

Probeklausur “Grundlagen des Maschinellen Lernens” WS 2022/23

Mittowch, 25. Januar 2023

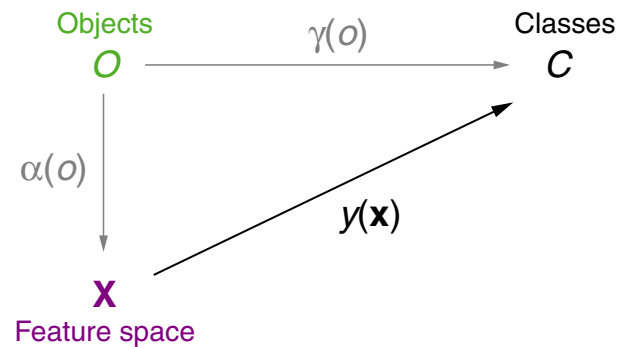
Studiengang:

- Die Bearbeitungszeit beträgt 60 Minuten. Zu Beginn werden 10 Minuten Einlesezeit gewährt.
- Prüfen Sie die Vollständigkeit Ihres Klausurexemplares (Deckblatt, 6 Aufgaben, 13 Blätter).
- Schreiben Sie zu Beginn der Klausur auf das Deckblatt und jedes weitere Blatt Ihren Namen, Ihren Vornamen und Ihre Matrikelnummer.
- Verwenden Sie keinen Rotstift, keinen Bleistift und kein eigenes Konzeptpapier. Benutzen Sie nur den Platz unter den Aufgabenstellungen und die Rückseiten der Blätter für Ihre Lösungen. Sollten Sie zusätzliches Konzeptpapier benötigen, so wenden Sie sich bitte an die Aufsicht.
- Werden zu einer (Teil-)Aufgabe mehrere alternative Lösungen angegeben, gilt die Aufgabe als nicht gelöst. Nicht zu bewertende Bestandteile sind eindeutig zu kennzeichnen.
- Als Hilfsmittel ist ein einseitig eigenhändig handbeschriebenes DIN-A4 Blatt zugelassen. Dieses Blatt wird mit eingesammelt; schreiben Sie auch auf dieses Blatt Ihren Namen und ihre Matrikelnummer. Außerdem ist ein nicht-programmierbarer Taschenrechner erlaubt.
- Viel Erfolg!

Aufgabe	1	2	3	4	5	6	Σ
Erreichbare Punkte	5	6	8	8	10	10	50
Erreichte Punkte							

Aufgabe 1 : Grundlagen & Evaluation (2+3=5 Punkte)

- (a) Gegeben ist die folgende Abbildung, welche die Bestandteile der Spezifikation von Lernaufgaben aufzeigt. Vervollständigen sie die darunter stehende Tabelle am Beispiel des Problems der Klassifikation von Tieren zu verschiedenen Arten anhand von erhobenen Messwerten.



Bestandteil	Beispiel
Objects O	zu klassifizierende Tiere numerische Messwerte zu Gewicht, Größe, und Körpertemperatur Menge der zu unterscheidenden Tierarten fachkundige:r Zoolog:in, welche:r Tiere identifiziert Waage, Thermometer, Maßstab
$\gamma(\mathbf{x})$	zu lernender Klassifikator

Name:	Matrikelnummer.:	3 / 13
-------	------------------	--------

- (b) Gegeben ist die folgende Menge D von Beispielen, die Tiere (c) durch ihr Körpergewicht in Kilogramm (X) charakterisieren:

Beispiel	X	c
1	3	Katze
2	4	Hund
3	2	Katze
4	1	Hund
5	2	Katze
6	5	Hund

Sie möchten ein Klassifikationsmodell mit k -facher Kreuzvalidierung auf diesem Datensatz evaluieren, wobei der Wert k auf 3 gesetzt ist.

- (b1) Wie viele Klassifikatoren werden trainiert?
- (b2) Wie viele Beispiele befinden sich je in der *Trainingsmenge* D_{tr} und *Testmenge* D_{test} für jeden dieser Klassifikatoren?
- (b3) Nehmen Sie an, dass Ihr Klassifikationsmodell nach der folgenden Gleichung funktioniert:

$$y(x) = \begin{cases} \text{Katze, wenn} & x < 3 \\ \text{Hund, wenn} & x \geq 3 \end{cases}.$$

Berechnen Sie die wahre Fehlerrate $Err^*(y)$ [*True Misclassification Rate*] dieses Klassifikators unter der Annahme, dass alle Beispiele in D zur Evaluation verwendet werden.

