

1. Welche der folgenden Aussagen zur Abgrenzung von klassischer Programmierung und Maschinellen Lernen sind korrekt?
  - A. In der klassischen Programmierung werden Regeln aus Daten automatisch gelernt.
  - B. Beim Maschinellen Lernen werden Regeln implizit aus Daten und Zielwerten abgeleitet.
  - C. Maschinelles Lernen ist besonders geeignet für Probleme mit vielen schwer formulierbaren Regeln.
  - D. Maschinelles Lernen benötigt keine Daten, solange ein geeignetes Modell gewählt wird.
2. Welche der folgenden Aussagen zum Training und zur Evaluation von Machine-Learning-Modellen sind korrekt?
  - A. Ein sehr geringer Fehler auf den Trainingsdaten garantiert keine gute Generalisierung.
  - B. Die Evaluation auf unabhängigen Testdaten ist notwendig, um die Modellgüte realistisch einzuschätzen.
  - C. Ein Modell mit vielen Parametern generalisiert immer besser als ein einfaches Modell.
  - D. Overfitting tritt häufig auf, wenn ein Modell zu komplex im Verhältnis zur Datenmenge ist.
3. Welche der folgenden Aussagen zu Daten und Merkmalen (Features) im Machine Learning sind korrekt?
  - A. Nicht repräsentative Daten können dazu führen, dass ein Modell systematisch falsche Vorhersagen trifft.
  - B. Irrelevante oder verrauschte Features können die Lernleistung eines Modells verschlechtern.
  - C. Mehr Features führen immer zu besseren Modellen, da mehr Information zur Verfügung steht.
  - D. Die Qualität der Daten ist oft entscheidender für die Modelleleistung als die Wahl des Algorithmus.
4. Welche der folgenden Aussagen zu Lernparadigmen im Machine Learning sind korrekt?
  - A. Beim unüberwachten Lernen werden keine expliziten Zielwerte (Labels) verwendet.
  - B. Reinforcement Learning basiert auf Interaktion mit einer Umgebung und Belohnungssignalen.
  - C. Beim überwachten Lernen werden keine Eingabedaten benötigt.
  - D. Clustering ist eine typische Aufgabe des unüberwachten Lernens.

5. Sie arbeiten an einem Machine-Learning-Projekt mit Sensordaten. Warum sollten Sie den Testdatensatz bereits vor der detaillierten Datenanalyse abtrennen?
  - A. Damit die Trainingsdaten schneller verarbeitet werden können
  - B. Um zu vermeiden, dass implizit Wissen aus dem Testdatensatz in das Modell einfließt
  - C. Damit der Testdatensatz größer wird als der Trainingsdatensatz
  - D. Um eine realistische Abschätzung der Generalisierungsfähigkeit zu erhalten
  - E. Damit der Testdatensatz für Feature Engineering genutzt werden kann
6. Ein Datensatz enthält Zeitreihendaten eines industriellen Prozesses. Wie sollte der Trainings- und Testdatensatz sinnvoll aufgeteilt werden?
  - A. Zufällige Aufteilung, um eine gleichmäßige Verteilung sicherzustellen
  - B. Ältere Daten für Training und neuere Daten für Test verwenden
  - C. Nur Daten mit extremen Werten in den Testdatensatz legen
  - D. Testdaten sollten zukünftige Daten simulieren, um reale Einsatzbedingungen abzubilden
  - E. Trainings- und Testdaten sollten identisch sein, um Vergleichbarkeit zu gewährleisten
7. Sie analysieren ein Merkmal mit stark schiefer Verteilung. Welche Aussagen zur Transformation sind korrekt?
  - A. Eine log-Transformation verschlechtert immer die Modelleleistung
  - B. Eine log-Transformation kann helfen, eine annähernd gaußförmige Verteilung zu erreichen
  - C. Manche ML-Algorithmen profitieren von normalverteilten Merkmalen
  - D. Eine log-Transformation entfernt automatisch Ausreißer vollständig
  - E. Eine log-Transformation ist nur für kategoriale Daten sinnvoll
8. Sie verwenden den OneHotEncoder für ein kategoriales Merkmal. Welche Aussagen treffen zu?
  - A. Kategorien werden als fortlaufende Zahlen kodiert, wodurch eine Ordnung entsteht
  - B. Jede Kategorie wird als eigener binärer Vektor dargestellt
  - C. Der Abstand zwischen Kategorien ist gleich
  - D. OneHotEncoding reduziert die Anzahl der Features immer
  - E. OneHotEncoding darf nur bei ordinalen Daten verwendet werden
9. In einem Datensatz finden sich fehlende Werte und Ausreißer. Welche Vorgehensweisen sind sinnvoll im Rahmen des Data Cleaning?
  - A. Fehlende Werte können durch Imputation (z.B. Median) ersetzt werden
  - B. Ausreißer sollten analysiert werden, da sie auch relevante Informationen enthalten können
  - C. Alle Datensätze mit fehlenden Werten müssen immer gelöscht werden
  - D. Ausreißer sollten grundsätzlich entfernt werden, unabhängig vom Kontext
  - E. Fehlende Werte können auch durch Entfernen von Features behandelt werden
10. Sie evaluieren ein Modell mittels k-facher Kreuzvalidierung. Welche Aussagen sind korrekt?
  - A. Der Trainingsdatensatz wird in k disjunkte Teilmengen aufgeteilt
  - B. In jedem Durchlauf wird eine Teilmenge zur Validierung verwendet
  - C. Der Testdatensatz wird in die Kreuzvalidierung integriert
  - D. Es entstehen k Trainings- und Validierungsdurchläufe
  - E. Kreuzvalidierung ersetzt vollständig die Nutzung eines Testdatensatzes

