

Josh Starmer

Machine Learning → visuell lernen ←

von StatQuest

Mit Bildern ganz einfach
lernen und verstehen



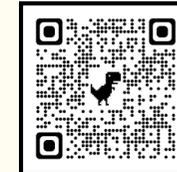
mitp



Über den Autor

Josh Starmer ist das Gesicht hinter dem bekannten Youtube-Kanal »StatQuest with Josh Starmer«. Seit 2016 erklärt er allen Interessierten mit seinem innovativen und einzigartigen visuellen Stil Konzepte aus den Bereichen Statistik, Data Science, Machine Learning und Algorithmen.

StatQuest hilft Menschen auf der ganzen Welt dabei, Wettbewerbe zu gewinnen, Prüfungen zu bestehen, neue Berufe zu ergreifen und befördert zu werden.



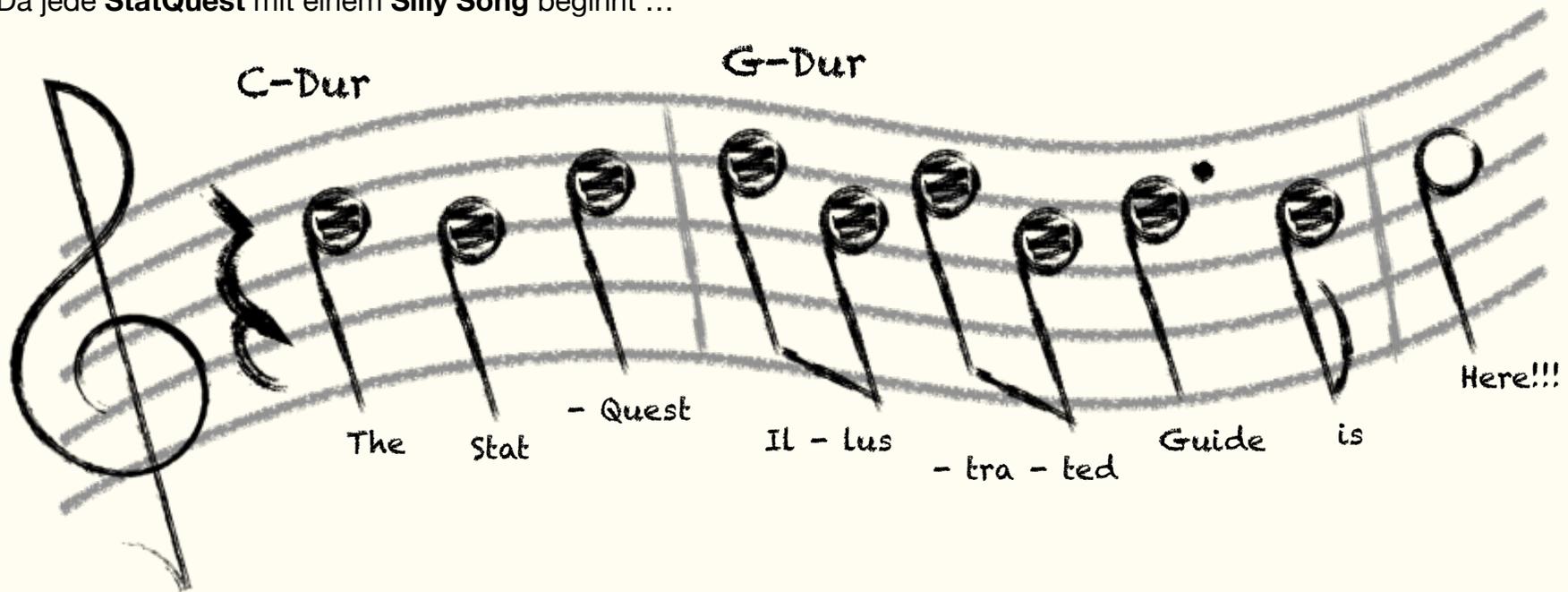
Scanne, klicke oder tippe auf diesen QR-Code, um [StatQuest.org](https://www.statquest.org) zu besuchen!!!

Über den Fachkorrektor der deutschen Ausgabe

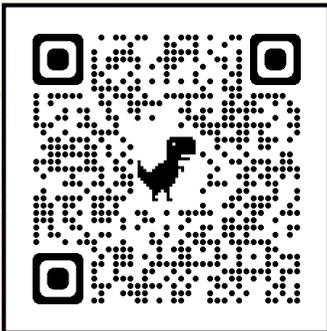
Friedhelm Schwenker ist Professor für Informatik (Fachgebiet: Machine Learning) an der Universität Ulm. Er hat im Bereich der Angewandten Mathematik promoviert und ist seit vielen Jahren im Bereich Machine Learning in Forschung und Lehre tätig. Seine Forschungsgebiete sind Pattern Recognition, Data Mining und Machine Learning mit Schwerpunkt Neuronale Netze. In jüngster Zeit befasst er sich auch mit Anwendungen des Machine Learning im Affective Computing. Er ist Editor von 19 Proceedingsbänden und Special Issues sowie Autor von 200+ Journal- und Konferenzartikeln.

Da jede **StatQuest** mit einem **Silly Song** beginnt ...

C-Dur



Scanne, klicke oder tippe auf diesen QR-Code, um den **Silly Song** zu hören!!!

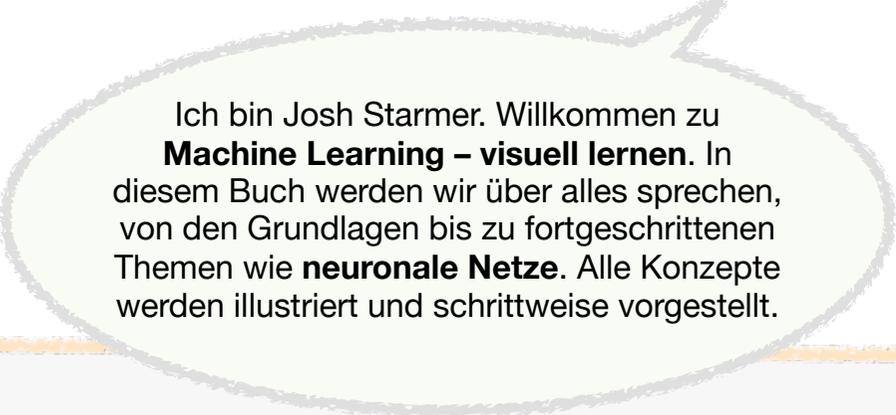


Hurra!!!

StatQuest!!!



Hallo!!!