Name:	Matrikelnummer.:	4 / 13
-------	------------------	--------

Aufgabe 2 : Concept Learning (6 Punkte)

Betrachten Sie den gegebenen Merkmalsraum für das Zielkonzept *EnjoySurfing* mit den folgenden sechs Merkmalen und ihren jeweiligen Domänen:

$Sky = \{sunny, rainy\}$, $Temperature = \{warm, cold\}$, $Humidity = \{normal, high\}$,
 $Wind = \{strong, weak\}$, $Water = \{cold, warm\}$ und $Forecast = \{same, change\}$.

Mit dem Kandidaten-Eliminierungs-Algorithmus [*Candidate Elimination*] soll das Konzept *EnjoySurfing* gelernt werden, das den Wert 0 für negative Merkmalsvektoren \mathbf{x} [*feature vectors*] annimmt und den Wert 1 für positive. Der Lernalgorithmus hat bereits Beispiele verarbeitet, die zu den folgenden Mengen H_S und H_G geführt haben:

$H_S = \{(sunny, warm, normal, ?, warm, same), (sunny, ?, normal, weak, ?, same)\}$
 $H_G = \{(sunny, warm, ?, ?, ?, same), (sunny, ?, normal, ?, ?, same)\}$

Der nächste Merkmalsvektor hat die folgenden Werte:

$\mathbf{x}_1 = (sunny, cold, normal, weak, cold, same) \quad EnjoySurfing(\mathbf{x}_1) = 0$

Führen Sie den nächsten Schritt des Kandidaten-Eliminierungs-Algorithmus [*Candidate Elimination*] am Beispiel \mathbf{x}_1 aus. Schreiben Sie den Endzustand von H_S und H_G auf, sowie jeden Zwischenschritt, um zu begründen, warum Sie Hypothesen zu einer dieser Mengen hinzufügen oder entfernen.

Name:	Matrikelnummer.:	5 / 13
-------	------------------	--------

Aufgabe 3 : Lineare Modelle (6+2=8 Punkte)

- (a) Betrachten Sie den Incremental-Gradient-Descent-Algorithmus (IGD) für das Trainieren linearer Modelle wie unten dargestellt.

IGD(D, η)

```

1. initialize_random_weights( $\mathbf{w}$ ),  $t = 0$ 
2. REPEAT
3.    $t = t + 1$ 
4.   FOREACH  $(\mathbf{x}, c) \in D$  DO
5.      $y(\mathbf{x}) =$  
6.      $\delta = c(\mathbf{x}) - y(\mathbf{x})$ 
7.      $\Delta \mathbf{w} = \eta \cdot \delta \cdot \mathbf{x}$ 
8.      $\mathbf{w} = \mathbf{w} + \Delta \mathbf{w}$ 
9.   ENDDO
10. UNTIL (convergence( $D, y(), t$ ))
11. return( $\mathbf{w}$ )

```

- (a1) Die Modellfunktion $y(\mathbf{x})$ erscheint in Zeile 5 des Algorithmus. Schreiben Sie die Gleichung für die Modellfunktion auf, wenn $y(\mathbf{x})$ ein logistisches Regressionsmodell ist.

- (a2) Beschreiben Sie kurz den Zweck der Zeilen 6 und 7 im Algorithmus.

Name:	Matrikelnummer.:	6 / 13
-------	------------------	--------

- (a3) Für die lineare Regression lautet die Modellfunktion $y(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$. Erklären Sie, warum bei der Anpassung eines linearen Regressionsmodells anstelle eines logistischen Regressionsmodells mit IGD der Algorithmus – und insbesondere die Zeilen 5 und 6 – gleich bleibt.

- (b) Betrachten Sie das reellwertige lineare Regressionsmodell $y(\mathbf{x}) = w_0 + \sum_{j=1}^p w_j \cdot x_j = \mathbf{w}^T \mathbf{x}$.

- (b1) Sei $\gamma()$ eine Zielfunktion mit Werten aus der Menge $\{-1, 1\}$. Überführen Sie $y()$ in einen Klassifikator $\hat{y}()$, der ebenfalls ausschließlich Werte aus dieser Menge annimmt.

$$\gamma(\mathbf{x}) \approx \hat{y}(\mathbf{x}) = \boxed{}$$

- (b2) Geben Sie einen mathematischen Term (Name und Formel) an, um die Anpassungsgüte [*goodness of fit*] des linearen Regressionsmodells zu beurteilen.

