

Willkommen beim Arbeitsmaterial zu kiVerstehen. Einige Aufgaben sind interaktiv. Diese Aufgaben werden durch das Jupyter-Symbol gekennzeichnet. Diese Aufgaben erledigst du, indem du das Jupyter-Notebook über ausführst. Zudem sind Zusatzaufgaben mit einem kleinen Sternchen gekennzeichnet.



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?

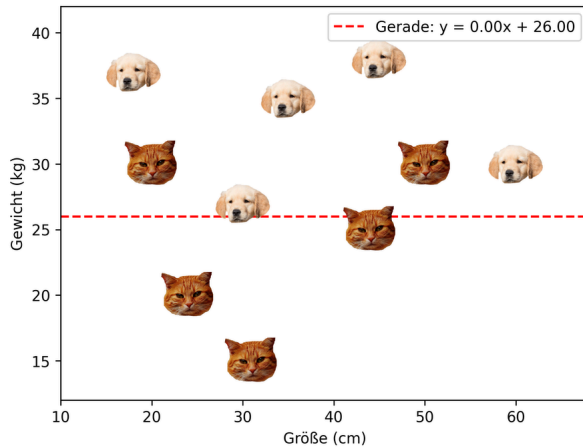


Abb. 1: Hunde bevorzugen

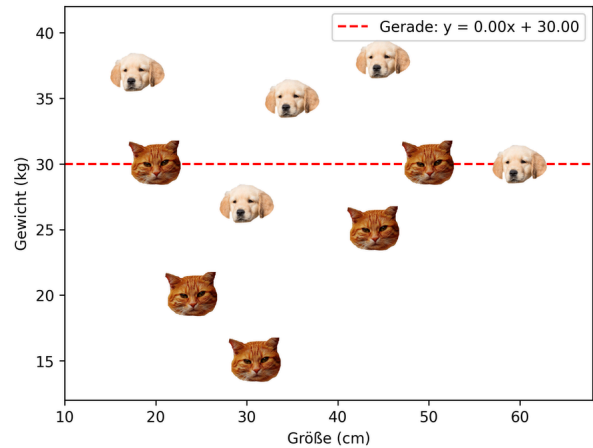
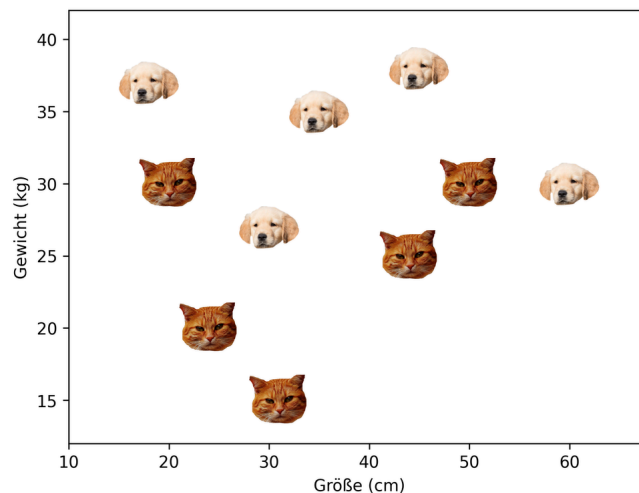


Abb. 2: Katzen bevorzugen

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 eine Katze und einen Hund falsch kategorisiert.
- Begründe, ob dies jetzt fairer ist.



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.

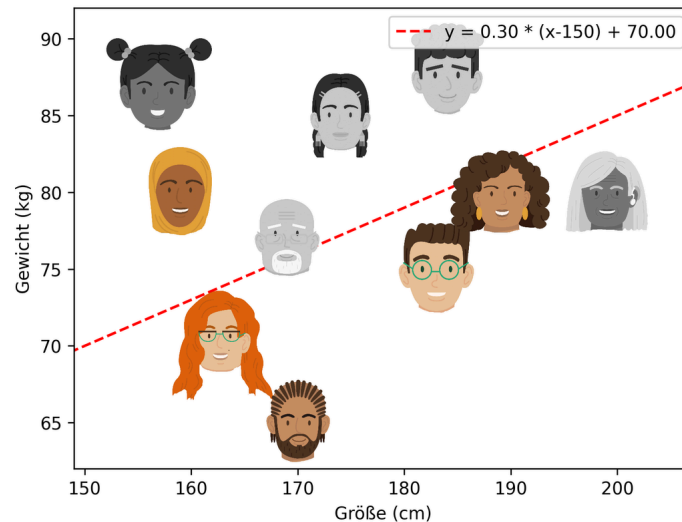


Abb. 3: Kategorisierung von kranken und gesunden Menschen

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.