Ich bin Josh Starmer. Willkommen zu **Machine Learning – visuell lernen**. In diesem Buch werden wir über alles sprechen, von den Grundlagen bis zu fortgeschrittenen Themen wie **neuronale Netze**. Alle Konzepte werden illustriert und schrittweise vorgestellt.

Inhaltsverzeichnis

01 Grundlegende Konzepte des Machine Learnings!!!	8
02 Kreuzvalidierung!!!	21
03 Grundlegende Konzepte der Statistik!!!	30
04 Lineare Regression!!!	75
05 Gradientenabstiegsverfahren!!!	83
06 Logistische Regression!!!	108
07 Naive Bayes!!!	120
08 Bewertung der Modellleistung!!!	136
09 Überanpassung verhindern durch Regularisierung!!!	164
10 Entscheidungsbäume!!!	183
11 Support-Vektor-Klassifikatoren und -Maschinen (SVMs)!!!	218
12 Neuronale Netze!!!	234
Anhänge!!!	271

Wie man mit dem Buch arbeitet

1

HINWEIS: Bevor wir beginnen, wollen wir uns anhand einer Beispielseite ein wenig über die Funktionsweise dieses Buches informieren.

2

Jede Seite beginnt mit einer Überschrift, die genau angibt, um welches Konzept es geht.

3

Auf jeder Seite siehst du eingekreiste Zahlen wie diese, ...

... man muss ihnen nur der Reihe nach folgen und jedes Konzept wird genau erklärt.

4

BÄM!! Jetzt, wo du weißt, wie dieses Buch funktioniert, können wir loslegen!!!

Machine Learning: Die Grundideen

1

Hey **Normalsaurus**, kannst du das gesamte Machine Learning in einem einzigen Satz zusammenfassen?



Klar, **StatSquatch!** **Machine Learning (ML)** ist eine Sammlung von Werkzeugen und Techniken, die Daten in (hoffentlich gute) Entscheidungen umwandeln, indem sie **Klassifizierungen** vornehmen, z. B. ob jemand einen Film mögen wird oder nicht, oder **quantitative Vorhersagen**, z. B. wie groß jemand ist.



2

Norm, willst du damit sagen, dass es beim Machine Learning nur um zwei Dinge geht?
1) Wir können es nutzen, um etwas zu **klassifizieren**, und 2) wir können es nutzen, um **quantitative Vorhersagen** zu treffen?

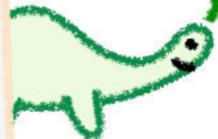


Das ist richtig, **Squatch!** Es geht nur um diese beiden Dinge. Wenn wir Machine Learning einsetzen, um Dinge zu **klassifizieren**, nennen wir das **Klassifizierung**. Und wenn wir **quantitative Vorhersagen** machen, nennen wir das **Regression**.



3

Lass uns also zunächst über die Grundideen sprechen, wie Machine Learning für die **Klassifizierung** eingesetzt wird.



BÄM!



Kapitel 01

Grundlegende

Konzepte des

Machine Learnings!!!

Machine Learning: Die Grundideen

1

Hey **Normalsaurus**, kannst du das gesamte Machine Learning in einem einzigen Satz zusammenfassen?

Klar, **StatSquatch!** **Machine Learning (ML)** ist eine Sammlung von Werkzeugen und Techniken, die Daten in (hoffentlich gute) Entscheidungen umwandeln, indem sie *Klassifizierungen* vornehmen, z. B. ob jemand einen Film mögen wird oder nicht, oder *quantitative Vorhersagen*, z. B. wie groß jemand ist.

2

Norm, willst du damit sagen, dass es beim Machine Learning nur um zwei Dinge geht?

1) Wir können es nutzen, um etwas zu *klassifizieren*, und 2) wir können es nutzen, um *quantitative Vorhersagen* zu treffen?

Das ist richtig, **Squatch!** Es geht nur um diese beiden Dinge. Wenn wir Machine Learning einsetzen, um Dinge zu *klassifizieren*, nennen wir das **Klassifizierung**. Und wenn wir *quantitative Vorhersagen* machen, nennen wir das **Regression**.

3

Lass uns also zunächst über die Grundideen sprechen, wie Machine Learning für die **Klassifizierung** eingesetzt wird.

BÄM!

Klassifizierung durch Machine Learning: Grundideen

1

Das Problem: Wir haben einen großen Haufen Daten, die wir für *Klassifizierungen* verwenden wollen.

Wir treffen zum Beispiel diese Person und wollen sie als jemanden **klassifizieren**, der **StatQuest** mögen wird oder nicht.



2

Eine Lösung: Wir können unsere Daten verwenden, um einen **Entscheidungsbaum** zu erstellen (für Details siehe **Kapitel 10**), um eine Person als jemanden zu **klassifizieren**, der **StatQuest** mögen wird oder nicht.

a

Sobald der **Entscheidungsbaum** erstellt ist, kann man ihn benutzen, um **Klassifizierungen** vorzunehmen, indem man an der Spitze beginnt und die Frage stellt: »Bist du an Machine Learning interessiert?«

b

Wenn du *nicht* an Machine Learning interessiert bist, geh bitte nach **rechts**, ...

g

BÄM!!!

Lernen wir nun die wichtigsten Ideen kennen, wie Machine Learning für die **Regression** verwendet wird.

Bist du an Machine Learning interessiert?

Ja

Nein

Magst du Silly Songs?

Ja

Nein

c

... und jetzt fragen wir: »Magst du **Silly Songs**?«

f

Dann wirst du **StatQuest** mögen!!!

Und wenn du dich für Machine Learning interessierst, dann sagt der **Entscheidungsbaum** voraus, dass du **StatQuest** magst!!!

Dann wirst du **StatQuest** mögen!!!

:(

e

Wenn du andererseits **Silly Songs** magst, dann sagt der **Entscheidungsbaum** voraus, dass du **StatQuest** mögen wirst!!!

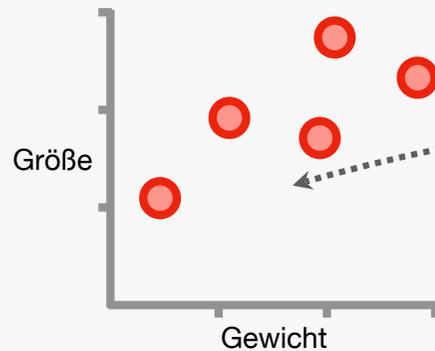
d

Wenn du dich nicht für Machine Learning interessierst und keine **Silly Songs** magst, dann ist das schade!

Machine Learning - Regression: Grundideen

1

Das Problem: Wir haben einen weiteren Haufen Daten und wollen diesen nutzen, um *quantitative Vorhersagen* zu treffen, d. h. wir wollen Machine Learning für die **Regression** einsetzen.



