

In einem Tierheim finden streunende Katzen und Hunde ein vorübergehendes Zuhause. Die Tiere werden beim Empfang gewogen und gemessen. Zudem wird vermerkt, ob es sich bei den Tieren um Hund oder Katze handelt. Folgende Graphen zeigen die Daten der Tiere. Um Zeit und Geld zu sparen, soll das Kategorisieren nun automatisiert werden. Als Entscheidungskriterium wird erstmal das Gewicht der Tiere verwendet. Doch wo genau setzt man die Entscheidungsgrenze?

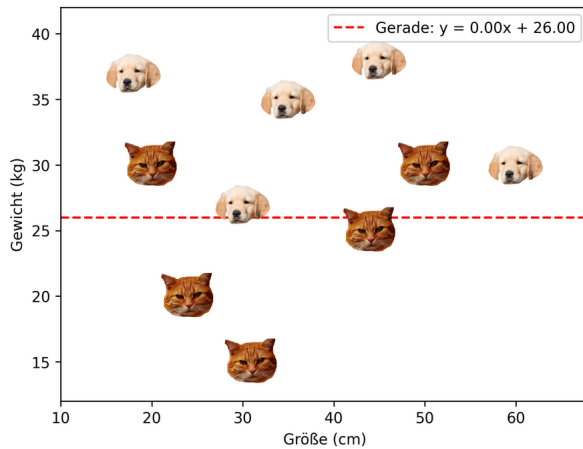


Abb. 1: Hunde bevorzugen

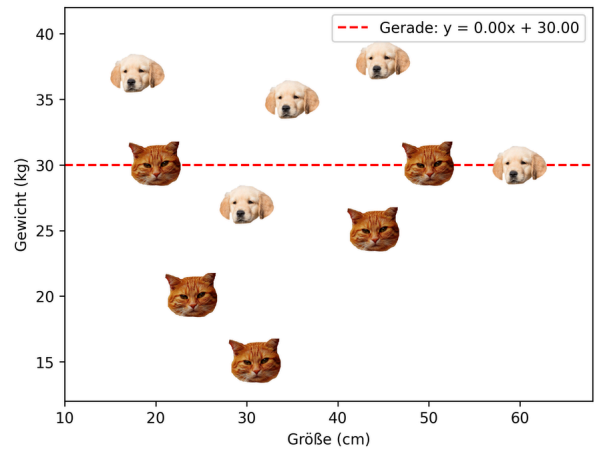
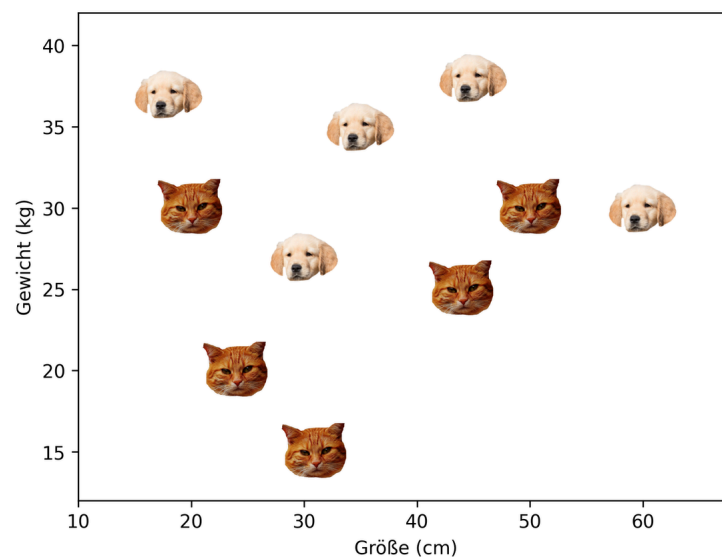


Abb. 2: Katzen bevorzugen

In Abbildung 1 kategorisieren wir alle Tiere über 26kg als Hunde. Dadurch werden alle Hunde richtig kategorisiert, zwei Katzen werden allerdings zu Hunden! In Abbildung 2 setzen wir die Entscheidungsgrenze auf 30kg. Wir kategorisieren jetzt alle Katzen richtig, aber zwei Hunde falsch! Geht das auch besser?

