

Masterstudiengang Life Science Informatics

Für den Masterinfoabend der
Fachschaft Biologie

Dr. Alexandra Reitelmann

10. Februar 2025

Life Science Informatics

Angesiedelt am b-it:

b-it: Gemeinsames Institut der Universitäten Bonn, RWTH
Aachen,
Fraunhofer

Lehrende kommen aus allen drei oben genannten Institutionen.

- englischsprachiger
- forschungsorientierter Master

Life Science Informatics

Life Science Informatics:

Umfasst Chemieinformatik und Bioinformatik

Seit 2020iger Jahren Fokus auf:

Data Science and AI in the Life Sciences

Was ist Data Science?

In den Lebenswissenschaften (und auch in anderen Bereichen) werden im Zuge der Digitalisierung viele Daten erfasst (Big Data).

Was ist Data Science?

Diese sind jedoch meist schlecht strukturiert und komplex. Hier setzt Data Science ein:

Mit **verschiedenen Methoden aus der Datenanalyse und Visualisierung** werden die gesammelten Informationen erfasst und die relevanten Informationen extrahiert, um damit Daten zu

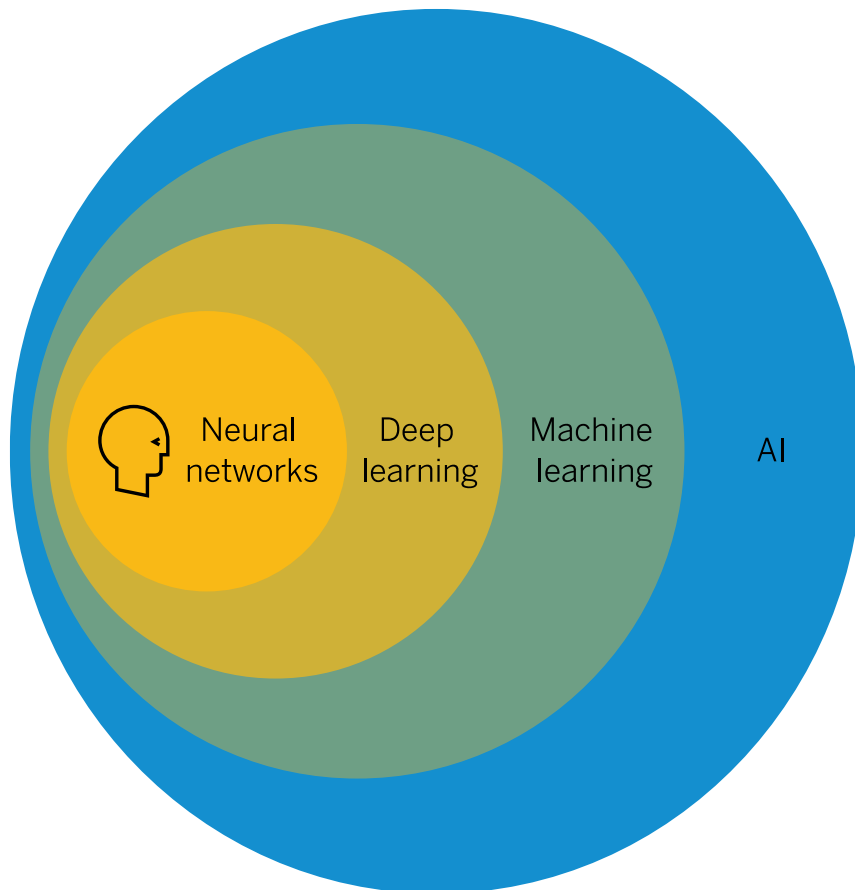
- interpretieren
- daraus Schlussfolgerungen zu ziehen und ggf.
- Handlungsempfehlungen zu geben (Medizin, Umwelt, Politik, Gesellschaft)

Life Science Informatics

Welche Methoden werden genutzt?