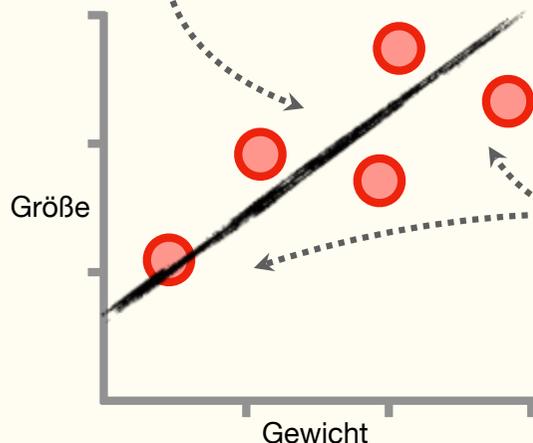
Hier haben wir zum Beispiel die **Körpergröße** und das **Gewicht** von 5 verschiedenen Personen gemessen. Da in den Daten ein Trend zu erkennen ist - je größer der Wert für das Gewicht, desto größer die Person - erscheint es sinnvoll, die Körpergröße anhand des Gewichts vorherzusagen.

Wenn also jemand neu zu uns kommt und uns sein Gewicht mitteilt, möchten wir diese Information nutzen, um seine Körpergröße vorherzusagen.



2

Eine Lösung: Mit der Methode der **linearen Regression** (für Details siehe **Kapitel 4**) können wir eine **Gerade** an die gesammelten Originaldaten anpassen und diese verwenden, um quantitative Vorhersagen zu treffen.

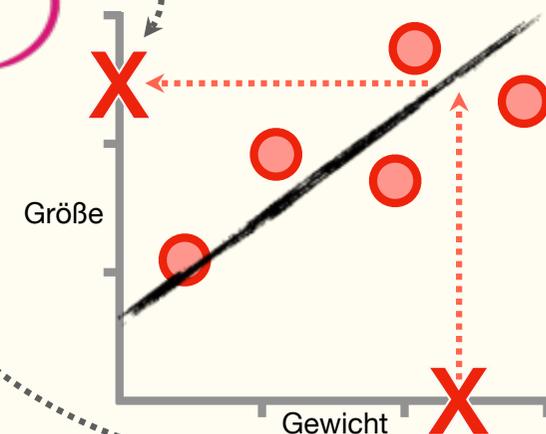


Die **Gerade**, die mit zunehmendem Gewicht ansteigt, fasst den Trend zusammen, den wir in den Daten gesehen haben: Wenn das Gewicht einer Person zunimmt, nimmt im Allgemeinen auch ihre Größe zu.

... dann könnten wir die **Gerade** benutzen, um vorherzusagen, dass deine Größe ist. **BÄM!!!**

Da es viele Methoden des Machine Learnings gibt, lass uns darüber sprechen, wie wir die beste Methode für unser Problem wählen.

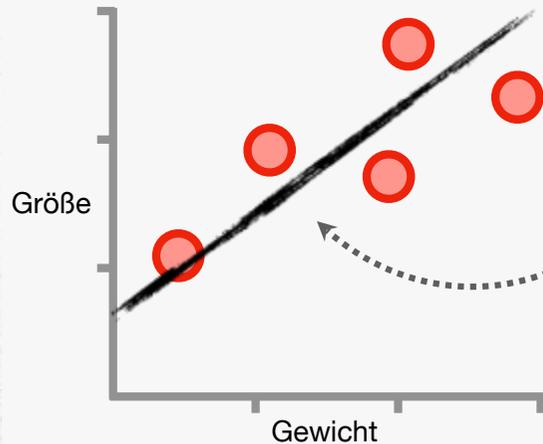
Wenn du mir sagen würdest, dass das dein Gewicht ist, ...



Methoden des Machine Learnings im Vergleich: Grundideen

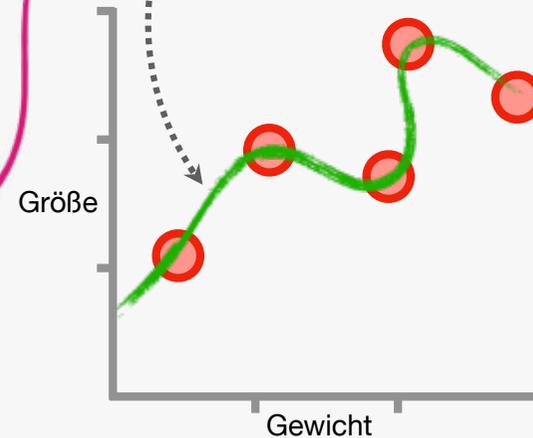
1

Das Problem: Wie du in diesem Buch lernen wirst, besteht Machine Learning aus vielen verschiedenen Methoden, die es ermöglichen, **Klassifizierungen** vorzunehmen oder **quantitative Vorhersagen** zu treffen. Wie entscheiden wir, welche wir verwenden?



Man könnte zum Beispiel diese **schwarze Gerade** verwenden, um die Größe aus dem Gewicht vorherzusagen, ...

... oder wir könnten diesen **grünen Schnörkel** benutzen, um die Größe aus dem Gewicht vorherzusagen.



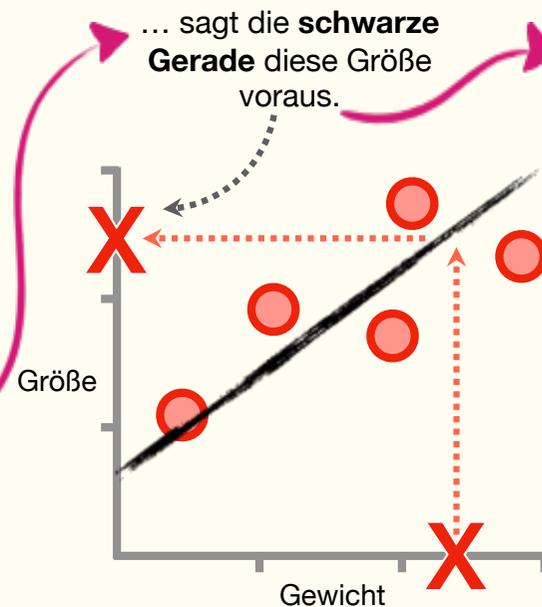
Wie entscheidet man, ob die **schwarze Gerade** oder der **grüne Schnörkel** verwendet werden soll?

2

Eine Lösung: Beim Machine Learning bedeutet die Entscheidung für eine Methode oft, sie einfach auszuprobieren und zu sehen, wie gut sie abschneidet.