11. Betrachten Sie das in Abbildung 1 dargestellte Modellierungsergebnis. Dabei stellt die Linie die ermittelte Entscheidungslinie zwischen zwei Klassen dar und jeder Punkt einen Datenpunkt aus den Trainingsdaten. Die roten Kreise gehören zur einen Klasse und die grünen Quadrate zur anderen Klasse.



Abbildung 1: Entscheidungslinie für ein Klassifikationsproblem.

- Bitte markieren Sie die Aussagen, die für das gegebene Modellierungsergebnis zutreffen.
- A. Die Entscheidungslinie wurde mit einem linearen Modellierungsverfahren erstellt.
  - B. Die Modellkomplexität ist zu hoch.
  - C. Das Modell wird gut auf neue Daten generalisieren.
  - D. Das Modell zeigt eine hohe Genauigkeit auf den visualisierten Trainingsdaten.
12. Bitte markieren Sie jene der folgenden Eigenschaften, die für klassische Multi-Layer-Perceptrons zutreffen!
- A. Es gibt zwischen allen Neuronen direkte Verbindungen.
  - B. Für das Anpassen der Gewichte werden die Vorhersagefehler im Netz vorwärts propagiert.
  - C. Der Output eines einzelnen Neurons wird von den Ausgangswerten der mit ihm verbundenen Input-Neuronen, den Gewichten der Eingangsverbindungen und der Aktivierungsfunktion des Neurons bestimmt.
  - D. Das MLP besteht aus einem Input-Layer, mindestens einem Hidden-Layer und einem Output-Layer.
13. Stellen Sie sich vor, Sie arbeiten für ein Reisebüro. Das Reisebüro gibt Ihnen den Auftrag ein Machine Learning Modell zu entwickeln, das für einen Kunden des Reisebüros das perfekte Urlaubsziel vorhersagt. Überlegen Sie sich und beschreiben Sie ausführlich
- welches Modellierungsverfahren (Klassifikation, Regression, Clustering) Sie nutzen würden,
  - welche Eingabedaten Sie benutzen würden,
  - welche Ausgabe das Modell machen soll und
  - wie das Modell in der Vorhersagesituation genutzt werden kann.
14. Ein Türhersteller möchte eine KI gestützte Tür entwickeln, die erkennt, ob die Person vor der Haustür zur Familie oder zum Freundeskreis gehört, oder ob es sich um eine unbekannte Person handelt.
- (a) Welchen Typ von Modellierungsverfahren verwenden Sie dafür?  
 Klassifikation    Clustering    Regression
- (b) Begründen Sie Ihre Antwort ausführlich, indem Sie beschreiben,
- welche Eingabedaten Sie benutzen würden,
  - welche Ausgabe das Modell machen soll und
  - wie das Modell in der Vorhersagesituation genutzt werden kann.