Name:	Matrikelnummer.:	7 / 13
-------	------------------	--------

Aufgabe 4 : Bayesian Learning (1+2+5=8 Punkte)

(a) Zeigen sie den Zusammenhang von Prior-, Posterior-, und Likelihood-Wahrscheinlichkeiten.

(b) Sei \mathbf{X} ein p -dimensionaler Merkmalsraum [*feature space*], sei C die Menge der k Klassen eines Zielkonzepts, und sei D eine Menge von Beispielen über $\mathbf{X} \times C$. Hierbei bezeichnen A_1, \dots, A_k Ereignisse der Art $C = c_i$, und $B_j, j = 1, \dots, p$ bezeichnen Ereignisse der Art $X_j = x_j$.

Welche Wahrscheinlichkeiten sind zu schätzen, um einen Naive Bayes-Klassifikator für das Zielkonzept zu konstruieren?

(c) Gegeben seien die folgenden fünf Trainingsbeispiele in einem zweidimensionalen Merkmalsraum [*feature space*]:

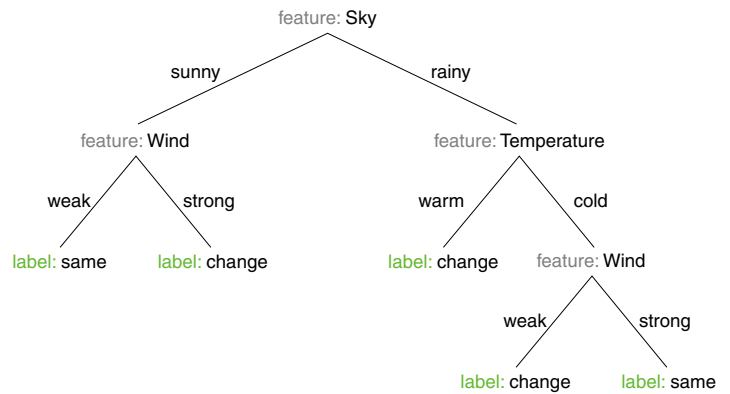
Example	x_1	x_2	c
1	1	0	0
2	0	1	1
3	1	0	1
4	0	1	0
5	1	0	0

Berechnen Sie auf der Grundlage dieses Datensatzes die Prior- und Posterior-Wahrscheinlichkeiten, die zur Spezifikation eines Naive Bayes-Klassifikators erforderlich sind.

Aufgabe 5 : Entscheidungsbäume (2+5+2+1=10 Punkte)

- (a) Gegeben ist der folgende Datensatz sowie ein daraus konstruierter Entscheidungsbaum für Wettervorhersagen.

Sky	Temp.	Wind	Forecast
sunny	warm	weak	same
sunny	warm	strong	change
sunny	cold	weak	same
sunny	cold	strong	change
rainy	warm	weak	change
rainy	warm	strong	change
rainy	cold	weak	change
rainy	cold	strong	same



Geben Sie für diesen Entscheidungsbaum die folgenden Größen an:

- (a1) Anzahl der Blattknoten [*leaf node number*]:
 - (a2) Höhe des Baumes [*tree height*]:
 - (a3) Externe Pfadlänge [*external path length*]:
 - (a4) Gewichtete externe Pfadlänge [*weighted external path length*]:
- (b) Die Interpretation einer Booleschen Formel hängt von den Wahrheitszuweisungen ihrer booleschen Variablen ab und ist entweder “wahr” (1) [*true*] oder “falsch” (0) [*false*]. Betrachten Sie den Wahrheitswert der Formel als das Klassenlabel für den jeweiligen Vektor x der Wahrheitszuweisungen für die Booleschen Variablen. Ihre Aufgabe ist es, einen Entscheidungsbaum zu konstruieren, der die Vektoren der Wahrheitszuweisungen für die Boolesche Formel $(A \vee \neg B) \wedge C$ korrekt klassifiziert.
- (b1) Stellen Sie die Trainingsmenge für Ihren Entscheidungsbaum [*decision tree*] als eine Tabelle von Beispielen mit den Merkmalen A , B und C und dem jeweiligen Klassenwert zusammen.