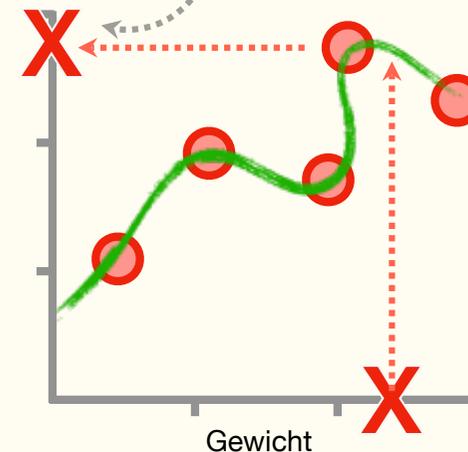


Zum Beispiel, wenn man das Gewicht dieser Person kennt, ...



... sagt die **schwarze Gerade** diese Größe voraus.

Im Gegensatz dazu sagt der **grüne Schnörkel** voraus, dass die Person *etwas größer* ist.



Wir können diese beiden Vorhersagen mit der *tatsächlichen* Größe der Person vergleichen, um die Güte der einzelnen Vorhersagen zu bestimmen.

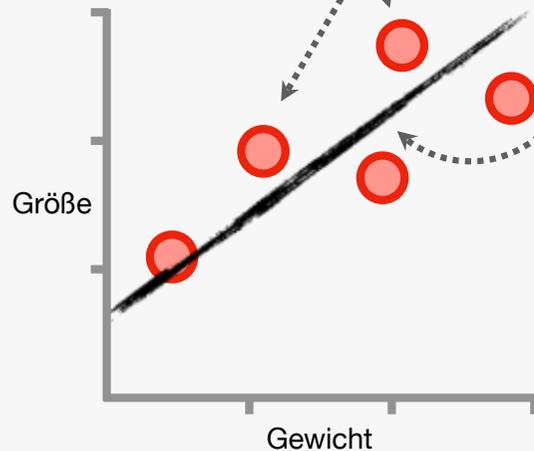
BÄM!!!

Nachdem wir nun die **Grundideen** hinter den Methoden des Machine Learnings verstanden haben, wollen wir uns ein Bild davon machen, wie man dieses in der Praxis anwendet.

Methoden des Machine Learnings im Vergleich: Einblick Teil 1

1

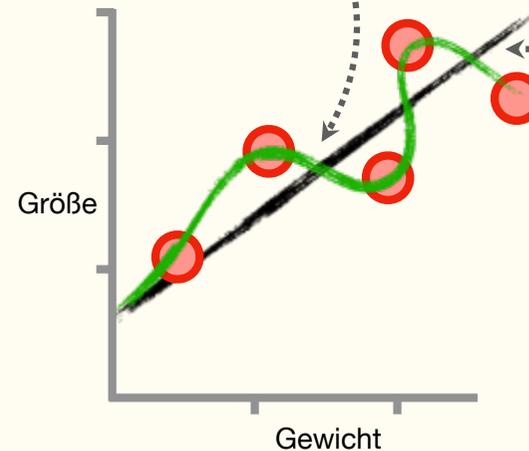
Die Daten, die wir zur Beobachtung des Trends und zur Anpassung der **Gerade** verwenden, werden **Trainingsdaten** genannt.



Mit anderen Worten: Die **schwarze Gerade** wird an die **Trainingsdaten angepasst**.

2

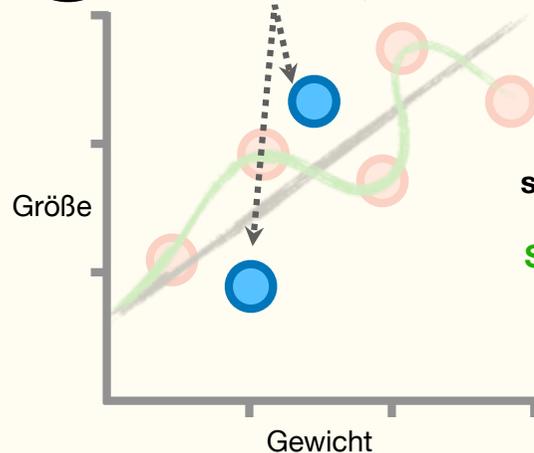
Alternativ hätte man auch einen **grünen Schnörkel** in die **Trainingsdaten** einfügen können.



Der **grüne Schnörkel** passt besser zu den **Trainingsdaten** als die **schwarze Gerade**, aber bedenke, dass das Ziel des Machine Learnings darin besteht, **Vorhersagen** zu treffen. Wir brauchen also eine Möglichkeit, um festzustellen, ob die **schwarze Gerade** oder der **grüne Schnörkel** bessere Vorhersagen macht.

3

Wir sammeln also weitere Daten, die sogenannten **Testdaten**, ...



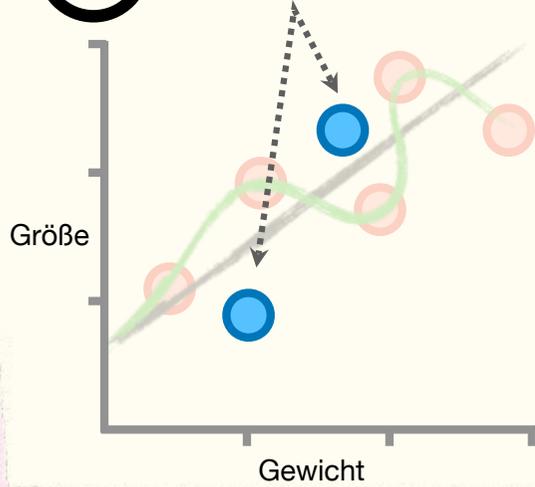
... und verwenden diese **Testdaten**, um die Vorhersagen der **schwarzen Gerade** mit den Vorhersagen des **grünen Schnörkels** zu vergleichen.

Hey **Normalsaurus**, würdest du dir nicht wünschen, dass wir eine Warnung bekommen, wenn neue Begriffe wie **Trainingsdaten** und **Testdaten** eingeführt werden?