Gibt es eine ideale Gerade? Das kommt darauf an, was wir mit ideal meinen.

Wir brauchen eine Möglichkeit zu beurteilen, wie ideal unsere Lösung ist. Eine erste Idee ist es, die falsch kategorisierten Tiere einfach zu zählen. Je weniger falsch kategorisierte Tiere gezählt werden, umso besser ist unsere Lösung. Die Funktion, die dies berechnet, wird **Verlustfunktion** genannt. Je kleiner der Verlust, desto besser.

- ③ Finde eine möglichst gute Lösung in Jupyter Notebook, indem du die Geradenparameter variierst und den entstehenden Verlust beurteilst. Je kleiner der Verlust, desto besser die Lösung.

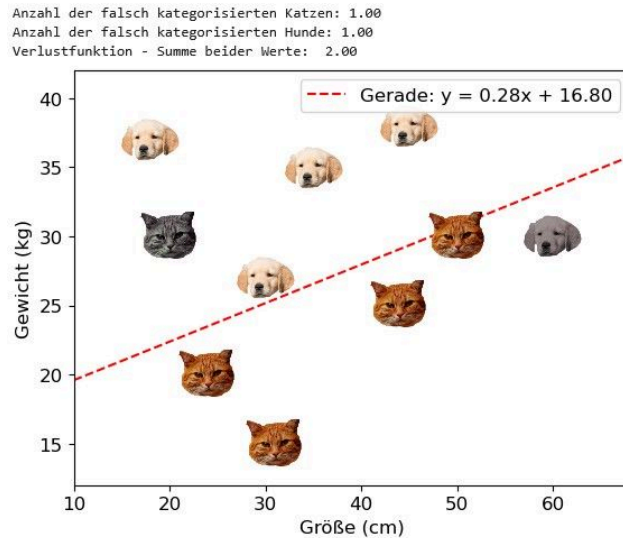


Abb. 4: Falsch kategorisierte Tiere zählen

Eine weitere Idee ist es, die Verlustfunktion so zu verändern, dass sie die Gewichtsunterschiede der falsch kategorisierten Tiere zur Geraden misst und aufaddiert.

- ④ Finde eine möglichst gute Lösung in Jupyter Notebook, indem du die Geradenparameter variierst und den entstehenden Verlust beurteilst.