① Zeichne eine Gerade, die nur einen Hund und eine Katze falsch kategorisiert.

- Ist das fairer? Begründe!



Bei Katzen und Hunden ist das ja nicht so wild, wenn mal ein Tier falsch kategorisiert wird. Wenn Algorithmen über Menschen entscheiden, kann das aber sehr schwere Konsequenzen für den Einzelnen haben!

Die GeldVorMensch-Krankenkasse hat in ihren Daten eine Korrelation zwischen Diabetes Typ 2-Erkrankten, ihrer Körpergröße und ihrem Körpergewicht entdeckt. In Abbildung 3 werden die Daten visualisiert. Schwarz-weiß eingefärbte Personen sind dabei an Diabetes Typ 2 erkrankt.

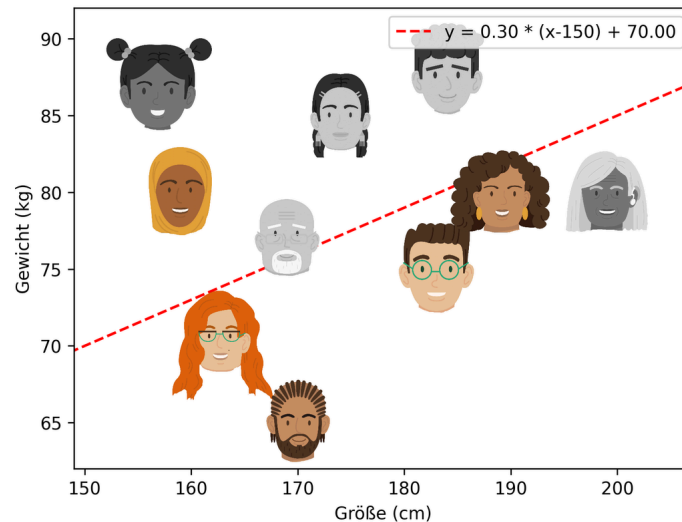


Abb. 3: Kategorisierung von kranken und gesunden Menschen

Alle Menschen mit einem Gewicht größer als  $y = 0,3 * (x-150) + 70$  (wobei  $x$  die Körpergröße in cm ist) sollen einen höheren Beitrag bezahlen, weil sie ein größeres Diabetes Typ 2-Risiko aufweisen. Die MenschVorGeld-Krankenkasse hat dieselben Datensätze ausgewertet, sie bietet gefährdeten Personen kostenfreie Vorsorgeuntersuchung an und verlangt keine Zusatzbeiträge.

- ② Bewerte, wann es sinnvoll sein könnte, dass Maschinen Entscheidungen über Menschen treffen und wann nicht!