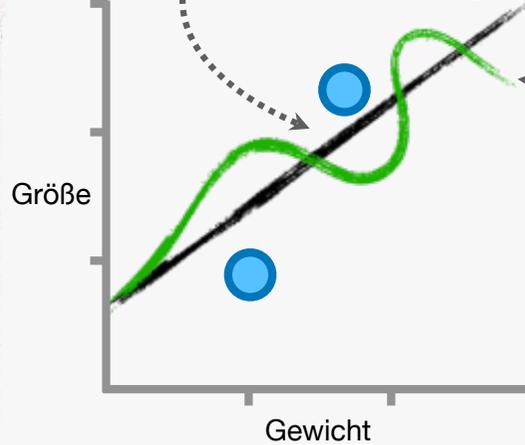
Das wäre toll, **StatSquatch!** Halte von nun an Ausschau nach dem gefürchteten **Terminologie-Alarm!!!**

Methoden des Machine Learnings im Vergleich: Einblick Teil 2

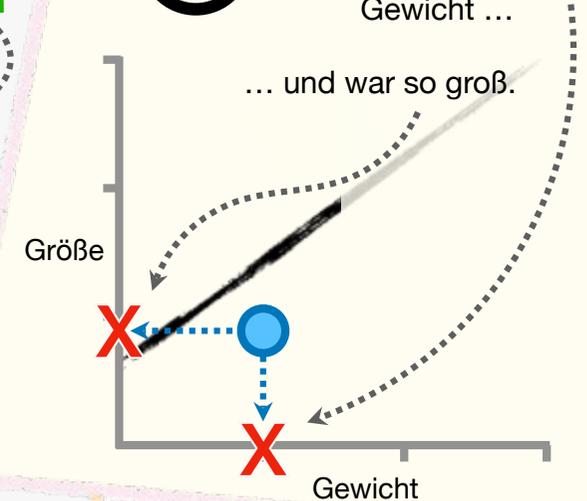
4 Wenn diese **blauen Punkte** nun die **Testdaten** sind, ...



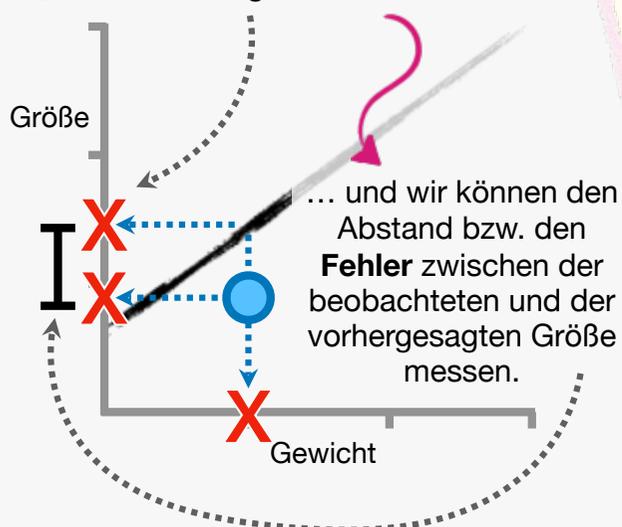
5 ... dann können wir die **beobachteten** Größen mit den Größen vergleichen, die durch die **schwarze Gerade** und den **grünen Schnörkel** vorhergesagt werden.



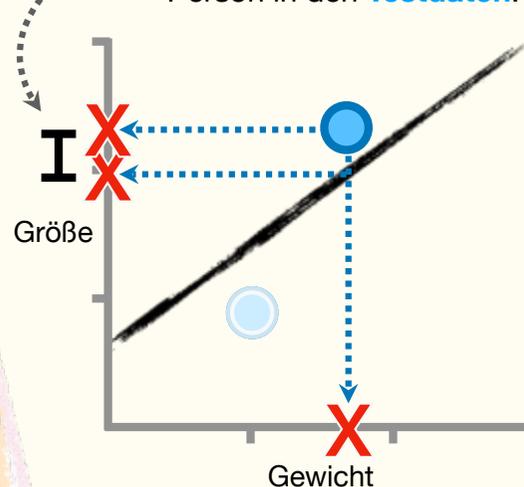
6 Die erste Person in den **Testdaten** hatte dieses Gewicht ...



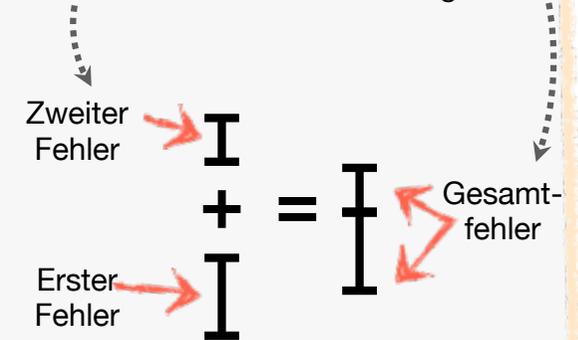
7 Die **schwarze Gerade** sagt jedoch voraus, dass sie größer ist, ...



8 Ebenso messen wir den **Fehler** zwischen den beobachteten und den vorhergesagten Werten für die zweite Person in den **Testdaten**.

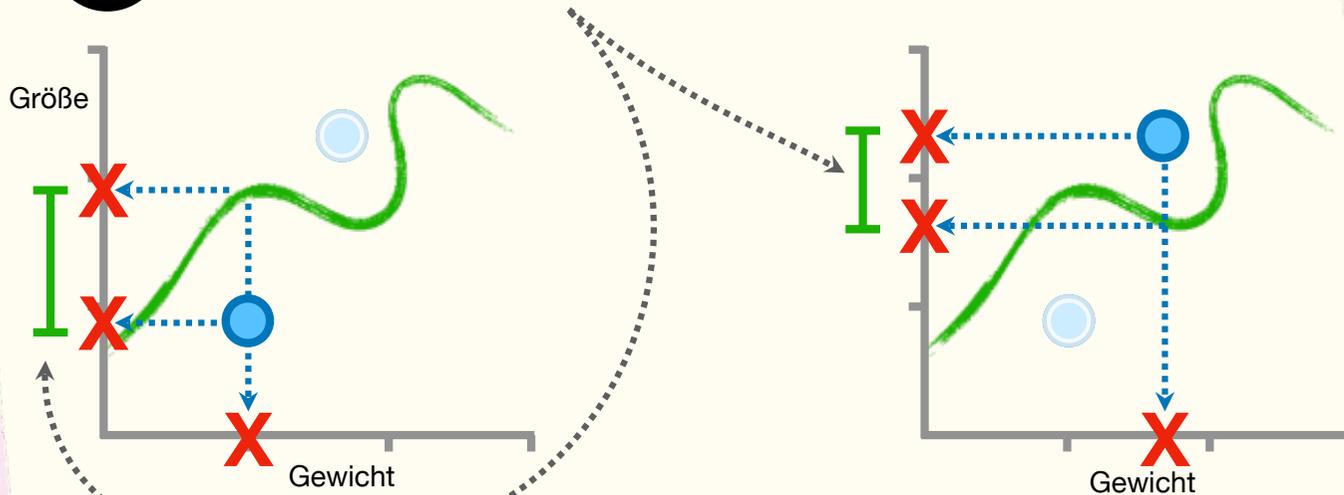


9 Wir können dann die beiden **Fehler** addieren, um ein Gefühl dafür zu bekommen, wie nahe die beiden Vorhersagen an den beobachteten Werten für die **schwarze Gerade** liegen.