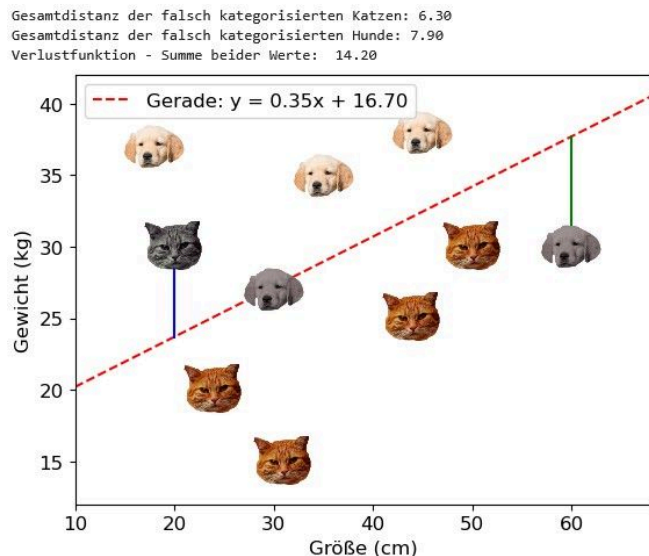


Abb. 5: Gewichtsunterschiede messen

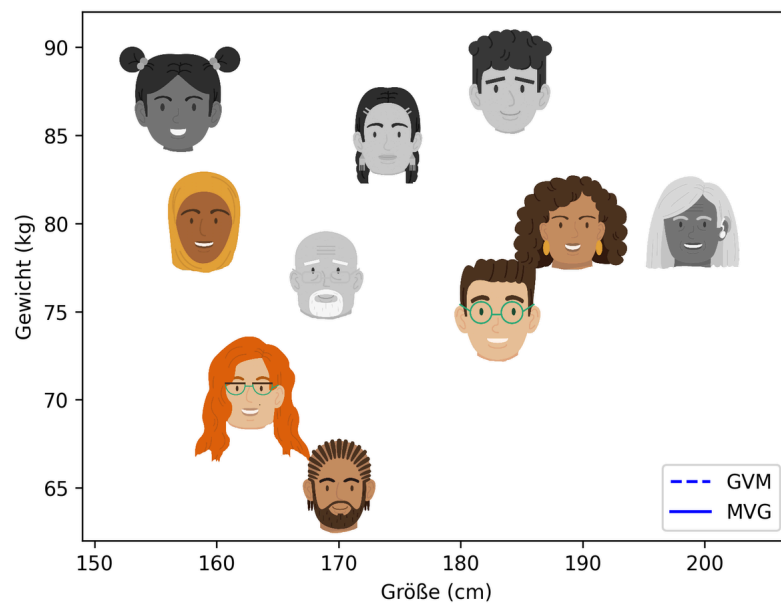
- ⑤ Bewerte, welche Verlustfunktion besser geeignet ist, um eine ideale Lösung zu finden.

Eine Verlustfunktion kann beliebig gewählt werden. Ist es uns wichtig, alle Katzen richtig zu kategorisieren, könnten wir beispielsweise die Gewichtsunterschiede der falsch kategorisierten Katzen in der Kostenfunktion stärker gewichten.

- ⑥ Können die Krankenkassen auch bessere Geraden finden, um die Menschen zu kategorisieren?
Zur Erinnerung: die GeldVorMensch-Krankenkasse (GVM) will, dass Diabetes-gefährdete Personen einen höheren Beitrag bezahlen. Die MenschVorGeld-Krankenkasse (MVG) möchte gefährdeten Personen eine kostenfreie Vorsorgeuntersuchung anbieten.

- Zeichne entsprechende Geraden ein.
- Bewerte die Fairness deiner Lösungen.

★ Beschreibe, wie die Verlustfunktionen aussehen könnten, um deine Lösungen abzubilden.



GeldVorMensch-
Krankenkasse:

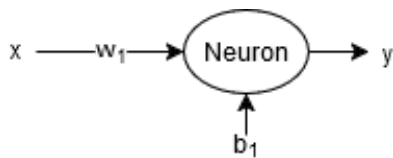
MenschVorGeld-
Krankenkasse:

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. Um Geraden miteinander zu verknüpfen, verwenden wir sogenannte **Neuronen**.

Schauen wir uns das Neuron genauer an:



Das Neuron funktioniert ähnlich wie eine Gerade. Eine Eingangsgröße x wird mit dem **Gewicht** w_1 multipliziert. Das ist so wie die Steigung einer Geraden. Der **Bias** b_1 bestimmt, wann das Neuron „feuert“, also ein Signal weitergibt. Den Bias kannst du auch mit dem y-Achsenabschnitt einer Geraden vergleichen. Solange der Wert der Geradengleichung $w_1 * x + b_1$ kleiner als 0 ist, wird das Ergebnis auf 0 gesetzt. Ansonsten wird das Ergebnis weitergegeben. Das Ganze kannst du in der nebenstehenden Abbildung beobachten. In blau ist die Übertragungsfunktion des Neurons dargestellt.

