

Von einigen Patienten wissen wir, ob sie an einer gewissen Krankheit erkrankt sind oder nicht. Zudem kennen wir die Blutwerte A und B dieser Patienten. Diese scheinen gute Indikatoren dafür zu sein, ob der Patient krank ist oder nicht. Ziel des Arbeitsblattes ist es, basierend auf den Blutwerten von Patienten, von denen wir wissen, ob sie erkrankt sind oder nicht, ein Klassifizierungsmodell zu entwickeln. Dieses soll dann genutzt werden, um bei neuen Patienten schnell bestimmen zu können, ob diese vermutlich krank sind oder nicht.

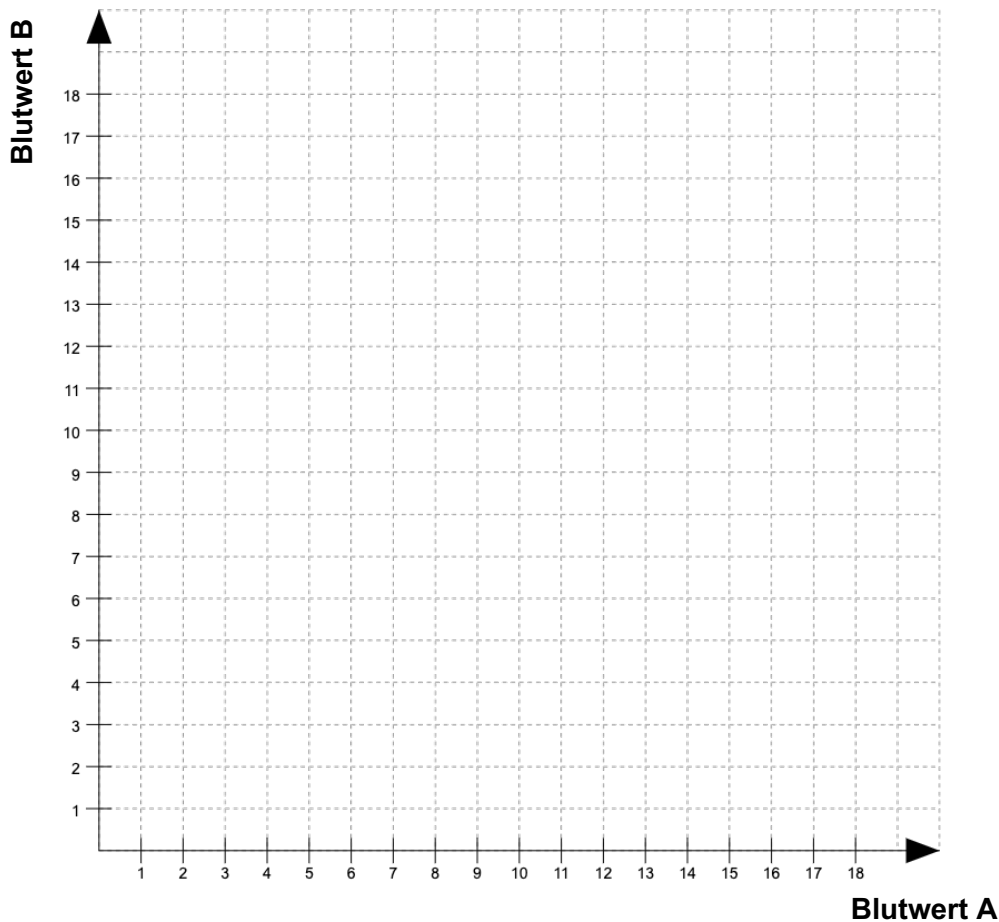


Aufgabe 1

Übertrage die **Lerndaten** (nicht die Testdaten) aus der Tabelle in das Koordinatensystem. Nutze beim Einzeichnen der Datenpunkte für jede Klasse eine eigene Farbe.

Tabelle 1: Der Datensatz (G = gesund, K = erkrankt)

	Lerndaten														Testdaten					
Blutwert A	4	4	5	5	5	5,5	6	6	10	10	10,5	11	11	11,5	12	5,5	11	10	8	5
Blutwert B	14	15	14	15	16	14,5	15	16	11	12	10	12	11	10,5	12	15,5	13	14	15	17
Klasse (Label)	G	G	G	G	G	G	G	G	K	K	K	K	K	K	K	G	K	K	G	K



Aufgabe 2

- Zeichne eine Gerade in das Koordinatensystem. Die Gerade soll die Datenpunkte der beiden Klassen voneinander trennen. Bestimme die Funktionsvorschrift deiner Trenngeraden und notiere sie in Tabelle 2. Wir nennen die Gerade auch **Trenngeraden**.

- b. Zeichne zusätzlich die Gerade mit der Funktionsvorschrift $y = \frac{5}{3}x + 4$ und die Gerade mit der Funktionsvorschrift $y = 13,5$ in das Koordinatensystem ein.