15. Um den Belegungsgrad einer Fähre zu messen, soll mit Hilfe einer Kamera automatisch für jedes an Bord kommende Objekt entschieden werden, ob es sich um eine Fußgänger\*in, Fahrzeug oder Motorrad handelt. Nehmen Sie an, dass die Kamera am Eingang der Fähre angebracht ist und immer nur genau ein Objekt oder gar kein Objekt im Bild der Kamera zu erkennen ist.
- (a) Welchen Typ von Modellierungsverfahren verwenden Sie dafür?  
 Klassifikation    Clustering    Regression
- (b) Begründen Sie Ihre Antwort ausführlich, indem Sie beschreiben,
- welche Eingabedaten Sie benutzen würden,
  - welche Ausgabe das Modell (bspw. ein neuronales Netz) machen soll,
  - wie das Modell in der Vorhersagesituation genutzt werden kann und
  - wie das Training aussehen würde (wie sehen die Trainingsdaten aus? Haben wir Label und falls ja wie sehen diese aus?).
- (c) Sie möchten ein Convolutional neural network (CNN) zur Lösung des Problems nutzen. Geben Sie die Anzahl der Neuronen im Eingangslayer und die Anzahl Neuronen im Ausgangslayer des CNN an. Nehmen Sie dabei an, dass die Kamera eine Auflösung von 640x480 Pixel hat. Um was für eine Art von Layer handelt es sich normalerweise bei dem Ausgangslayer eines CNN?
16. Sie sollen für ein Versicherungsunternehmen die Anzahl der Vertragsabschlüsse im nächsten Jahr vorhersagen. Welchen Typ von Modellierungsverfahren verwenden Sie dafür?  
A. Klassifikation   B. Clustering   C. Regression

17. Betrachten Sie das in Abbildung 2 dargestellte Diagramm eines Trainingsprozesses für ein Regressionsmodell. Dabei stellt die durchgezogene Kurve den quadratischen Vorhersagefehler auf den Trainingsdaten dar, die gestrichelte Kurve den quadratischen Fehler auf den Validierungsdaten und die gepunktete, horizontale Linie den Vorhersagefehler des aktuell im Einsatz befindlichen Regressionsmodells, das durch ein neues, besseres Modell ersetzt werden soll.

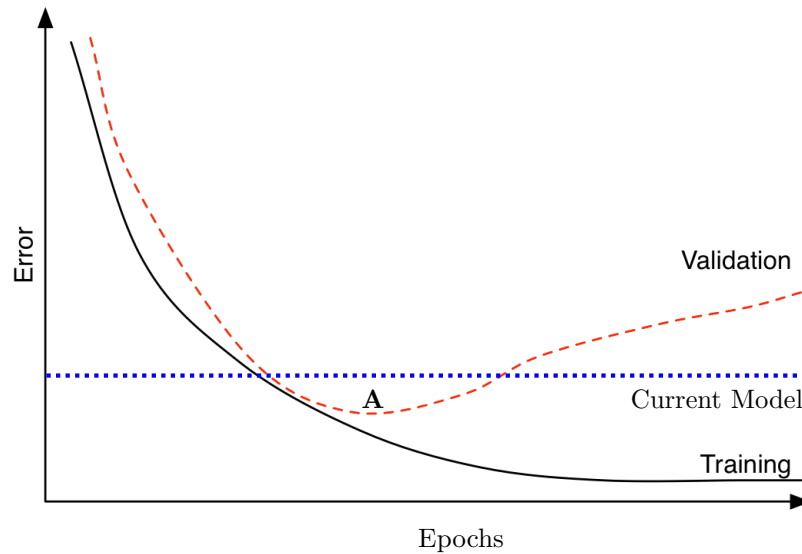


Abbildung 2: Verlauf des Trainings- und Validierungsfehlers eines Regressionsmodells über den Trainingszeitraum.

Bitte markieren Sie, welche **zwei** Eigenschaften für den markierten Punkt A zutreffend sind. Wenn Sie mehr als zwei Antworten ankreuzen, kann die Aufgabe nicht bewertet werden.

- Das betrachtete Modell ist im Overfitting.
- Die Modellkomplexität ist zu niedrig.
- Das Modell wird gut auf neue Daten generalisieren.
- Das Modell hat eine hohe Vorhersagegenauigkeit auf den Trainingsdaten.