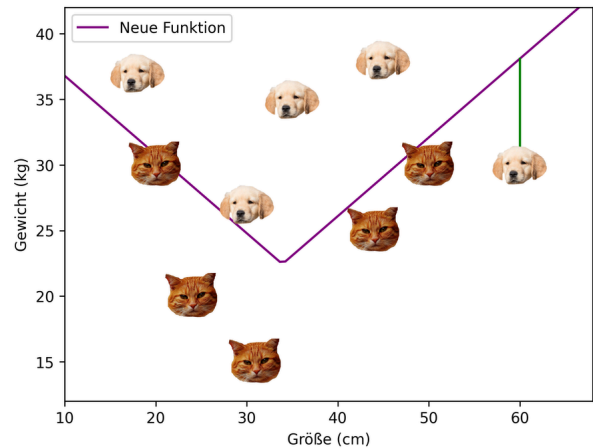


Abb. 6: Neue Funktion

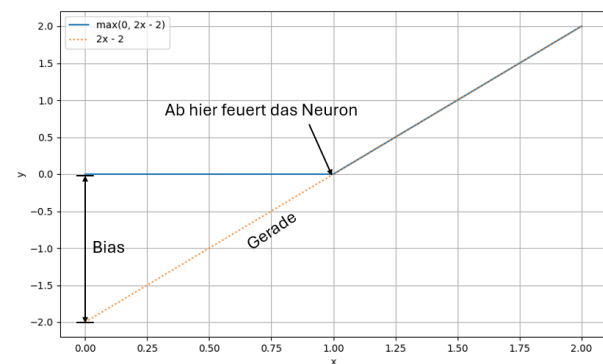


Abb. 7: Übertragungsverhalten eines Neurons mit einem Eingang x

Das funktioniert wie folgt: Wir definieren eine Funktion $\max(a,b)$, die den größeren Wert zweier Eingangswerte als Ergebnis zurückgibt. Rechts kannst du den Pseudocode sehen.

max(a,b)
 if $a > b$:
 return a
 else:
 return b

⑦ Berechne die Lösungen zu den folgenden Aufgaben.

$$\max(-2, 4) = 4$$

$$\max(5, 3) =$$