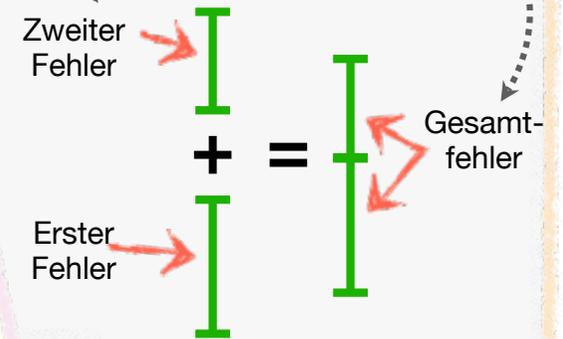


Methoden des Machine Learnings im Vergleich: Einblick Teil 3

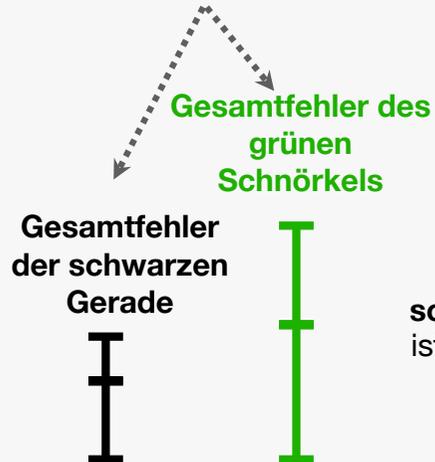
10 Ebenso kann man die Abstände bzw. **Fehler** zwischen den durch den **grünen Schnörkel** vorhergesagten Größen und den beobachteten Größen messen.



11 Wir können dann die beiden Fehler addieren, um ein Gefühl dafür zu bekommen, wie nahe die Vorhersagen an den beobachteten Werten für den **grünen Schnörkel** liegen.

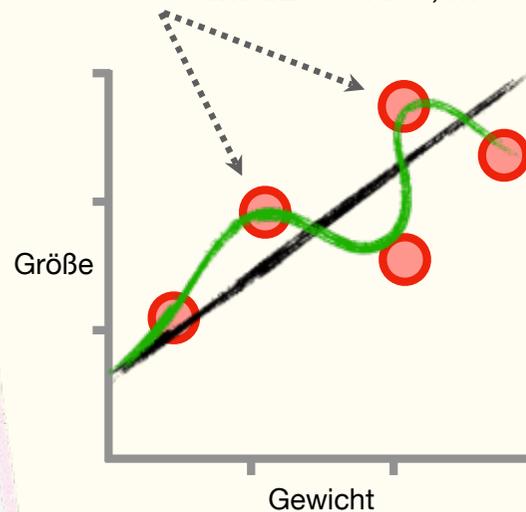


12 Nun kannst du die Vorhersagen der **schwarzen Gerade** mit den Vorhersagen des **grünen Schnörkels** vergleichen, indem du die Summen der **Fehler** vergleichst.

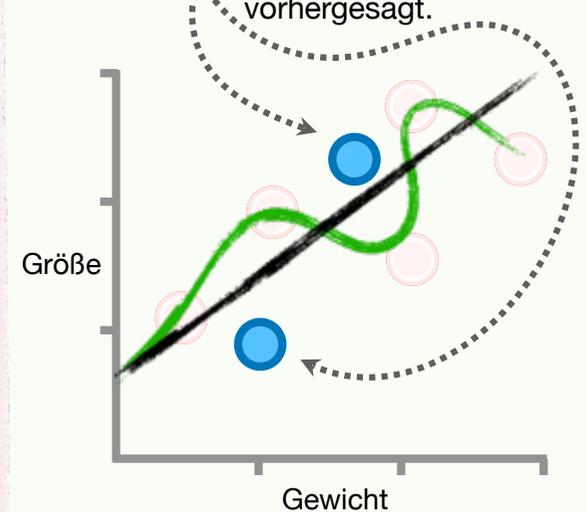


Wir sehen, dass die Summe der **Fehler** der **schwarzen Gerade** kürzer ist, was bedeutet, dass sie bessere Vorhersagen macht.

13 Mit anderen Worten: Auch wenn der **grüne Schnörkel** viel besser zu den **Trainingsdaten** passt als die **schwarze Gerade**, ...



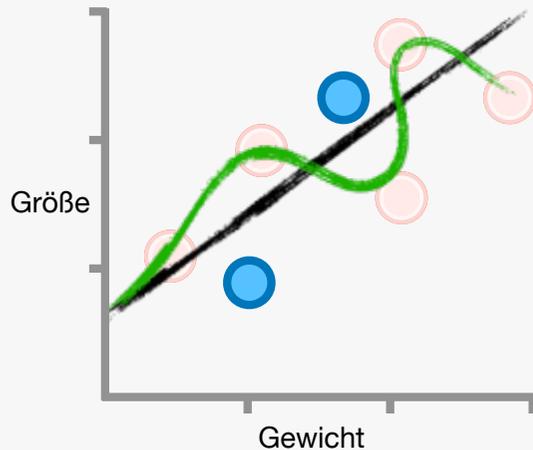
14 ... so hat die **schwarze Gerade** die **Größe** für die **Testdaten** besser vorhergesagt.



Methoden des Machine Learnings im Vergleich: Einblick Teil 4

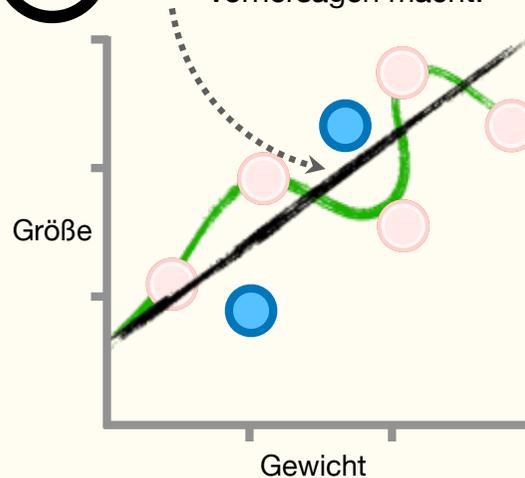
15

Wenn wir also zwischen der schwarzen Gerade und dem grünen Schnörkel wählen müssten, um Vorhersagen zu treffen, ...