18. Um die Empfängerinformation vom Versandetikett auslesen zu können und mit einer Datenbank bekannter Empfänger:innen zu vergleichen, wird das FTS mit einer Kamera ausgestattet. Nehmen Sie an, die Kamera ist so am Greifarm befestigt, dass sie aufnimmt worauf der Greifarm zeigt und somit sichergestellt werden kann, dass nur das Versandetikett eines Päckchens im Bild der Kamera zu sehen ist. Ihre Aufgabe ist es ein Modell zu trainieren, dass die einzelnen Buchstaben eines Namens erkennt. Ein anderer Algorithmus erkennt dabei wo im Bild der Name steht und liefert Ihnen kleine Bildausschnitte der Größe 50x50 Pixel zurück, in denen immer nur ein Buchstabe zu sehen ist.
- (a) Welchen Typ von Modellierungsverfahren verwenden Sie, um den Buchstaben in den 50x50 großen Bildausschnitten zu erkennen?
- Klassifikation    Clustering    Regression
- (b) Begründen Sie Ihre Antwort ausführlich, indem Sie beschreiben,
- welche Eingabedaten Sie benutzen würden,
  - welche Ausgabe das Modell (bspw. ein neuronales Netz) machen soll,
  - wie das Modell in der Vorhersagesituation genutzt werden kann und
  - wie das Training aussehen würde (wie sehen die Trainingsdaten aus? Haben wir Label und falls ja wie sehen diese aus?).
- (c) Sie möchten ein Convolutional Neural Network (CNN) zur Lösung des Problems der Buchstabeerkennung nutzen. Geben Sie die Anzahl der Neuronen im Eingangslayer und die Anzahl der Neuronen im Ausgangslayer des CNN an. Nehmen Sie dabei an, dass die Bildchen, wie oben angegeben, eine Größe von 50x50 Pixeln haben. Um was für eine Art Layer handelt es sich normalerweise bei dem Layer, welches bei einem CNN nach dem Eingangslayer genutzt wird?

19. Neben der Erkennung der Empfängerinformation möchten Sie auch eine Personenerkennung per Spracheingabe realisieren. Sie haben ein neuronales Netz zur Erkennung der Person trainiert und erhalten das in Abbildung 3 dargestellte Diagramm eines Trainingsprozesses. Dabei stellt die durchgezogene Kurve den Loss (bspw. Cross Entropy) auf den Trainingsdaten dar, die gestrichelte Kurve den Loss auf den Validierungsdaten und die gepunktete, horizontale Linie den Loss des aktuell im Einsatz befindlichen Klassifikationsmodells, das durch Ihr neues, besseres Modell ersetzt werden soll.

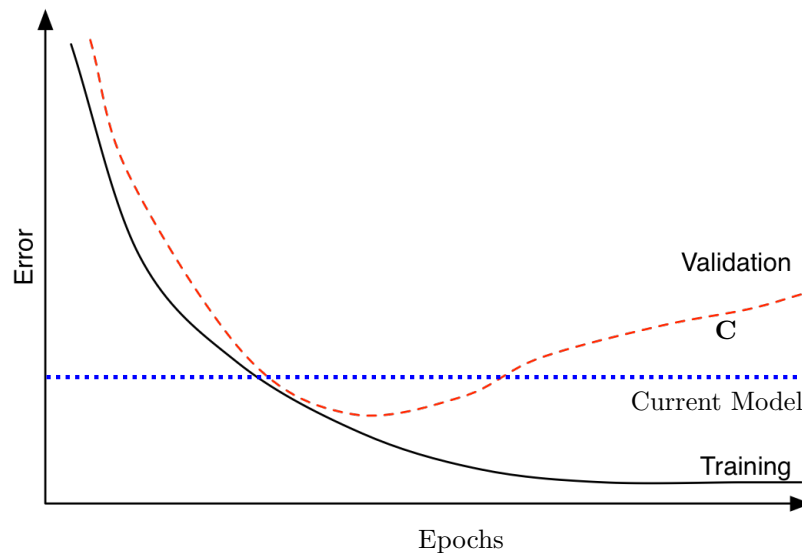


Abbildung 3: Verlauf des Trainings- und Validierungsfehlers eines Klassifikationsmodells über den Trainingszeitraum.

Bitte markieren Sie, welche **zwei** Eigenschaften für den markierten Punkt C zutreffend sind. Wenn Sie mehr als zwei Antworten ankreuzen, kann die Aufgabe nicht bewertet werden.