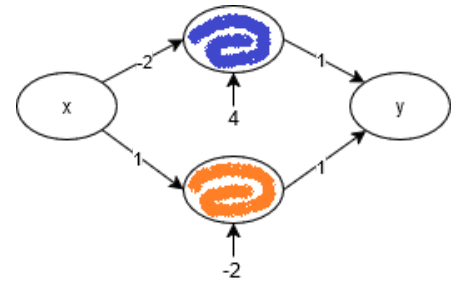
$$\max(0, 6x + 3) = \quad \text{für } x = 1$$

$$\max(0, 6x + 3) = \quad \text{für } x = -1$$

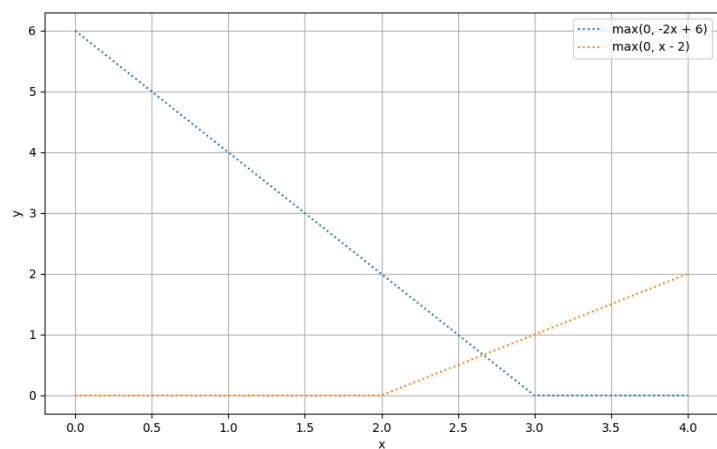
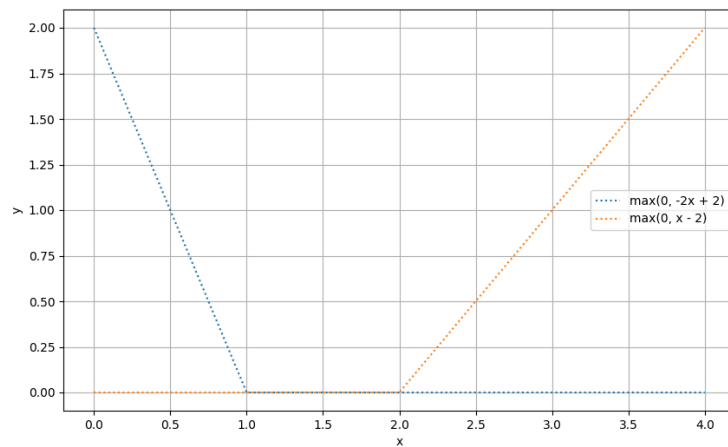
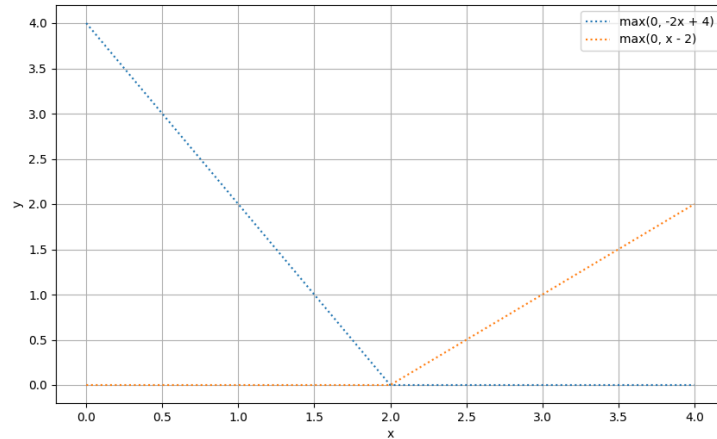
Das Übertragungsverhalten eines Neurons wird durch die Formel $\max(0, w_1 * x + b_1)$ modelliert. Wir übergeben der max-Funktion also die Werte 0 und die Geradengleichung mit der Steigung bzw. dem Gewicht w_1 und dem y-Achsenabschnitt bzw. dem Bias b_1 .