---

---

---

---

---

---

---

---

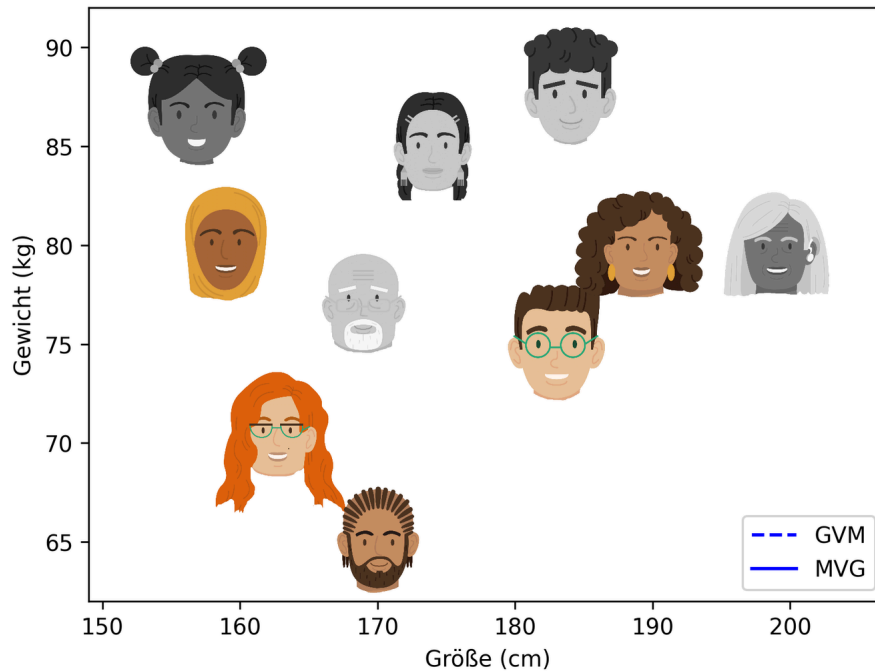
---

---



- ④ Wie könnten die Krankenkassen eine fairere Gerade finden?  
Zur Erinnerung: die GeldVorMensch-Krankenkasse will, dass Diabetes-gefährdete Personen einen höheren Beitrag bezahlen. Die MenschVorGeld-Krankenkasse möchte gefährdeten Personen eine kostenfreie Vorsorgeuntersuchung anbieten.

- Wie könnten die Geraden aussehen? Zeichne ein!
- Beschreibe jeweils eine mögliche Kostenfunktion in eigenen Worten.



GeldVorMensch-  
Krankenkasse:

---

---

---

---

MenschVorGeld-  
Krankenkasse:

---

---

---

---

Egal ob Mensch oder Tier, mit einer Geraden ist es schwierig alle Individuen richtig einzuordnen.

### Wie wäre es mit zwei Geraden?

Damit wir zwei Geraden so aneinanderschweißen können wie in der Abbildung, können wir die Geraden nicht einfach addieren. Wir müssen beide Geradengleichungen etwas anpassen. Wir stecken die Geradengleichungen jeweils mit dem Wert 0 in die  $\max()$ -Funktion. Das führt dazu, dass unsere Geraden keine Werte kleiner null annehmen können! Erst dann können wir sie addieren. Die gesamte Gleichung ergibt sich zu:

$$y = \max(0, w_1 \cdot x + b_1) + \max(0, w_2 \cdot x + b_2)$$



**max(x,y)**

Gibt den größeren Wert der beiden Eingangsparametern x und y zurück

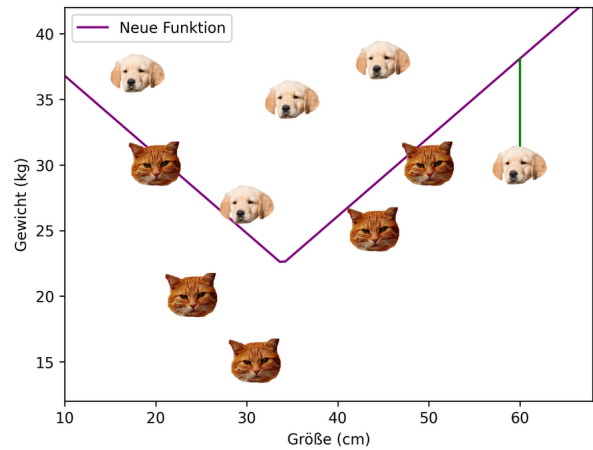
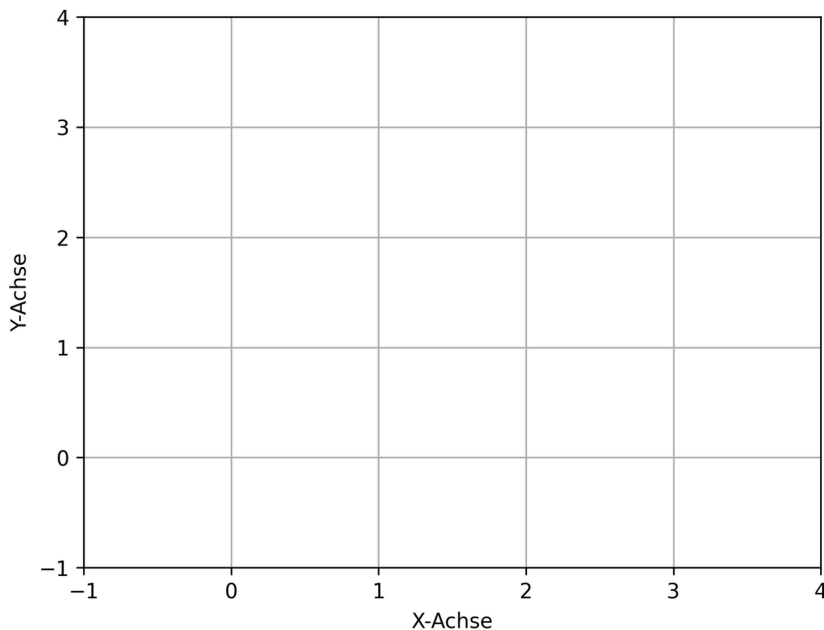