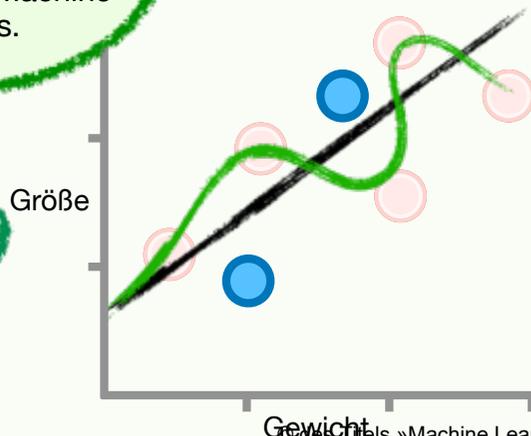
16

... würden wir die schwarze Gerade wählen, weil sie bessere Vorhersagen macht.



BÄM!!!

Das Beispiel, das wir gerade durchgespielt haben, zeigt 2 wichtige Konzepte des Machine Learnings.



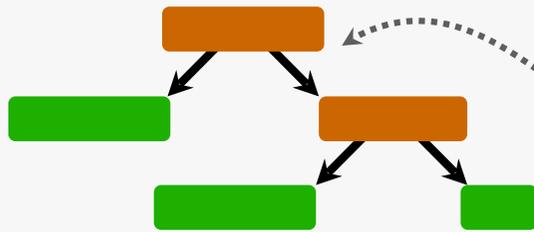
Erstens: Wir verwenden **Testdaten**, um Methoden des Machine Learnings zu bewerten.

Zweitens: Nur weil eine maschinelle Lernmethode gut zu den **Trainingsdaten** passt, bedeutet das nicht, dass sie auch bei den **Testdaten** gut funktioniert.

TERMINOLOGIE-ALARM!!!

Wenn eine Machine-Learning-Methode *sehr gut zu den Trainingsdaten* passt, aber *schlechte Vorhersagen* macht, spricht man von einer **Überanpassung** der Methode an die **Trainingsdaten**. Die **Überanpassung** einer Machine-Learning-Methode hängt mit dem **Bias-Varianz-Dilemma** zusammen, auf das wir später noch näher eingehen werden.

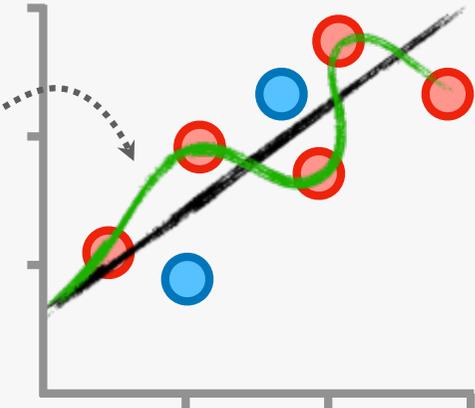
Die wichtigsten Konzepte des Machine Learnings: Zusammenfassung



Vielleicht fragst du dich jetzt, warum wir dieses Buch mit einem super einfachen **Entscheidungsbaum** begonnen haben, ...

... und einer einfachen **schwarzen Gerade** und einem albernem **grünen Schnörkel** anstelle eines ...

... **Deep Learning Convolutional Neural Networks** oder einer **[hier die neueste, ausgefallenste Methode des Machine Learnings einfügen]**.



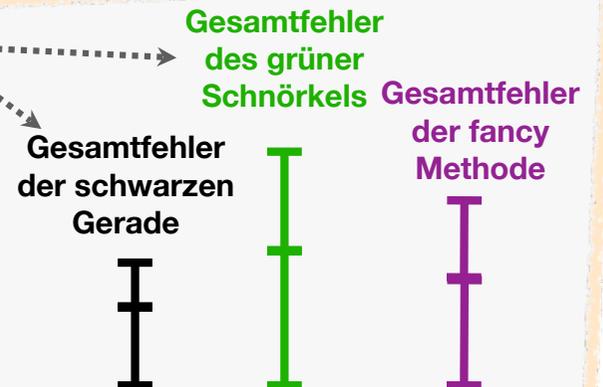
Es gibt viele coole Machine-Learning-Methoden. In diesem Buch lernst du etwas über ...

- Regression**
- Logistische Regression**
- Naive Bayes**
- Entscheidungsbäume**
- Regressionsbäume**
- Support-Vektor-Maschinen**
- Neuronale Netze**

Es gibt eine Vielzahl hochtrabender Methoden im Machine Learning wie **Deep Learning Convolutional Neural Networks** und jedes Jahr kommt etwas Neues und Aufregendes auf den Markt. Aber egal, was du benutzt, das Wichtigste ist, wie die Methode mit den **Testdaten** funktioniert.

BÄM!!!

Jetzt, wo du einige der wichtigsten Ideen des Machine Learnings verstanden hast, solltest du dir ein paar ausgefallene Begriffe aneignen, damit du dich schlau anhörst, wenn du auf Partys über **diese Dinge sprichst**.



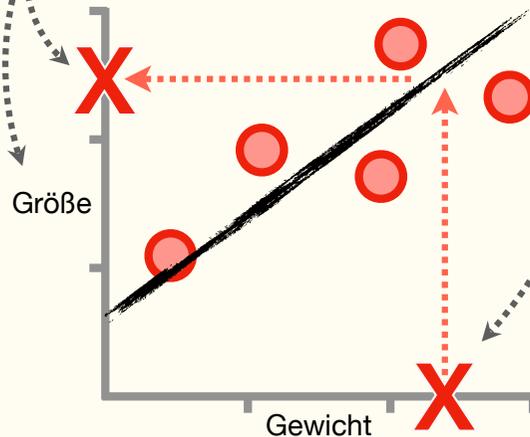
Terminologie-Alarm!!! Unabhängige und abhängige Variablen

1

Bis jetzt haben wir die Größe ...

... aus Gewichtsmessungen vorhergesagt, ...

... und die Daten wurden alle in einem schönen Diagramm dargestellt. Wir können die Daten auch in einer Tabelle organisieren.



Unabhängig davon, ob man die Daten im Diagramm oder in der Tabelle betrachtet, erkennt man, dass das Gewicht von Person zu Person *variiert*, und daher wird das Gewicht als **Variable** bezeichnet.

Ebenso *variiert* die Körpergröße von Person zu Person, sodass die Körpergröße auch als **Variable** bezeichnet wird.

Gewicht	Größe
0,4	1,1
1,2	1,9
1,9	1,7
2,0	2,8
2,8	2,3

2

Wir können auch die Arten von **Variablen**, die Größe und Gewicht darstellen, genauer beschreiben.

Da die Vorhersagen zur Körpergröße von den Gewichtsmessungen *abhängen*, nennen wir die Körpergröße eine **abhängige Variable**.

Da wir jedoch das Gewicht nicht vorhersagen und das Gewicht somit nicht von der Größe abhängt, bezeichnen wir das Gewicht als **unabhängige Variable**. Alternativ kann das Gewicht auch als **Merkmal** bezeichnet werden.