Name:	Matrikelnummer.:	9 / 13
-------	------------------	--------

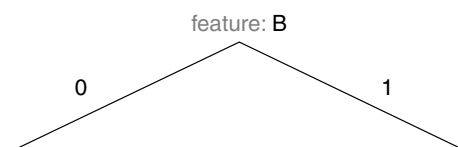
- (b2) Identifizieren Sie für die Teilung an der Wurzel [*root split*] des Baums (aus den Merkmalen *A*, *B* und *C*) das Merkmal mit der maximalen Unreinheitsreduktion [*impurity reduction*] $\Delta\iota$. Verwenden Sie das Unreinheitsmaß [*impurity function*] basierend auf der Fehlklassifikationsrate [*misclassification rate*], $\iota_{misclass}$, als Teilungskriterium [*splitting criterion*].

Hinweis:

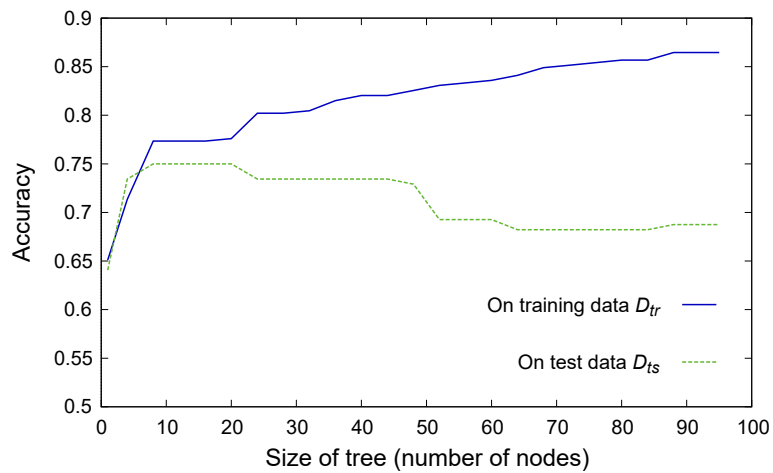
$$\iota_{misclass}(D) = 1 - \max \left\{ \frac{|\{(\mathbf{x}, c(\mathbf{x})) \in D : c(\mathbf{x}) = c_1\}|}{|D|}, \frac{|\{(\mathbf{x}, c(\mathbf{x})) \in D : c(\mathbf{x}) = c_2\}|}{|D|} \right\}$$

$$\Delta\iota = \iota_{misclass}(D) - \left(\frac{|D_1|}{|D|} \cdot \iota_{misclass}(D_1) + \frac{|D_2|}{|D|} \cdot \iota_{misclass}(D_2) \right)$$

- (b3) Der folgende Entscheidungsbaum verwendet Merkmal *B* für die erste Aufteilung. Vervollständigen Sie die Konstruktion dieses Entscheidungsbaums, so dass er alle Beispiele korrekt klassifiziert, und verwenden Sie dabei Merkmale Ihrer Wahl für die restlichen Aufteilungen. D.h., zeichnen Sie den resultierenden Baum und markieren Sie die Teilungsmerkmale an den inneren Knoten, die Werte der Teilungsmerkmale entlang der entsprechenden Kanten und die Klassen an den Blättern.



(c) Bei der Konstruktion eines Entscheidungsbaums können wir die folgende Abbildung erstellen:



(c1) Auf welches Problem weist die zunehmende Differenz zwischen der blauen und der grünen Kurve hin?

(c2) Erklären Sie, wie das Stutzen [*pruning*] von Entscheidungsbäumen für die Lösung dieses Problems verwendet werden kann. Verwenden Sie die Abbildung aus der obigen Frage wieder.

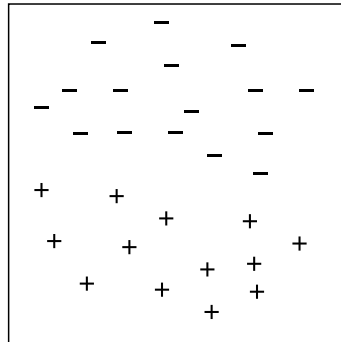
(d) Ist es angesichts eines beliebigen endlichen widerspruchsfreien Datensatzes D immer möglich, einen Entscheidungsbaum zu konstruieren, der alle Beispiele aus D korrekt klassifiziert? Erläutern Sie Ihre Antwort.

Name:	Matrikelnummer.:	11 / 13
-------	------------------	---------

Aufgabe 6 : Neurale Netzwerke (1+1+8=10 Punkte)

- (a) Erfordert ein Multilayer-Perzeptron, dass eine Beispielmeng D linear separierbar [*linearly separable*] ist, um sie korrekt zu klassifizieren?