Abb. 5: Neue Funktion

⑤ Zeichne die Funktion  $y_1 = \max(0, -2 \cdot x + 4)$  und  $y_2 = \max(0, 1 \cdot x - 2)$  und deren Summe  $y = y_1 + y_2$  ein.

- Wie kann ich eine beliebige Funktion bauen? Begründe!





Was passiert wenn wir die Anzahl der Neuronen in unserer Gleichung erhöhen? Abbildung 7 zeigt ein Neuronales Netz. Die Neuronen werden in sogenannten **Schichten** angeordnet. Beachte, dass jetzt alle Neuronen einer Schicht mit allen Neuronen der nächsten Schicht über Gewichte miteinander verbunden sind! Wir haben jetzt 12 Parameter, die wir verändern können, um eine möglichst gute Gleichung zu finden! Wenn die Rede von einer KI oder einem KI-Modell ist, dann steckt meist so ein Neuronales Netz dahinter.

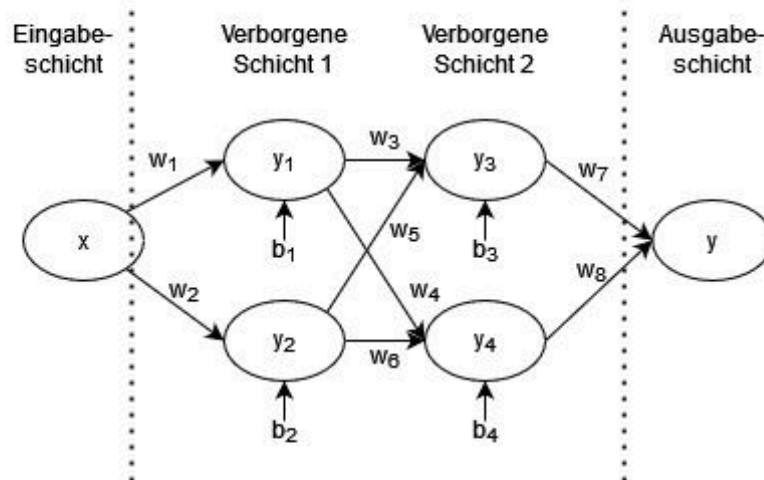


Abb. 7: Neuronales Netz mit 12 Parametern

Das geht auch noch besser! In der Eingabeschicht können wir mehr als nur eine Eingabe spezifizieren. Hier verwenden wir die Größe  $x$  und das Gewicht  $y$  der Tiere. Als Ausgabe soll das Modell uns einen Wert von 1 für einen Hund und einen Wert von 0 für eine Katze ausgeben. Wir können die Anzahl der verborgenen Schichten, sowie die Anzahl der Neuronen beliebig groß machen! Je mehr Neuronen, desto mehr Gewichte und Biase, die wir einstellen müssen. Das benötigt mehr Rechenzeit, kann aber auch zu einem besseren Modell führen.