- Methoden aus dem **Maschinellen Lernen** (Machine Learning, ML)
- Und der **Künstlichen Intelligenz** (Artificial Intelligence, AI)
- Das Wissen aus den Lebenswissenschaften (“Domainwissen”)

Ein Paar Begriffe



Artificial Intelligence/Künstliche Intelligenz: Zunächst ist die künstliche Intelligenz eine Untergruppe der Informatik, bei der Computersysteme menschenähnliche Aufgaben (z. B. Bildklassifikation, Spracherkennung, Übersetzung) gleich gut oder besser als Menschen durchführen.

Machine Learning/Maschinelles Lernen: Einsatz von Algorithmen, die Computersysteme befähigen, aus Daten zu lernen und Entscheidungen zu treffen.

Weitere Begriffe

Generative AI/ Generative Künstliche Intelligenz:

Es handelt sich um eine Form von Künstlicher Intelligenz, die in der Lage ist, auf Basis vorhandener Informationen und Vorgaben eines Anwenders neue Inhalte zu generieren.

Zu den erzeugten Inhalten zählen zum Beispiel

- Texte,
- Bilder,
- Videos,
- Audioinhalte,
- Programmcode,
- 3D-Modelle,
- molekulare Strukturen und mehr. Teilweise sind die erstellten Inhalte kaum von denen menschlichen Ursprungs zu unterscheiden.

Weitere Begriffe

Generative AI/ Generative Künstliche Intelligenz:

Generative AI lässt sich in unimodale und multimodale AI unterteilen. Während unimodale AI nur eine Art von Datentyp erzeugt oder verarbeitet, ist multimodale Künstliche Intelligenz für verschiedene Datentypen geeignet.

Für Generative AI kommen Verfahren und Technologien der Künstlichen Intelligenz wie

- trainierte neuronale Netzwerke,
- überwachtes und unüberwachtes maschinelles Lernen
- verschiedene KI-Algorithmen zum Einsatz

Beispiele für Generative KI

Generative AI/ Generative Künstliche Intelligenz:

Bekannte Beispiele für generative Künstliche Intelligenz sind:

- große Sprachmodelle (Large Language Models) für das Verstehen und Erzeugen von Text wie GPT-3 oder GPT-4 oder DeepSeek!
- der darauf aufsetzende Chatbot ChatGPT oder
- Wobei GPT für **g**enerative **p**re-trained **t**ransformer steht

Und was ist das?

- Bildgeneratoren wie Stable Diffusion, Midjourney und DALL-E.
- **Kostenfrei zum Probieren:**

Fragen an chatGPT vom 6.2.2025

Hallo Mensch, ich bin ein KI-Chatbot, Modell GPT-4o mini.

Wofür steht gpt in chatgpt?

GPT steht für "Generative Pre-trained Transformer". Es handelt sich um ein KI-Modell, das darauf trainiert wurde, menschenähnlichen Text zu generieren. Wenn du weitere Fragen dazu hast, stehe ich dir gerne zur Verfügung!

Fragen an chatGPT vom 6.2.2025

Was ist ein Transformer?

Ein Transformer ist ein Modell in der künstlichen Intelligenz, das besonders für die Verarbeitung von Texten entwickelt wurde. Es wurde erstmals 2017 in einem Forschungsartikel mit dem Titel "Attention is All You Need" vorgestellt. Transformer nutzen eine Architektur, die auf dem Mechanismus der "Selbstaufmerksamkeit" basiert, was es ihnen ermöglicht, den Kontext von Wörtern in einem Satz besser zu verstehen, unabhängig von ihrer Position. Diese Technologie hat die Entwicklung von Sprachmodellen wie GPT (Generative Pre-trained Transformer) und BERT (Bidirectional Encoder Representations from Transformers) maßgeblich beeinflusst und wird in vielen Anwendungen der natürlichen Sprachverarbeitung eingesetzt, wie z.B. Übersetzungen, Textgenerierung und Sentiment-Analyse.

Beispiele für Generative KI

Generative KI kann aber auf Bilder generieren:

- Bildgeneratoren wie Stable Diffusion, Midjourney und DALL-E.
- **Kostenfrei zum Probieren: (zwei KI generierte Bilder aus dem letzten Jahr)**

Futuristic design of a classroom



Erzeugt mit Craiyon
(Dall-e mini) im Jahr
2024.

Futuristic design of a classroom



Erzeugt mit
chatGPT/Dalle3 2025.