- (b) Die folgende Abbildung zeigt eine Menge von Punkten in einem zweidimensionalen Merkmalsraum [*feature space*], die zu zwei verschiedenen Klassen gehören.



Wird der Perzeptron-Algorithmus (PT) schlussendlich zu einer Hyperebene konvergieren, die alle Punkte in diesem Datensatz korrekt klassifiziert? Erläutern Sie kurz Ihre Antwort.

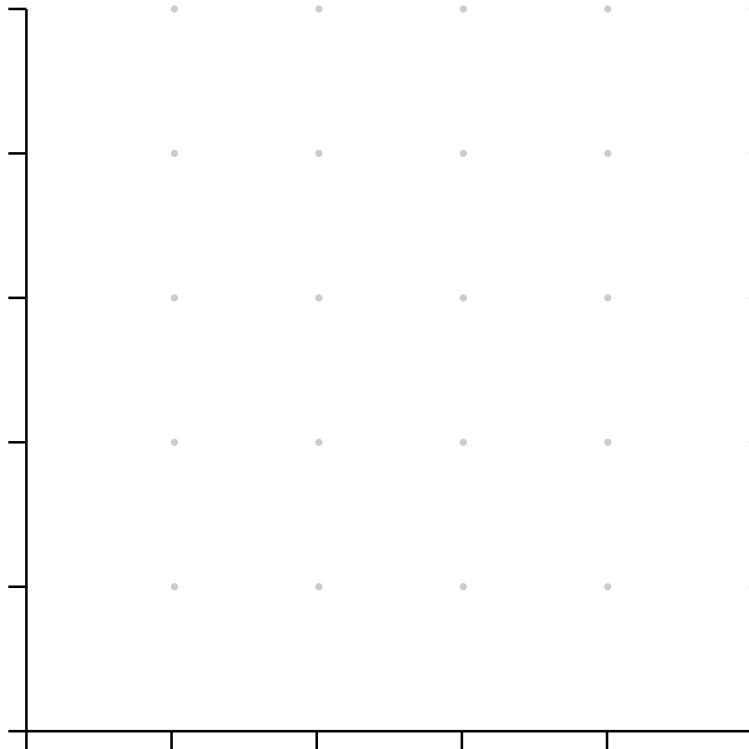
Name:	Matrikelnummer.:	12 / 13
-------	------------------	---------

- (c) Betrachten Sie ein Perzeptron mit zwei Eingängen und der Heaviside-Funktion H als Aktivierungsfunktion [*activation function*]. Gegeben ist weiterhin der folgende Datensatz, der 4 Punkte \mathbf{x}_i und entsprechende Klassen c_i enthält:

$$\begin{aligned}\mathbf{x}_1 &= (2, 1), & c_1 &= 1 \\ \mathbf{x}_2 &= (4, 1), & c_2 &= 0 \\ \mathbf{x}_3 &= (2, 3), & c_3 &= 1 \\ \mathbf{x}_4 &= (4, 3), & c_4 &= 1\end{aligned}$$

Die Gewichte des Perzeptrons werden initialisiert durch $(w_0, w_1, w_2)^T = (-3, -1, 3)^T$

- (c1) Skizzieren Sie die Datenpunkte und die Hyperebene, wie sie durch das initialisierte Perzeptron definiert sind, in das Koordinatensystem. (Sie können die Klasse 1 mit einem Pluszeichen und 0 mit einem Minuszeichen im Graph eintragen).



- (c2) Erklären Sie das Bild in Bezug auf die Klassifikation und die Fehlklassifizierungsrate [*misclassification rate*].

Name:	Matrikelnummer.:	13 / 13
-------	------------------	---------

(c3) Betrachten Sie den Perceptron Training-Algorithmus. Der Punkt \mathbf{x}_1 wird ausgewählt. Berechnen Sie einen Aktualisierungsschritt für die Gewichte, wie er durch den Perceptron Training-Algorithmus definiert ist. Nehmen Sie eine Lernrate von $\eta = 0.5$ an.

(c4) Zeichnen Sie die aktualisierte Hyperebene in das Koordinatensystem von Teilaufgabe 1 ein und interpretieren Sie, was sich in Bezug auf die Klassifikation und Fehlklassifikationsrate [*misclassification rate*] geändert hat.