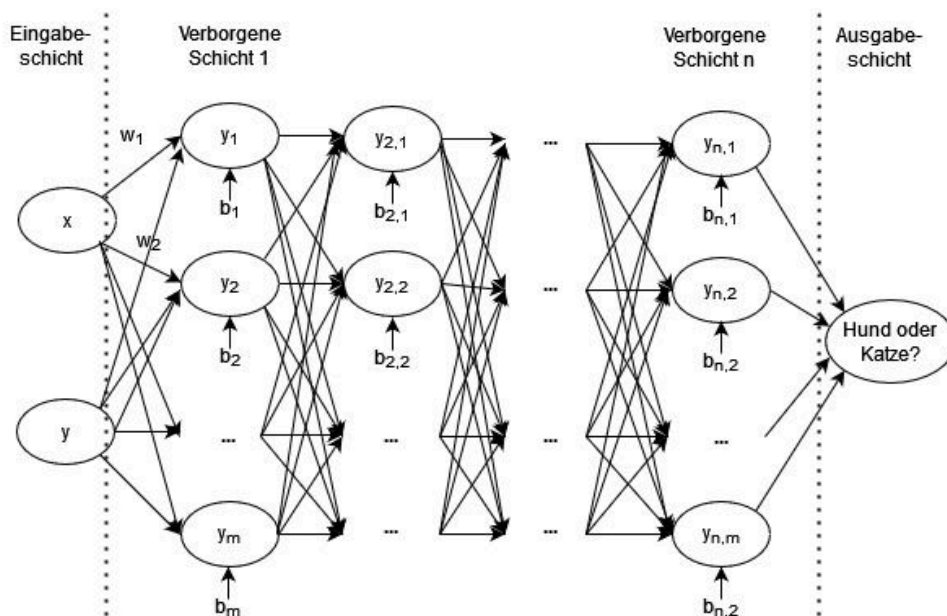


Abb. 8: Neuronales Netz mit  $n$  verborgenen Schichten mit jeweils  $m$  Neuronen

Die Gewichte und Biase der Neuronen werden so gewählt, dass die Kostenfunktion so klein wie möglich wird. Das Optimieren der Parameter wird als **Training** bezeichnet. Dem trainierten Modell können wir nun Gewicht und Größe des Tiers übergeben (wie bei einer Funktion). Als Rückgabe wird ein Wert berechnet, der zwischen 1 (Hund) und 0 (Katze) liegt.

Das Tierheim hat ihr Modell mit neuen Daten trainiert. Folgende Abbildung zeigt die Trainingsdaten und die Einschätzungen des KI-Modells.