Tabelle 2: Trenngeraden

	Funktionsvorschrift
Gerade 1 (deine Gerade)	$y =$
Gerade 2	$y = 13,5$
Gerade 3	$y = \frac{5}{3}x + 4$

Aufgabe 3

- a. Welche der Geraden trennt die beiden Klassen **am besten**? Begründe deine Antwort.

- b. Notiere für deine Mitschüler:innen ein Kriterium / eine Arbeitsanweisung mithilfe derer sie die beste Gerade finden können.

Wähle die Gerade, die ...

- c. Zeichne die **Testdatenpunkte** aus Tabelle 1 in das Koordinatensystem ein. Welcher Klasse teilen die drei Geraden die jeweiligen Testdatenpunkte zu? Trage die Anzahl der richtigen und falschen Vorhersagen der Geraden in Tabelle 3 ein. Welche Gerade liefert die besten Ergebnisse bezüglich der Testdaten? Begründe.

Tabelle 3: Anzahl richtig / falsch klassifizierter Testdatenpunkte

Gerade 1	Richtig:	Falsch:
Gerade 2	Richtig:	Falsch:
Gerade 3	Richtig:	Falsch:

Deine Begründung:

Bisher haben wir durch „Hinschauen“ bestimmt, welcher Klasse ein Datenpunkt von einer Trenngeraden zugeordnet wird. Das Klassifizieren durch „Hinschauen“ ist in unserem kleinen Beispiel noch möglich. Müssen jedoch **sehr viele** Datenpunkte klassifiziert werden, sollte die Klassifizierung automatisiert werden.

Aufgabe 4

- a. Notiere in eigenen Worten, wie mithilfe der besten Trenngeraden eine Voraussage für einen neuen Punkt gemacht werden kann.
- b. Berechne mithilfe der besten Trenngeraden, welcher Klasse der Punkt (8 | 21) zugeordnet wird. **Hinweis:** Für diese Teilaufgabe liegen zwei Tippkarten bereit.

Fazit: Mithilfe der **besten** Trenngeraden können wir die Klassenzuordnung von Datenpunkten vorhersagen, bei denen die Zuordnung vorher noch gänzlich unbekannt ist. Also von Patienten, bei denen wir zuvor nicht wissen, ob sie erkrankt sind oder nicht. Auf diesem Arbeitsblatt habt ihr wichtige Schritte kennengelernt, die ganz typisch sind für Methoden aus dem Bereich **Künstliche Intelligenz!** Diese Schritte werden in der nächsten Aufgabe nochmal zusammengefasst.

Aufgabe 5

Bringe die folgenden Schritte in eine sinnvolle Reihenfolge. Der erste Schritt ist bereits eingetragen.

- a. Mithilfe der Testdaten bewerten, wie gut das Modell zur Klassifizierung neuer Daten geeignet ist.
- b. Wenn das Modell gut funktioniert, kann dieses auf gänzlich unbekannte Daten angewandt werden.
- c. Unterteilen der Daten in Lerndaten und Testdaten.
- d. Festlegung eines Bewertungskriteriums (*Wann ist unsere Trenngerade „gut“?*)
- e. Bestimmung der optimalen Trenngerade.
- f. Entwicklung eines mathematischen Modells (in unserem Fall: Trenngerade)

1. **Schritt:** Sammeln / Erheben und Aufbereiten von Daten. Daten den Klassen zuordnen.¹

2. **Schritt:**

3. **Schritt:**

4. **Schritt:**

5. **Schritt:**

6. **Schritt:**

7. **Schritt:**

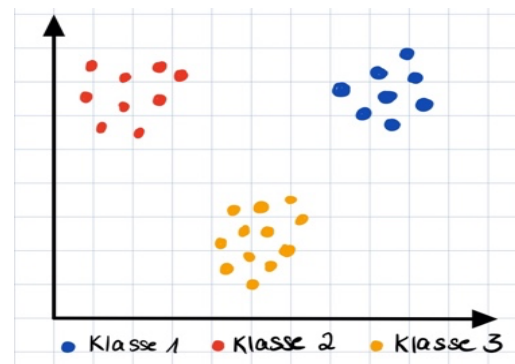
¹ Das Zuordnen zur entsprechenden Klasse wird auch als „Labeln der Daten“ bezeichnet.

Zusatzaufgabe 1

Notiere Ideen, wie wir neben der Zuordnung zu einer Klasse zudem berechnen können, wie sicher wir uns mit der Zuordnung sind.

Zusatzaufgabe 2

Notiere Ideen, wie unser bisheriges Vorgehen auf **Probleme mit drei Klassen** erweitert werden kann. Ein möglicher Datensatz ist in der folgenden Skizze dargestellt.



Zusatzaufgabe 3

Notiere Ideen, welche Probleme bei der Entwicklung eines Klassifizierungsmodells basierend auf Messdaten auftreten können.

Tipp: Schau dir bspw. den Testdatenpunkt (5 | 17) nochmal genauer an.