Durch das Addieren der unterschiedlichen Neuronenfunktionen können wir die Geraden verknüpfen. Werden Neuronenfunktionen in einem Netz miteinander summiert, spricht man von einem **Neuronalen Netz**.

Rechts ist ein Neuronales Netz mit zwei Neuronen dargestellt. Als Ausgabe wird $y = \max(0, -2x+4) + \max(0, x-2)$ berechnet. Beachte, dass auch die Ausgänge der Neuronen gewichtet werden können (hier mit 1).



⑧ Zeichne die Summe y der beiden Neuronenfunktionen (blau, orange) in die Graphen ein.

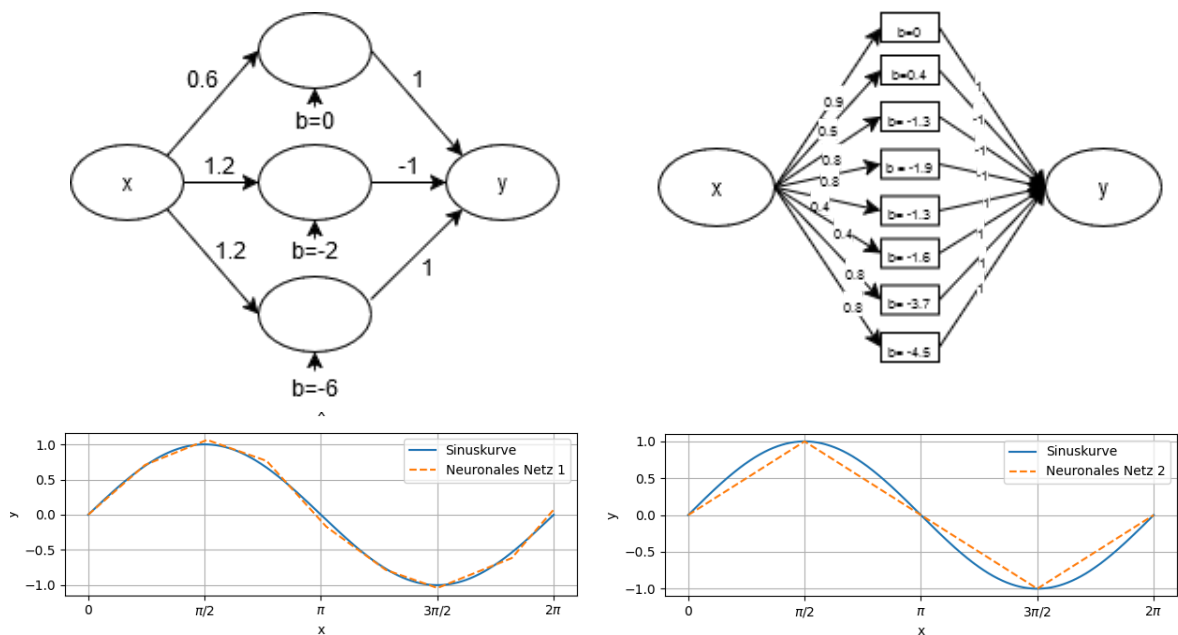


- ⑨ Finde eine möglichst ideale Funktion in Jupyter Notebook, um die Katzen und Hunde zu kategorisieren. Über den Verlust kannst du deine Lösung mit deinen Klassenkameraden vergleichen.
- Beschreibe, wie du bei der Lösungsfindung vorgegangen bist.

Du hast gerade die Parameter (Gewichte, Biase) eines kleinen Neuronalen Netzes optimiert. Das hast du wahrscheinlich gemacht, indem du die Parameter so verändert hast, dass die Verlustfunktion so klein wie möglich wird. Findet dieser Prozess automatisch statt, wird er als **Training** bezeichnet.

Theoretisch können wir jede beliebige Funktion mit genügend Neuronen nachbauen. Denn jede Funktion lässt sich durch das Aneinanderhängen kleiner Geradenstückchen annähern.

- ⑩ Für den Matheunterricht hat Tom eine Sinuskurve mit zwei Neuronalen Netzen angenähert. Dazu hat er einmal 3 und das andere mal 8 Neuronen verwendet. Ordne die Schaubilder der Neuronalen Netze den Graphen zu.



Ein Neuronales Netz besteht aus Neuronen, die in unterschiedlichen **Schichten** angeordnet sind. Bisher haben wir mit der Eingabeschicht, der Ausgabeschicht und einer verborgenen Schicht gearbeitet. Die Anzahl der verborgenen Schichten ist allerdings variabel. Alle Neuronen einer Schicht sind mit allen Neuronen der nächsten Schicht über Gewichte miteinander verbunden. Je mehr Schichten und Neuronen es gibt, desto mehr Parameter können verändert werden, um das Neuronale Netz auf die Problemstellung anzupassen. Mit einem Neuronalen Netz können wir jede beliebige Funktion nachbauen. Wieso erstellen wir also nicht eine Funktion die Gewicht und Größe der Tiere als Eingaben nimmt und direkt in Katzen und Hunde kategorisiert? 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 wählen. Je mehr Neuronen und Schichten, desto mehr Gewichte und Biase, die wir beim Training einstellen müssen. Das benötigt mehr Rechenzeit, kann aber auch zu einem besseren Modell führen.