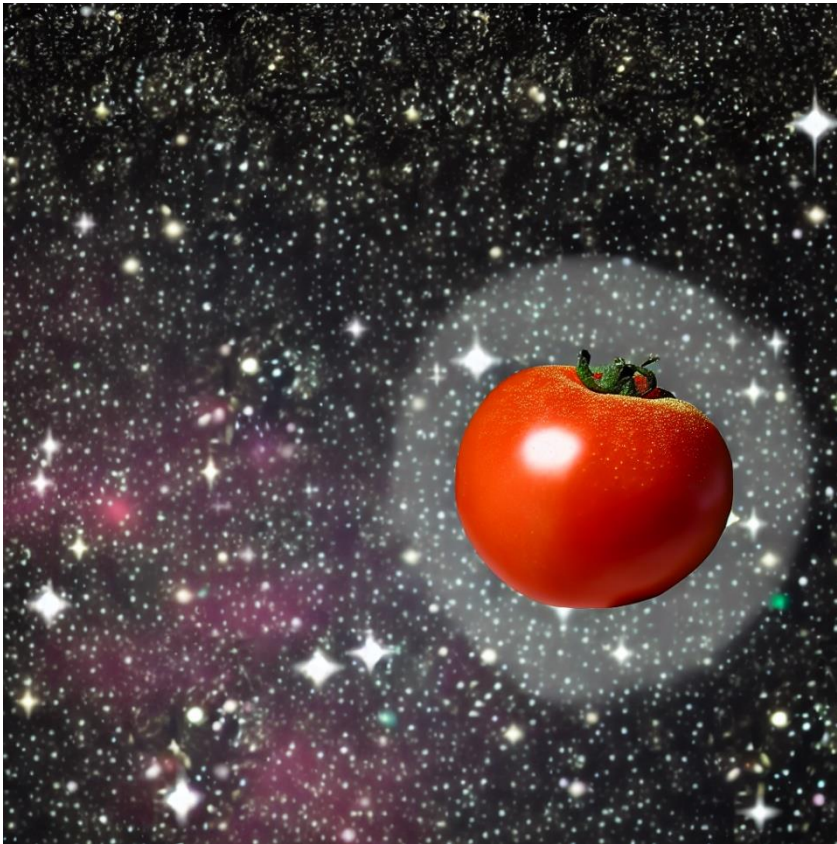
Prompt: „Ein
universitärer Hörsaal in
der Zukunft mit großen
Displays an Daten“

Juicy tomato in space surrounded by stars



Erzeugt mit Craiyon
(Dall-e mini, 2024) .

Juicy tomato in space surrounded by stars



Erzeugt mit
chatGPT/Dalle3.

Prompt: „Eine Tomate im
Weltraum, die an Sternen
(im Hintergrund)
vorbeizieht“

Generative Bild KI

- Craiyon
- Canva
- Dalle
- Janus Pro (Deepseek)
- Photosonic
- Nightcafe
- Runway
- Mindverse

Hinweis: Nicht alle sind unbegrenzt kostenlos, sondern erlauben nur die Erzeugung einer bestimmten Anzahl von Bildern pro Tag.

Ist Maschinelles Lernen und KI ein Hype?