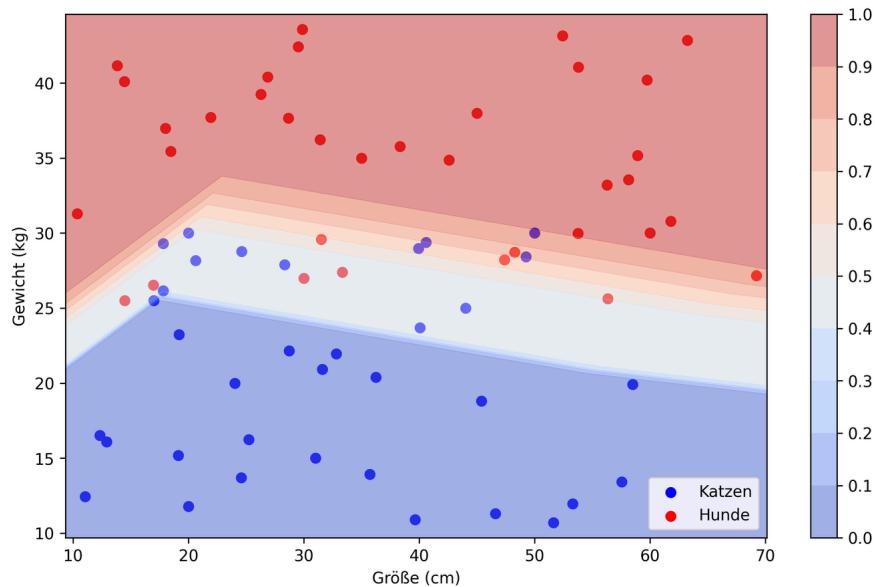
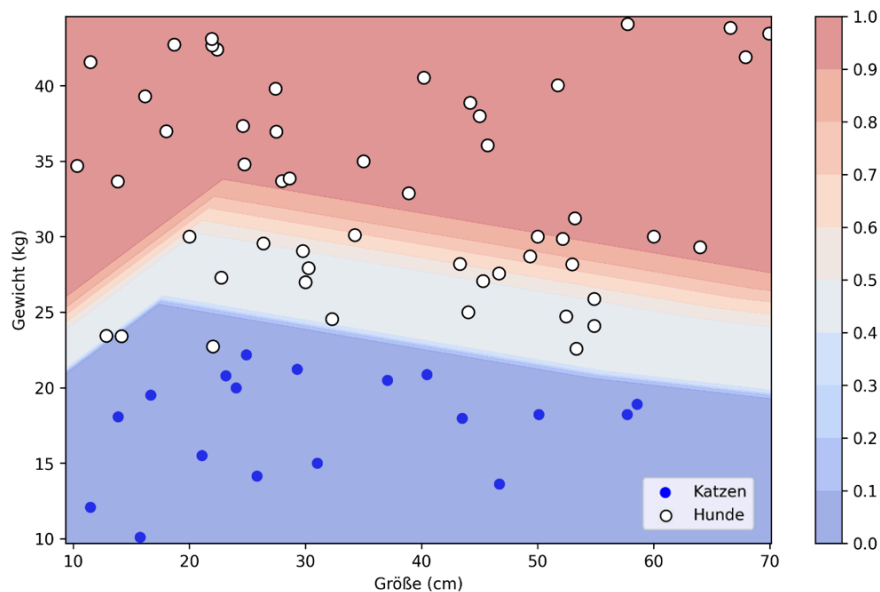


Abb. 9: Trainingsdaten und Einschätzungen der KI für bestimmte Größe und Gewicht eines Tiers

Die KI ist sich nicht immer sicher und gibt auch Werte zwischen 0 (Katze) und 1 (Hund) aus. Wir können die Werte aber nur in zwei Kategorien einordnen: Katzen oder Hunde. Dementsprechend werden alle Werte  $>0.5$  als Hund kategorisiert und  $<0.5$  als Katze.

- ⑧ Das Modell soll neue Tierdaten kategorisieren. Vervollständige die Daten, indem du die restlichen Tiere entsprechend kategorisierst. Male dazu alle als Katze kategorisierten Tiere aus.

- Vergleiche deine Ergebnisse mit Abb. 9. Was fällt dir auf?





## ⑨ Trainiere das Modell mit den selben Trainingsdaten in Jupyter Notebook.



- Beschreibe, was dir dabei auffällt.
- Begründe, wieso das so sein könnte.

---

---

---

---

---

In diesem Beispiel konnten wir uns alle Entscheidungen der KI visualisieren. Als Eingabegrößen gab es auch nur Gewicht und Größe eines Tiers. Wie eine KI in der echten Welt auf komplexe Situationen reagiert, kann man allerdings schlecht vorhersagen. Das müsste man eigentlich testen. Das ist allerdings gar nicht so leicht, denn unsere Eingaben haben meist deutlich mehr Parameter. Alle möglichen Kombinationen durchzutesten ist daher nicht praktikabel.