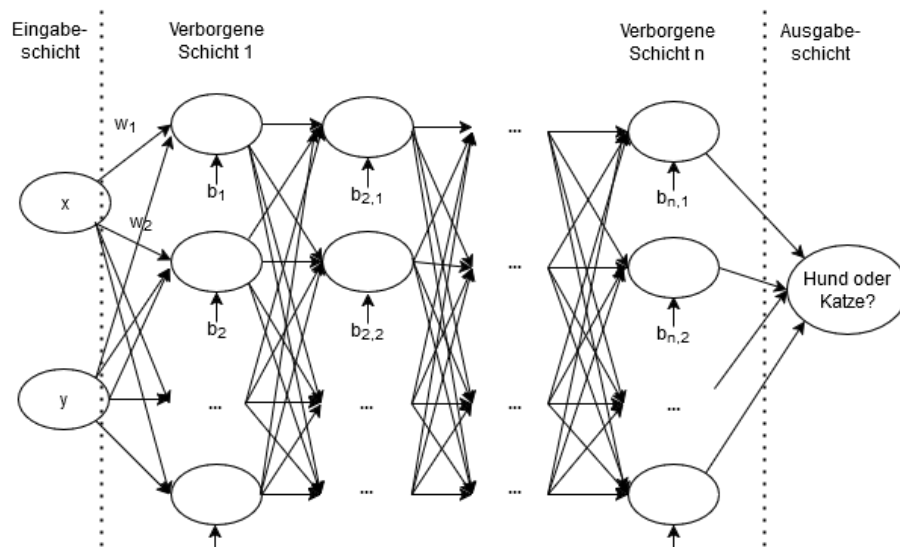


Abb. 8: Neuronales Netz mit n verborgenen Schichten mit jeweils m Neuronen

⑪ Das Tierheim hat ähnlich wie in Abb. 8 ein Neuronales Netz mit 19 verborgenen Schichten mit jeweils 10 Neuronen mit neuen Tierdaten trainiert.

- Teste das Modell mit den Testdaten.
- Kategorisiere die neu aufgenommenen Tiere mithilfe des Modells in Katzen und Hunde.
- Beschreibe, wie du den Ausgabewert des Modells zum Kategorisieren der Tiere interpretiert hast.



Testdaten

Größe: 20 cm

Gewicht: 25 kg

Größe: 40 cm

Gewicht: 40 kg

neue Tiere

?

Größe: 20 cm

Gewicht: 25 kg

Größe: 20 cm

Gewicht: 25 kg

?

Größe: 20 cm

Gewicht: 25 kg

?

Testet man das Modell auf alle Gewichts- und Größenkombinationen erhält man folgende Grafik. Die Punkte stellen die Trainingsdaten dar, die Farbwerte repräsentieren die Einschätzungen des KI-Modells.

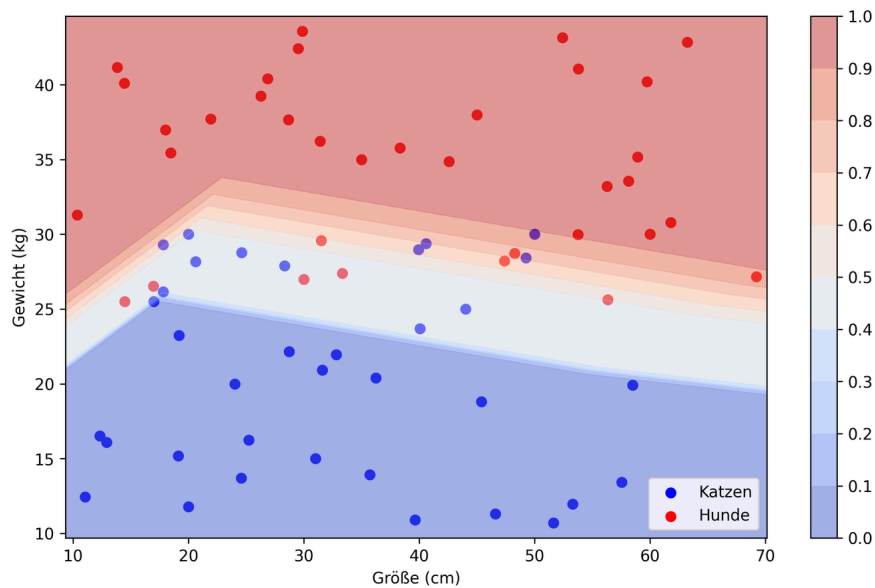
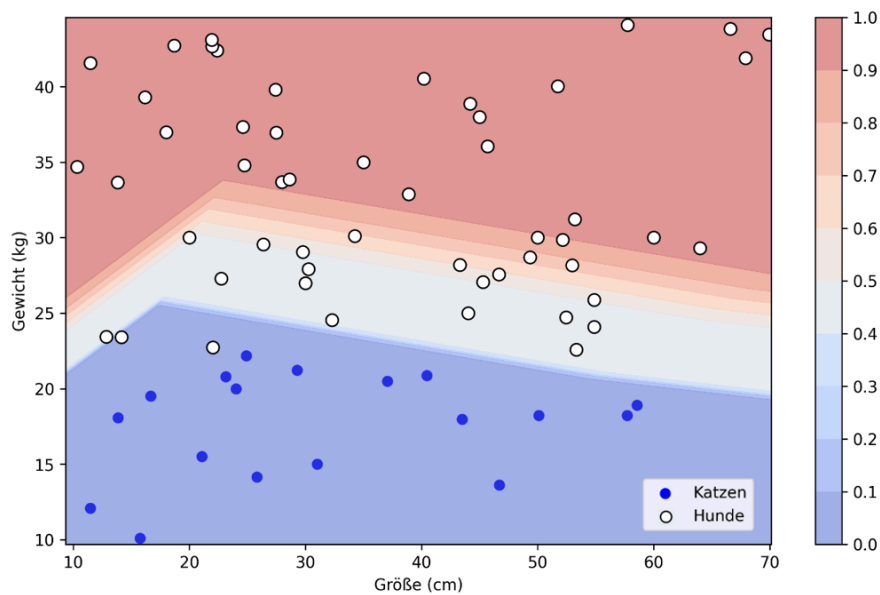