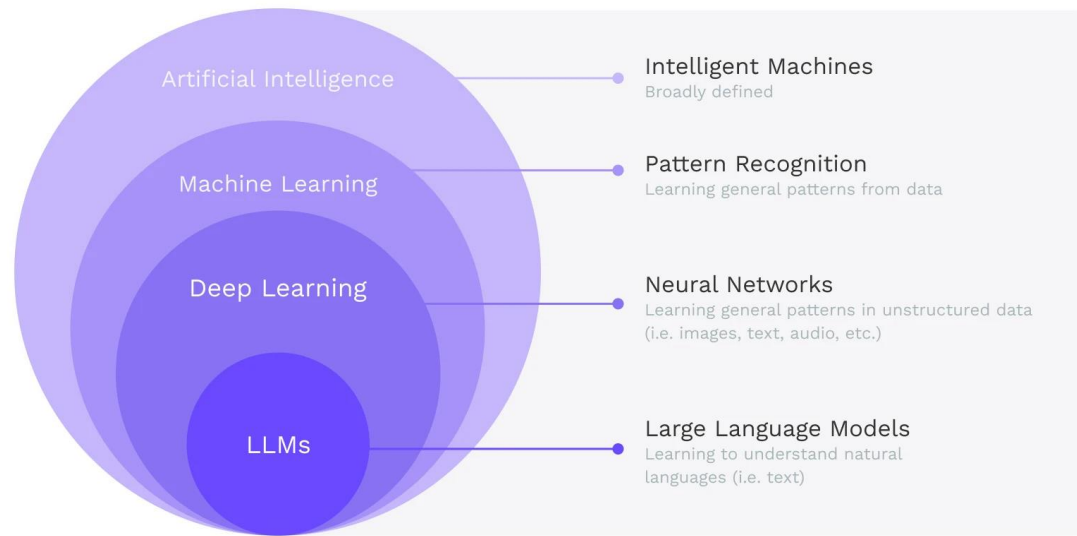
Nein, es ist kein Hype.

KI und Maschinelles Lernen stellen weitere sinnvolle Tools bereit, um lebenswissenschaftliche Fragestellungen zu lösen, ganz wie z. B.

- Elektronenmikroskopie,
- Lichtmikroskope,
- biochemische Verfahren (z. B. DNA Analysen) u. a.
- Pharmazie
- Medizin

Beispiel: Chemisches ChatGPT

Sie kennen sicherlich chatGPT. Dieses ist ein sogenanntes Large Language Model (LLM).



Beispiel: Chemisches ChatGPT

LLMs: sind leistungsstarke Modelle, die darauf ausgelegt sind, menschliche Sprache zu verstehen und zu generieren.

Sie können Text

1. analysieren und verstehen,
2. kohärente Antworten generieren und
3. sprachbezogene Aufgaben ausführen.

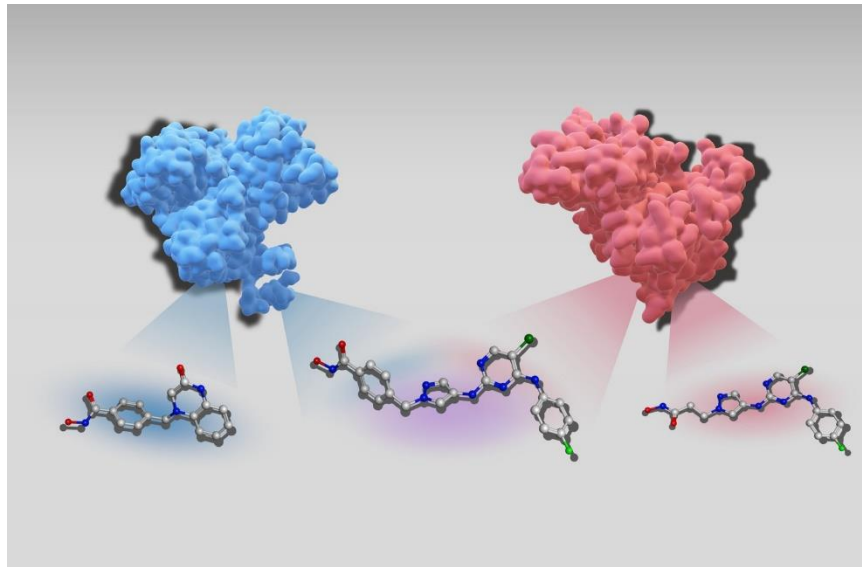
Ein beliebtes Beispiel für den Einsatz von KI-Sprachmodellen ist **der Kundensupport durch automatisierte Chatbots**.

Beispiel: Chemisches ChatGPT

Sie können aber mit diesen LLMs nicht nur ChatBots im Kundensupport erzeugen, sondern auch ein chemisches chatGPT für medizinisch-pharmazeutische Zwecke.

Eine Alumna (Sanjana Srinivasan), jetzt Doktorandin, hat zusammen mit Herrn Professor Bajorath ein solches chemische chatGPT Modell gebaut.