Die GeldVorMensch Krankenkasse hat ihr KI-Modell perfektioniert. Sie erkennt jetzt Menschen, die besonders gerne krank werden. Für diese Personengruppen erhöht sie den Beitrag. Außerdem verkauft sie ihr Modell an DieBöseJobAgenturGmbH. DieBöseJobAgenturGmbH nutzt das Krankheitsmodell, um jeder Person einen Berufsscore zuzuordnen. Menschen mit schlechtem Score bekommen keinen Job. Nur Menschen mit gutem Score werden an die zufriedenen Unternehmen weitervermittelt. Und DieBöseJobAgenturGmbH macht dabei ganz schön viel Geld!

## ⑩ Überlege, welche Unternehmen bereits genügend Daten von ihren Usern besitzen, um KI-Modelle zu trainieren.

- Hältst du die Gefahr des Second Use hier für groß? Bewerte!
- Wenn du Second Use von KI-Modellen bereits zum Opfer gefallen bist, wüsstest du das dann?



Werden trainierte KI-Modelle in einem anderen Kontext weiterverwendet, spricht man von **Second Use**.

---

---

---

---

---

Auch wenn du sensible Daten nicht angibst, könnte dein Profil ähnlich zu Personen sein, die sensible Daten angegeben haben! Die KI könnte beispielsweise lernen, welche sexuelle Orientierung du hast oder unter welcher Krankheit du leidest!

Ein Chihuahua wird im Tierheim abgegeben. Das Tierheim hinterfragt jetzt, ob Gewicht und Größe überhaupt entscheidend sind, um ein Tier richtig zu kategorisieren.

- ⑪ Überlege dir Kriterien, mit denen du Hunde und Katzen voneinander unterscheiden kannst.

- Welche Auswirkungen haben die verwendeten Kriterien auf das Modell? Begründe!

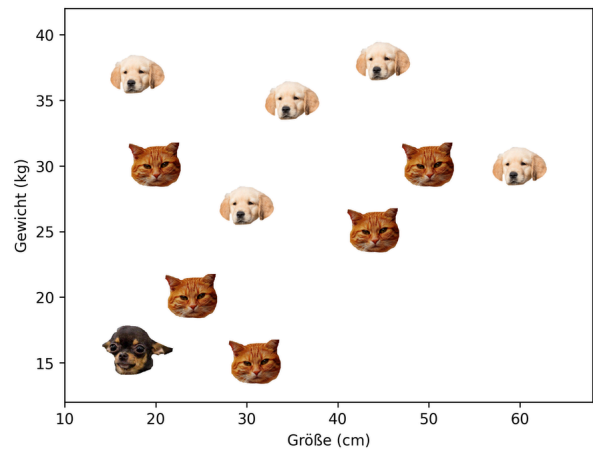


Abb. 10: Daten mit Chihuahua

- ⑫ Die Krankenkassen machen sich Gedanken. Ist das Gewicht und die Größe eines Menschen überhaupt ausreichend, um Krankheitswahrscheinlichkeiten vorherzusagen, auch wenn die Daten mit den Krankheiten korrelieren? Reflektiere!

Das Tierheim verwirft die Idee, die Tiere mittels Gewicht und Größe zu kategorisieren. Sie greift jetzt zur Bilderkennung mittels Neuronalen Netzen!



Jedes Pixel hat drei Farbwerte. Rot, grün und blau. Übergeben wir dem Modell ein Bild, so müssen wir also alle Pixel mit jeweils drei Farbwerten übergeben.



13 Berechne die Anzahl der Eingaben für die beiden Bilder. Vergleiche die Anzahl der Eingaben mit den bisherigen Eingaben des Tierheims.

- [illegible]



Seite 11/15




Es wurden vier unterschiedliche Modelle trainiert. Die Trainingsdaten wurden automatisch bei der Bildersuche mithilfe einer Suchmaschine heruntergeladen. Das Suchwort entspricht dem Label fürs Training. Beim „Hund oder Katze“-Modell wurden beispielsweise jeweils 150 Bilder mit dem Suchwort „hund“ und „katze“ heruntergeladen. Die beiden Suchwörter wurden dann als Label fürs Training verwendet.