3

Bisher haben wir in unseren Beispielen nur das Gewicht, eine einzelne **unabhängige Variable** bzw. ein **Merkmal** zur Vorhersage der Größe verwendet. Es ist jedoch üblich, mehrere **unabhängige Variablen** bzw. **Merkmale** zu nutzen, um Vorhersagen zu treffen. Zum Beispiel könnten wir, um die Körpergröße vorherzusagen, Gewicht, Schuhgröße und Lieblingsfarbe verwenden.

Gewicht	Schuhgröße	Lieblingsfarbe	Größe
0,4	3	Blau	1,1
1,2	3,5	Grün	1,9
1,9	4	Grün	1,7
2,0	4	Rosa	2,8
2,8	4,5	Blau	2,3

Bäm.

Wie in der Tabelle zu sehen ist, ist das Gewicht eine *numerische Größe* und die Lieblingsfarbe eine *diskrete Kategorie*, es handelt sich also um unterschiedliche Datentypen. Lies weiter, um mehr über diese Typen zu lernen!!!

Terminologie-Alarm!!! Diskrete und kontinuierliche Daten

1 Diskrete Daten ...
... sind **abzählbar** und nehmen nur **bestimmte** Werte an.

2 Man kann z. B. die Anzahl der Menschen zählen, die die Farbe **Grün** oder die Farbe **Blau** mögen.



4 Menschen mögen **Grün**



3 Menschen mögen **Blau**

Da wir einzelne Personen zählen und die Gesamtzahlen nur ganze Zahlen sein können, sind die Daten **diskret**.

3 Amerikanische Schuhgrößen sind **diskret**, es gibt zwar halbe Schuhgrößen wie $8 \frac{1}{2}$, aber keine Größen wie $8 \frac{7}{36}$ oder $9 \frac{5}{18}$.

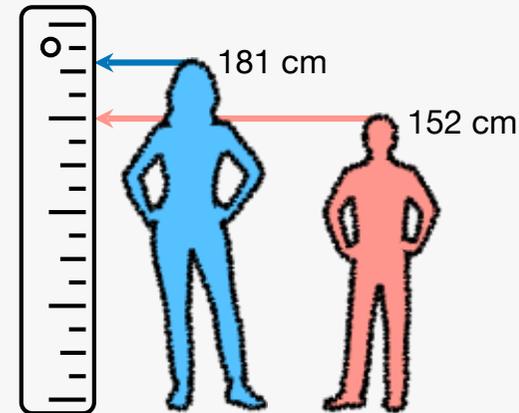


4 Ranglisten und andere Platzierungen sind ebenfalls **diskret**. Für den **1,68. Platz** gibt es keine Auszeichnung. Total blöd!



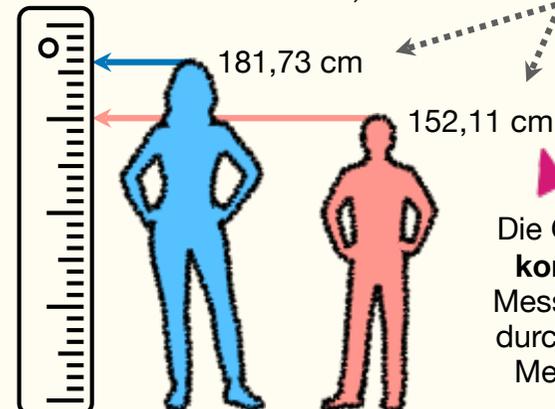
5 Kontinuierliche Daten ...
... sind **messbar** und können **jeden** numerischen Wert innerhalb eines Bereichs annehmen.

6 Zum Beispiel zählen Messungen der Körpergröße zu **kontinuierlichen** Daten.



Die Größe kann eine beliebige Zahl zwischen **0** und der Größe der größten Person auf der Welt sein.

7 **HINWEIS:** Wenn wir ein genaueres Lineal benutzen, ...
... dann werden die Messungen präziser.



Die Genauigkeit der **kontinuierlichen** Messungen wird nur durch die genutzten Messinstrumente begrenzt.



- Abhängige Variablen 18
- Aktivierungsfunktion 237, 239
- AUC 158-159
- Backpropagierung 252-268
- Bias 254
- Bias-Varianz-Dilemma 16, 221
- Binomialverteilung 38-44
- Blatt (Entscheidungsbaum) 185
- Datenleck 23
- Diskrete Daten 19
- Diskrete Wahrscheinlichkeitsverteilung 37
- Exponentialverteilung 54
- Falsch positiv 70
- Falsch-Positiv-Rate 146
- Gauß-(Normal-)Verteilung 49-51
- Gewichte 254
- Gini-Impurity 191-192
- Gleichverteilung 54
- Hidden Layer 238
- Histogramme 32-35
- Hypothesentest 71
- Impure (Unrein) 190
- Interner Knoten (Entscheidungsbaum) 185
- Konfusionsmatrix 138-142
- Knoten (Neuronale Netze) 237
- Kontinuierliche Daten 19
- Kontinuierliche Wahrscheinlichkeitsverteilung 48
- Kostenfunktion 88
- Lernrate 94
- Likelihood vs. Wahrscheinlichkeit 112-114
- Margin 224
- Merkmal 18
- Mittlerer quadratischer Fehler (MSE) 61-62
- Modelle 56-57
- Normal-(Gauß-)Verteilung 49-51
- Nullhypothese 71
- p -Werte 68-72
- Parameter 92
- Poisson-Verteilung 46
- Polynom-Kernel 227-231
- Präzision 144
- Präzision-Recall-Diagramm 161-162
- R^2 (R -Quadrat) 63-67
- Radial-Kernel 232
- Recall 144
- ReLU-Aktivierungsfunktion 239
- Residuum 58
- Richtig-Positive-Rate 146
- ROC 147-157
- Schichten (Neuronale Netze) 238
- Sensitivität 143
- Sigmoid-Aktivierungsfunktion 239
- Soft Margin 224
- SoftPlus-Aktivierungsfunktion 239
- Spezifität 143
- Stochastischer Gradientenabstieg 105
- Summe der quadrierten Residuen (SSR) 58-60
- Support-Vektor 224
- Support-Vektor-Klassifikator 224
- Tangente 291
- Testdaten 13
- Trainingsdaten 13
- Überanpassung 16
- Unabhängige Variable 18
- Underflow 118, 131
- Verlustfunktion 88
- Verzweigung (Entscheidungsbaum) 185
- Wahrscheinlichkeitsverteilung 36
- Wahrscheinlichkeit vs. Likelihood 112-114
- Wurzelknoten (Entscheidungsbaum) 185