

Einführung in die künstliche Intelligenz

Wintersemester 2022/23

Probeklausur



Vorwort: Diese Probeklausur entspricht der Klausur aus dem Vorjahr. Die diesjährige Klausur wird nur 60 Minuten lang.

Die vorgefertigten Codeteile liegen im aus den Übungen bekannten Gitlab unter https://git-kik.hs-ansbach.de/2022_wise_ki1/2022_ki1_probeklausur. Der Projektname darf beim Fork in die Gruppe <benutzername>_and_sg nicht geändert werden.

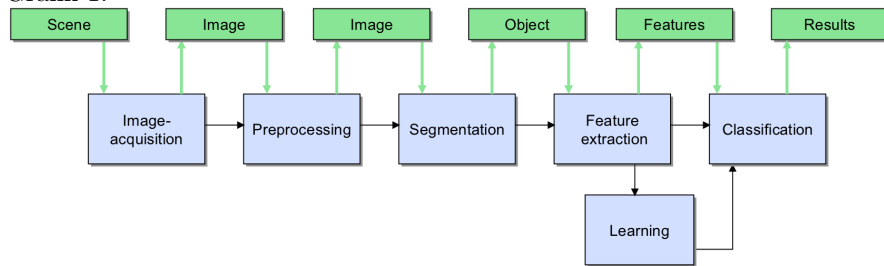
Alternativ kann das Projekt vor Klausurbeginn per USB Stick bezogen werden.

Zur Beantwortung der Theoriefragen ist die Datei `abgabe.csv` zu verwenden. Die Programmieraufgaben sind im jeweiligen Code zu bearbeiten.

Aufgaben:

1. Fülle die allgemeinen Angaben in `abgabe.csv` unter Aufgabe1_Organisatorisches korrekt aus. [1 Punkt]
2. Nenne zwei Beispiele für KI im Internet. [2 Punkte]
3. Was unterscheidet die in der Vorlesung besprochenen tiefen Neuronalen Netze **mindestens** von klassischen Neuronalen Netzen? [2 Punkte]
 - A: Anzahl der Parameter
 - B: Anzahl der Inputs
 - C: Anzahl der Hidden Layer
 - D: Anzahl der Neuronen
4. Welche der folgenden Punkte sind für die in der Vorlesung besprochene Definition von Machine Learning notwendig? [2 Punkte]
 - A: Trainingsdaten
 - B: Parameter
 - C: Anpassen der Parameter an Trainingsdaten
 - D: Modell
5. Wie viele numerische Werte kodieren ein in der Vorlesung besprochenes unkomprimiertes "schwarz-weißes" Graustufenbild mit einer Auflösung von 40x10 Pixeln? [1 Punkt]

Grafik 1:



6. Basierend auf Grafik 1: Beschreibe in zwei bis drei Sätzen, was "Features" sind und was ein gutes Feature ausmacht. [3 Punkte]
7. Basierend auf Grafik 1: Nenne zwei Algorithmen, die beim Preprocessing verwendet werden. [2 Punkte]
8. Basierend auf Grafik 1: Wie müssten die Schritte und Daten in der Grafik heißen, wenn Audiodaten statt Bildern verarbeitet werden? Bevorzuge wo möglich spezifische Begriffe für Audiodaten anstelle generischer Begriffe. Je ein Wort pro Rechteck, nicht alle Bezeichnungen müssen geändert werden. [5 Punkte]
9. Erkläre in ein bis zwei Sätzen, was ein linearer Filter ist. Nenne zwei Beispiele für lineare Filter. [2 Punkte]



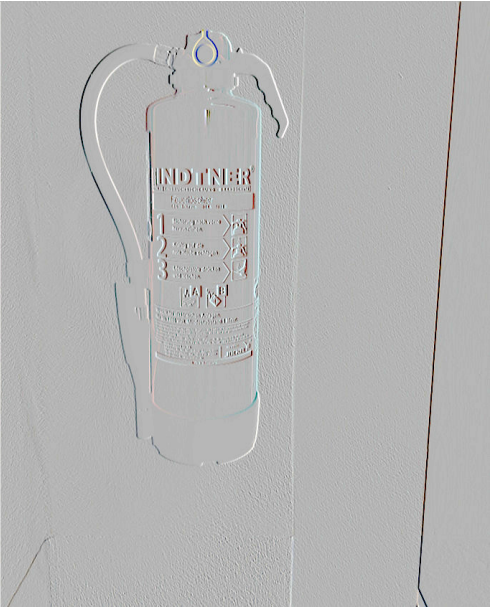
10. Warum evaluiert man in echten Anwendungen die Performance eines Machine Learning Algorithmus möglichst niemals auf Daten, auf denen der Algorithmus trainiert wurde? Wie ist die korrekte Herangehensweise?
Etwa drei Sätze. [3 Punkte]

Grafik 2

a:



b:

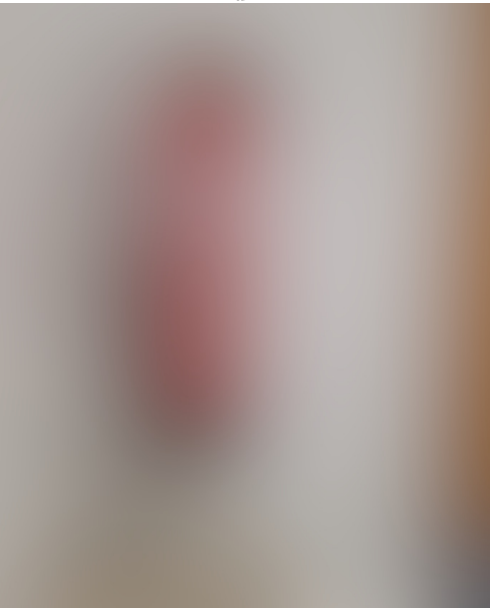


Grafik 2

c:



d:

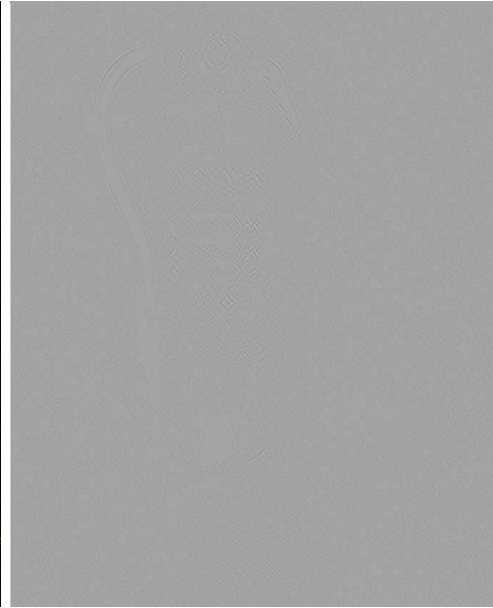


Grafik 2

e:



f:



11. Basierend auf Grafik 2: Identifiziere die verwendeten Filter:

- 2a: Kein Filter
- 2b:
- 2c:
- 2d:
- 2e:
- 2f:

Zur Auswahl stehen:

- Laplacefilter
- Tiefpassfilter 680x680
- Tiefpassfilter 3x3
- Tiefpassfilter 65x65
- Mehrfache Anwendung von Sobelfilter in x, dann y, dann x, dann y
- Sobelfilter x
- Sobelfilter y

Nicht alle Filter sind auf den Bildern gezeigt. Das verwendete Originalbild hat eine Auflösung von 684x845 Pixeln. Hilfreiche Details sind in der zoombaren PDF besser zu erkennen. [5 Punkte]



12. Featureauswahl per Expertenwissen: [2 Punkte]
 Welches der folgenden Feature ist am besten geeignet, um bei fester Distanz zwischen Kamera und Objekt in Bildern Äpfel von Kirschen zu unterscheiden?
- A: Anteil der roten Pixel im segmentierten Objekt (Rotwert > 240)
 - B: Rundheit des segmentierten Objektes
 - C: Anzahl der grünen Pixel (Grünwert > 240)
 - D: Größe des segmentierten Objektes in Pixeln

13. Tabelle 1: Confusion Matrix eines Teilchendetektors

	Bekannt	Unbekannt	Gesamt
Als bekannt klassifiziert	997175	1	997176
Als unbekannt klassifiziert	2823	1	2824
Gesamt	999998	2	1 000 000

Bei diesem Teilchendetektor wird eine Klassifikation eingesetzt, um bekannte Ereignisse von möglicherweise neuen Entdeckungen zu trennen. Neu bzw. unbekannt ist also die "Positive" Klasse.

- a) Berechne die True Positive Rate. [1 Punkt]
 - b) Berechne die True Negative Rate. [1 Punkt]
 - c) Trotz der relativ guten TPR von $\geq 45\%$ und der sehr guten TNR von $\geq 90\%$ ist es gut möglich, dass die korrekt als unbekannt klassifizierte Beobachtung unentdeckt bleibt. Warum? (Antwort in Conf. Matrix enthalten, ein bis zwei Sätze) [2 Punkte]
14. Programmieraufgabe in Datei *main.py*: Angenommen die Daten, die geladen und geplottet werden, sind Audiodaten: Bestimme einen Schwellwert *division_threshold*, der jedes gesprochene Wort mindestens einmal detektiert, aber möglichst nicht auf das Rauschen des Mikrofons anspringt.
 Hinweis: Die Antwort muss nur im korrekten Wertebereich liegen, nicht den best-möglichen Schwellwert darstellen. [3 Punkte]

15. Programmieraufgabe in *machine_learning.py*: [7 Punkte]

Die im Code bereits vorgegebene Funktion *dummy_divide* legt eine Gruppeneinteilung der Datenpunkte fest. Im Unterordner *./doc* findet sich die Dokumentation der Bibliothek scikit-learn bzw. sklearn zu drei Machine Learning Verfahren. Wähle unter diesen Verfahren den geeignetsten Algorithmus, um die Eigenschaften der Gruppen zu lernen und Daten den zwei Gruppen zuzuordnen. Implementiere das Verfahren unter Zuhilfenahme der Dokumentation und gib das Ergebnis als numpy array zurück.

16. Programmieraufgabe in *filter.py*:

- a) Schreibe in der Funktion *filter_data* einen eindimensionalen Medianfilter mit Breite 5. Hinweis: Code aus *helper_functions.py* könnte nützlich sein. Wende diesen Filter auf den eindimensionalen Input an und gib das Ergebnis als numpy array zurück. [6 Punkte]
- b) Erweitere den Filter in der Funktion *filter_with_edge* durch eine selbstgewählte Randbehandlung. Es gibt nur zwei Anforderungen: [3 Punkte]
 - Gleich viele Datenpunkte im gefilterten Output wie im Input
 - Gefilterte Daten abseits des vom Rand beeinflussten Bereichs haben die gleichen Werte wie ohne Randbehandlung gefilterten Daten

Fehler im Pycodestyle geben keinen Abzug, aber für volle Punktzahl muss der jeweilige Code ohne Laufzeitfehler ausführbar sein. Die Abgabe des Source Codes sowie der Datei *abgabe.csv* soll wie aus den Übungsaufgaben bekannt als Commit im geforkten Gitlabprojekt erfolgen [1 Punkt].

Alternativ kann auf Wunsch per USB Stick oder in Papierform abgegeben werden.

Bearbeitungszeit 90 Minuten.

Viel Erfolg!