⑭ Teste die Modelle mit den Testbildern in Jupyter Notebook.






- Markiere, ob die Bilder richtig kategorisiert wurden.
- Werfe einen Blick auf die Trainingsdaten und begründe, warum einige Bilder falsch kategorisiert wurden.






hund oder katze

Bilder		richtig (✓) oder falsch (x)?	Begründung
katze.jpg		✓	
hund.jpg			
chihuahua.jpg			
bär.jpg			



tomate oder apfel

Bilder		richtig (✓) oder falsch (x)?	Begründung
tomate.jpg			
apfel.jpg			
paprika.jpg			

arzt oder bauarbeiter

Bilder		richtig (✓) oder falsch (x)?	Begründung
arzt.jpg			
bauarbeiter.jpg			
arztMitHelm.jpg			
weißeTasse.jpg			
mannImAnzug.jpg			

cool oder uncool

Bilder		richtig (✓) oder falsch (x)?	Begründung
sonnenbrille.jpg			
vokuhila.jpg			
crocs.jpg			

⑮ Welche der trainierten Modelle eignen sich für ihre Problemstellung und warum?

---



---



---



---

Für das Trainieren der Modelle wurde ein bereits trainiertes Modell mit den entsprechenden Trainingsdaten feingestimmt (sogenanntes fine-tuning). Dabei werden lediglich die Neuronen der hintersten Schichten des Modells angepasst. Das spart Rechenzeit und wir benötigen nicht so viele Trainingsdaten.

Um ein KI-Modell von Grund auf zu trainieren, werden immens viele Daten benötigt. Das macht es besonders schwierig, alle Daten auf ihre Richtigkeit zu überprüfen! Des Weiteren ist es sehr schwierig, Bias in den Trainingsdaten aufzudecken! Das kann dazu führen, dass KIs diskriminieren.

- ⑩ Schau dir das Video an und begründe, wieso die Bildererkennung in den Beispielen so schlecht funktioniert!



[https://www.youtube.com/watch?v=QxuyfWoVV98&ab\\_channel=JoeyBuolamwini](https://www.youtube.com/watch?v=QxuyfWoVV98&ab_channel=JoeyBuolamwini)

---

---

---

---

---

---

---

---

- ⑪ Bias in den Trainingsdaten und Diskriminierung durch KI kann auch unsichtbar bleiben. Begründe, wieso es schwierig sein kann, Bias und Diskriminierung zu enttarnen.

- Was bedeutet das für dich? Begründe.

---

---

---

---

---

---

---

---

Ein Großkonzern trainiert ein eigenes Modell, um Bewerber zu klassifizieren. Die KI lernt aus den vergangenen Anstellungen des Unternehmens, welche Bewerber sich eignen und welche nicht. Aus den Daten geht hervor, dass das Unternehmen bisher Männer bevorzugt eingestellt hat.

- ⑮ Es gibt jeweils eine Bewerbung von einem Mann und einer Frau. Beide Personen sind gleich qualifiziert. Für wen entscheidet sich das trainierte KI-Modell? Bewerte!

---

---

---

---

Da die KI nun systematisch Männer bevorzugt, werden weniger Frauen eingestellt. Die KI trainiert fortlaufend mit den neuen Daten.

- ⑯ Wie stellt die KI zukünftige Bewerber ein? Begründe!

---

---

---

---

Amazon versuchte 2014 bis 2018 ein KI-Modell zur Sichtung von Bewerbungen zu trainieren. Das Modell lernte Bewerbungen von Frauen systematisch herabzustufen. Die KI hatte ein frauenfeindliches Vorurteil verinnerlicht. Das Projekt wurde schließlich abgebrochen, nachdem Amazon das Problem nicht beheben konnte.