM nimmt typischerweise die meiste Zeit in Anspruch?

- A) Business Understanding
- B) Data Preparation
- C) Modeling
- D) Deployment

Welche Phase des CRISP-DM nimmt typischerweise die meiste Zeit in Anspruch?

- A) Business Understanding
- B) Data Preparation
- C) Modeling
- D) Deployment

Lösung: B)

Data Preparation (inkl. Data Understanding) beansprucht oft 60-80% der gesamten Projektzeit.

Warum sollten Sie relative statt absolute Pfade in R verwenden?

- A) Relative Pfade sind schneller
- B) Relative Pfade ermöglichen Reproduzierbarkeit und Zusammenarbeit
- C) Absolute Pfade werden von R nicht unterstützt
- D) Es gibt keinen Unterschied

Warum sollten Sie relative statt absolute Pfade in R verwenden?

- A) Relative Pfade sind schneller
- B) Relative Pfade ermöglichen Reproduzierbarkeit und Zusammenarbeit
- C) Absolute Pfade werden von R nicht unterstützt
- D) Es gibt keinen Unterschied

Lösung: B)

Relative Pfade funktionieren auf jedem Computer und ermöglichen so Zusammenarbeit und Reproduzierbarkeit.

Welche dplyr-Funktion erstellt neue Spalten in einem Datensatz?

- A) filter()
- B) select()
- C) mutate()
- D) arrange()

Welche dplyr-Funktion erstellt neue Spalten in einem Datensatz?

- A) filter()
- B) select()
- C) mutate()
- D) arrange()

Lösung: C)

mutate() erstellt neue Spalten oder modifiziert bestehende. filter() wählt Zeilen, select() wählt Spalten, arrange() sortiert.

Frage 5: Regressionskoeffizient

In der Regression $\text{Umsatz} = 500 + 25 \times \text{Werbung}$ bedeutet der Koeffizient 25:

- A) Der Umsatz betraegt 25 CHF
- B) Pro zusaetzlichem CHF Werbung steigt der Umsatz um 25 CHF
- C) 25% des Umsatzes kommen von Werbung
- D) Die Korrelation betraegt 0.25

Frage 5: Regressionskoeffizient

In der Regression $\text{Umsatz} = 500 + 25 \times \text{Werbung}$ bedeutet der Koeffizient 25:

- A) Der Umsatz betraegt 25 CHF
- B) Pro zusaetzlichem CHF Werbung steigt der Umsatz um 25 CHF
- C) 25% des Umsatzes kommen von Werbung
- D) Die Korrelation betraegt 0.25

Lösung: B)

Der Koeffizient gibt die Änderung der abhängigen Variable (Umsatz) bei einer Einheit Erhöhung der unabhängigen Variable (Werbung) an.

Ein R^2 von 0.64 bedeutet:

- A) 64% der Datenpunkte liegen auf der Regressionslinie
- B) 64% der Varianz in Y wird durch das Modell erklärt
- C) Die Korrelation betraegt 0.64
- D) Das Modell hat 64 Variablen

Ein R^2 von 0.64 bedeutet:

- A) 64% der Datenpunkte liegen auf der Regressionslinie
- B) 64% der Varianz in Y wird durch das Modell erklärt
- C) Die Korrelation betraegt 0.64
- D) Das Modell hat 64 Variablen

Lösung: B)

R^2 misst den Anteil der Gesamtvarianz in Y, der durch die Prädiktoren erklärt wird. (Hinweis: Die Korrelation waere $r = \sqrt{0.64} = 0.8$)

Was bedeutet "ceteris paribus" bei der Interpretation von Regressionskoeffizienten?

- A) Alle Variablen sind gleich wichtig
- B) Der Effekt gilt nur für bestimmte Fälle
- C) Alle anderen Variablen werden konstant gehalten
- D) Die Koeffizienten sind standardisiert

Was bedeutet “ceteris paribus” bei der Interpretation von Regressionskoeffizienten?

- A) Alle Variablen sind gleich wichtig
- B) Der Effekt gilt nur für bestimmte Fälle
- C) Alle anderen Variablen werden konstant gehalten
- D) Die Koeffizienten sind standardisiert

Lösung: C)

Ceteris paribus (“alles andere gleich”) bedeutet, dass der Koeffizient den Effekt einer Variable zeigt, während alle anderen Variablen konstant gehalten werden.

Ein Interaktionsterm $\text{Preis} \times \text{Marke}$ in der Regression bedeutet:

- A) Preis und Marke sind korreliert
- B) Der Effekt des Preises hängt von der Marke ab
- C) Preis und Marke haben den gleichen Effekt
- D) Einer der beiden Effekte ist nicht signifikant

Frage 8: Interaktionsterm

Ein Interaktionsterm Preis×Marke in der Regression bedeutet:

- A) Preis und Marke sind korreliert
- B) Der Effekt des Preises hängt von der Marke ab
- C) Preis und Marke haben den gleichen Effekt
- D) Einer der beiden Effekte ist nicht signifikant

Lösung: B)

Ein Interaktionsterm modelliert, dass der Effekt einer Variable (Preis) unterschiedlich ist, je nach Ausprägung einer anderen Variable (Marke).

Overfitting tritt auf, wenn:

- A) Das Modell zu wenige Variablen hat
- B) Das Modell die Trainingsdaten zu gut lernt und schlecht generalisiert
- C) Die Daten zu wenig Varianz haben
- D) Die Koeffizienten zu klein sind

Overfitting tritt auf, wenn:

- A) Das Modell zu wenige Variablen hat
- B) Das Modell die Trainingsdaten zu gut lernt und schlecht generalisiert
- C) Die Daten zu wenig Varianz haben
- D) Die Koeffizienten zu klein sind

Lösung: B)

Overfitting bedeutet, dass das Modell das Rauschen in den Trainingsdaten mitlernt und dadurch auf neuen Daten schlecht performt.

Warum ist die Out-of-Sample-Performance wichtiger als die In-Sample-Performance?

- A) Out-of-Sample ist einfacher zu berechnen
- B) Out-of-Sample zeigt die Generalisierungsfähigkeit auf ungesehene Daten
- C) In-Sample-Metriken sind unzuverlässig
- D) Es gibt keinen Unterschied

Warum ist die Out-of-Sample-Performance wichtiger als die In-Sample-Performance?

- A) Out-of-Sample ist einfacher zu berechnen
- B) Out-of-Sample zeigt die Generalisierungsfähigkeit auf ungesehene Daten
- C) In-Sample-Metriken sind unzuverlässig
- D) Es gibt keinen Unterschied

Lösung: B)

In-Sample-Performance ist immer optimistisch, da das Modell die Daten "kennt". Out-of-Sample zeigt, wie gut das Modell auf neue, ungesehene Daten generalisiert.