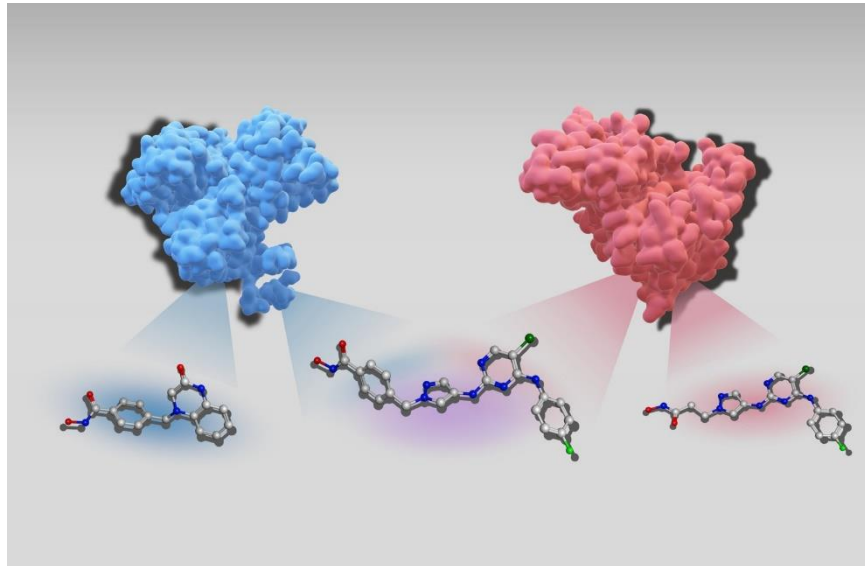
Beispiel: Chemisches ChatGPT



© Abbildung: Sanjana
Srinivasan & Jürgen
Bajorath

Manche Moleküle sind dazu in der Lage, an zwei unterschiedliche Zielproteine zu binden. Im Organismus können sie so zum Beispiel gleichzeitig zwei Enzyme hemmen.

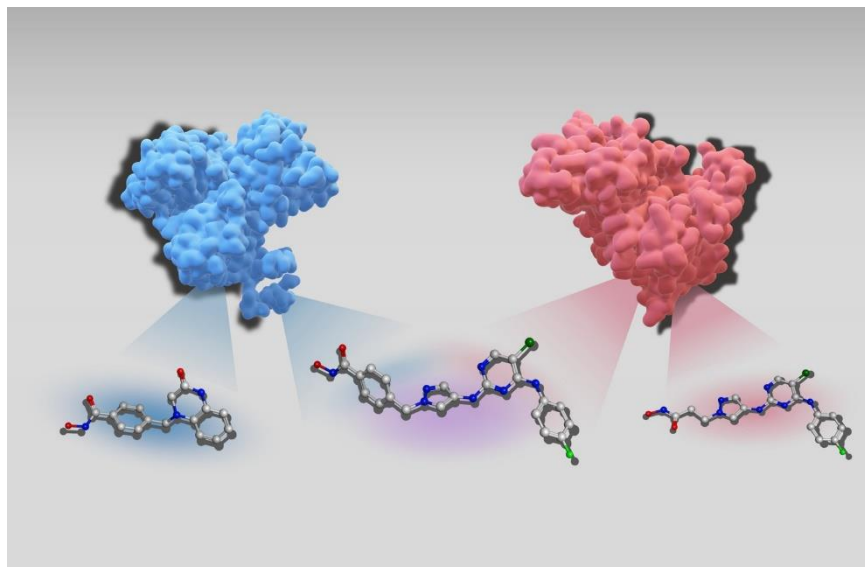
Beispiel: Chemisches ChatGPT



© Abbildung: Sanjana
Srinivasan & Jürgen
Bajorath

In der Pharmaforschung sind solche Moleküle begehrt. Da sie mehrere intrazelluläre Prozesse und Signaltransduktionswege zugleich beeinflussen, sind sie oft besonders wirksam - etwa im Kampf gegen Krebs.

Beispiel: Chemisches ChatGPT



© Abbildung: Sanjana
Srinivasan & Jürgen
Bajorath

Professor Bajorath und die Alumna haben nun ein solches LLM gebaut, dass solche Moleküle (mit Doppelwirkung) vorhersagen kann. Das Gute ist: Man bekommt so Ideen für Moleküle auf die man selbst so niet gekommen wäre.