Abb. 9: Trainingsdaten und Einschätzungen der KI für bestimmte Größe und Gewicht eines Tiers

⑫ 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 und beschreibe, was dir dabei auffällt.



⑬ 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!

⑭ Zähle einige Unternehmen auf, die bereits genügend Daten von ihren Usern besitzen, um KI-Modelle zu trainieren.

- Bewerte die Gefahr, dass diese KI-Modelle in einem anderen Kontext verwendet werden.
- Bewerte, ob du selbst von Second Use betroffen sein könntest.



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.

- ⑮ Zähle Kriterien auf, mit denen du Hunde und Katzen voneinander unterscheiden kannst.
- Begründe, welche Auswirkungen die verwendeten Kriterien auf das Verhalten des Modells haben.

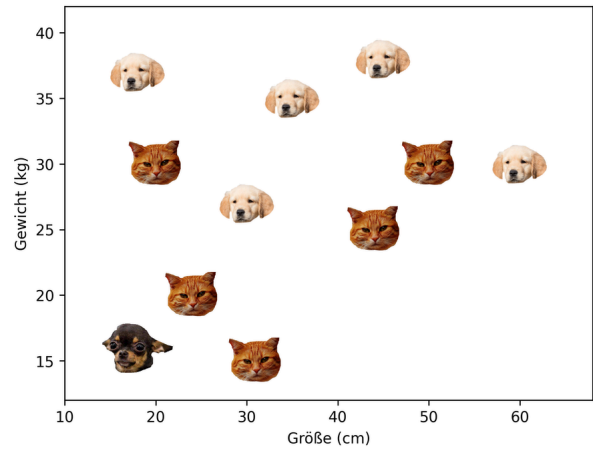


Abb. 10: Daten mit Chihuahua

- ⑯ 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? Begründe, zu welchen Erkenntnissen die Krankenkassen kommen sollten und warum.

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.



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

- 16x12 Pixel x3 Farbwerte
- 960x693 Pixel x3 Farbwerte

[illegible]

Das KI-Modell wird jetzt mit Bildern trainiert. Dazu brauchen wir viele Bilder mit **Label**. Ein Label ist die Beschreibung, was auf dem Bild zu sehen ist (bspw. Katze oder Hund). Erst durch das Label weiß die KI, ob sie ein Bild richtig erkannt hat. Wir trainieren das Modell nun so, dass möglichst viele Trainingsdaten richtig kategorisiert werden. Intern passiert das wie gehabt durch Minimierung einer Kostenfunktion.

Die KI kann nur die Zusammenhänge in den Bildern lernen, mit denen sie trainiert wird! Daher ist es ausgesprochen wichtig, sehr viele und sehr unterschiedliche Trainingsdaten (Bilder mit Label) zu verwenden.





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.

18) 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 (✗)?	Begründung
arzt.jpg			
bauarbeiter.jpg			
arztMitHelm.jpg			
weißeTasse.jpg			
mannImAnzug.jpg			

cool oder uncool

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

⑪ 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.

- 20 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

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

- Bewerte zudem, wie du persönlich damit umgehen kannst.

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.

- ②② Es gibt jeweils eine Bewerbung von einem Mann und einer Frau. Beide Personen sind gleich qualifiziert. Bewerte, für wen sich das trainierte KI-Modell entscheidet.

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

- ②③ Begründe, wie die KI zukünftige Bewerber einstellt.

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.