- Das betrachtete Modell ist im Overfitting.
  - Die Modellkomplexität ist zu niedrig.
  - Das Modell wird gut auf neue Daten generalisieren.
  - Das Modell hat eine hohe Vorhersagegenauigkeit auf den Trainingsdaten.
20. Um zu vermeiden, dass das FTS mit leerem Akku liegen bleibt, möchten Sie den Akkuladezustand ausgehend vom aktuellen Zeitpunkt für 10 Minuten in die Zukunft vorhersagen. Sie können davon ausgehen, dass der zukünftige Akkuladezustand vom Gesamtgewicht der geladenen Päckchen, der aktuellen Geschwindigkeit und dem aktuellen Akkuladezustand des FTS abhängt.
- (a) Welchen Typ von Modellierungsverfahren verwenden Sie, um den zukünftigen Akkuladezustand vorherzusagen?
- Klassifikation
  - Clustering
  - Regression
- (b) Begründen Sie Ihre Antwort ausführlich, indem Sie beschreiben,
- welche Eingabedaten Sie benutzen würden,
  - welche Ausgabe das Modell (bspw. ein neuronales Netz) machen soll,
  - ob die Trainingsdaten gelabelt sind und falls ja wie diese Label aussehen.
- (c) Sie möchten ein Multi-Layer Perceptron (MLP) zur Vorhersage des Akkuladezustandes nutzen. Geben Sie die Anzahl der Neuronen im Eingangslayer und die Anzahl der Neuronen im Ausgangslayer des MLP an. Begründen Sie Ihre Antwort sehr kurz.

21. Sie möchten zusätzlich eine Personenerkennung per Spracheingabe realisieren. Sie haben ein neuronales Netz zur Erkennung der Person trainiert und erhalten das in Abbildung 4 dargestellte Diagramm eines Trainingsprozesses. Dabei stellt die durchgezogene Kurve den Loss (bspw. Cross Entropy) auf den Trainingsdaten dar, die gestrichelte Kurve den Loss auf den Validierungsdaten und die gepunktete, horizontale Linie den Loss auf den Validierungsdaten des aktuell im Einsatz befindlichen Klassifikationsmodells, das durch Ihr neues, besseres Modell ersetzt werden soll.

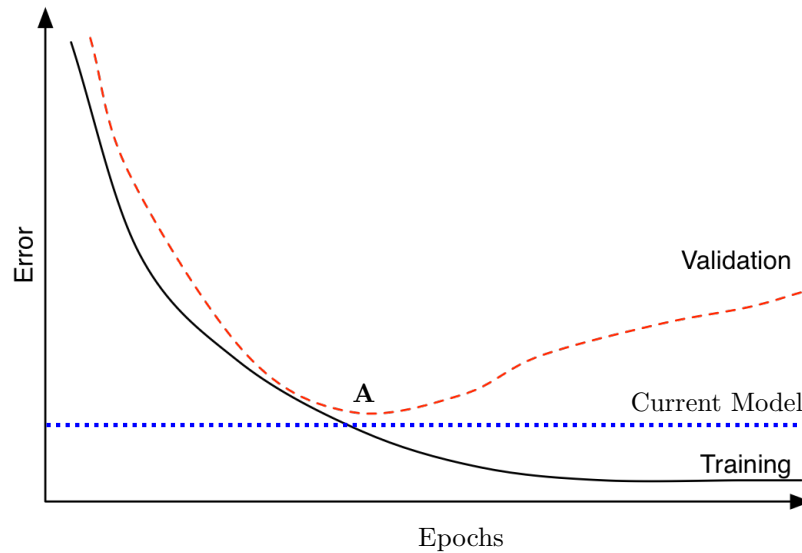


Abbildung 4: Verlauf des Trainings- und Validierungsfehlers eines Klassifikationsmodells über den Trainingszeitraum.

Bitte markieren Sie, welche **zwei** Dinge Sie tun könnten, um ein besser trainiertes neuronales Netz zu erhalten. Wenn Sie mehr als zwei Antworten ankreuzen, kann die Aufgabe nicht bewertet werden.

- Sie können dem neuronalen Netz einen Dropout Layer hinzufügen, um bessere Ergebnisse auf den Validierungsdaten zu erzielen.
- Sie können das trainierte neuronale Netz bei Epoche A (s. Abb.) abspeichern und das aktuell genutzte Modell durch dieses ersetzen.
- Sie können dem Trainingsdatensatz weitere Daten hinzufügen und das Training erneut starten.
- Sie können die Daten aus dem Validierungsdatensatz entfernen für die das neuronale Netz schlechte Ergebnisse liefert.