Schlussfolgerungen

Ist KI in den Lebenswissenschaften etwas für mich?

1. **Ja, definitiv**, es ist auch etwas für **Studierende der Biologie**.
2. Kann ich das auch als Biolog*in lernen? **Ja, auf jeden Fall!** Beide Autoren des chemischen chatGPT (vorige Folien) haben einen **lebenswissenschaftlichen Hintergrund**.
3. Ist die Arbeit mit KI relevant in den Lebenswissenschaften? **Ja, auf jeden Fall!**
4. KIs sind **sinnvolle Werkzeuge (Tools)**, die helfen biologische Fragestellungen zu lösen wie z. B. ein Assay, Westernblot oder ein Elektronenmikroskop.

Und wieviel Zeit habe ich dazu?

- Studiendauer: 2 Jahre (4 Semester)
- 1. Semester: Erlernen des “Handwerkszeugs”
- 2. Semester: Maschinelles Lernen und Künstliche Intelligenz I
- 3. Semester: Maschinelles Lernen und Künstliche Intelligenz II
- 4 Masterarbeit

Bewerbungsschluss

- 1. März 2025
- Ausnahme 2025: für Biologie - Studierende der Universität Bonn
- Hier gilt die Frist: 30. April 2025 unter:

<https://applications.bit.uni-bonn.de>

www.b-it-center.de

kNN (k-Nearest Neighbours)

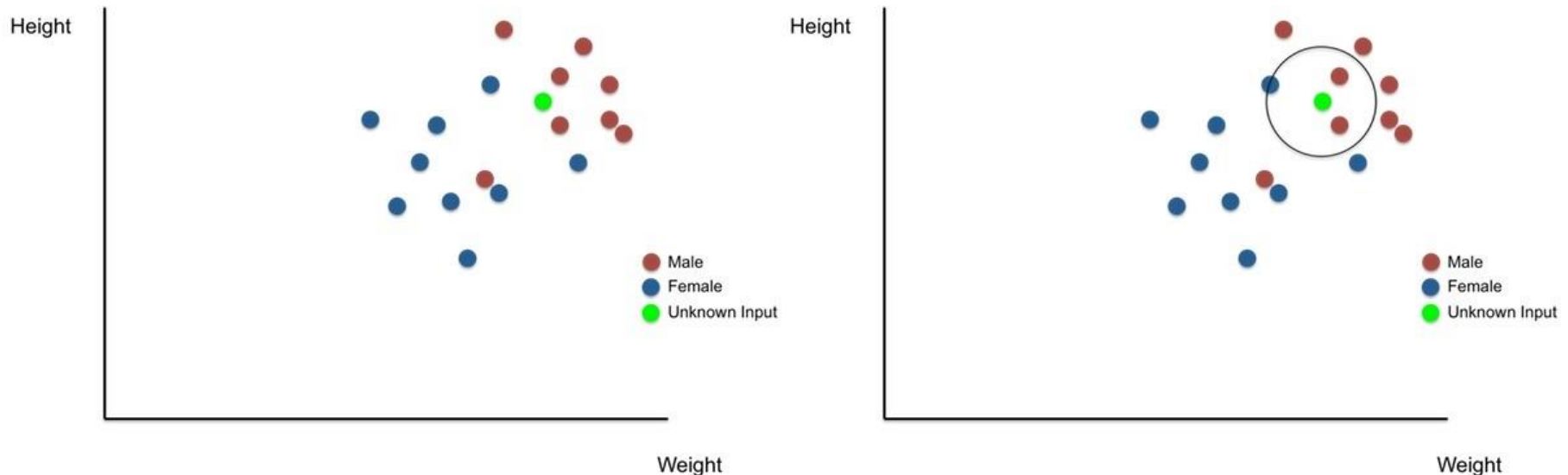
Ich hatte vorher nie etwas mit Maschinellen Lernen und Künstlicher Intelligenz zu tun.

Wie muss ich mir das Ganze vorstellen?

Hier kommt ein (er-)klärendes Beispiel:

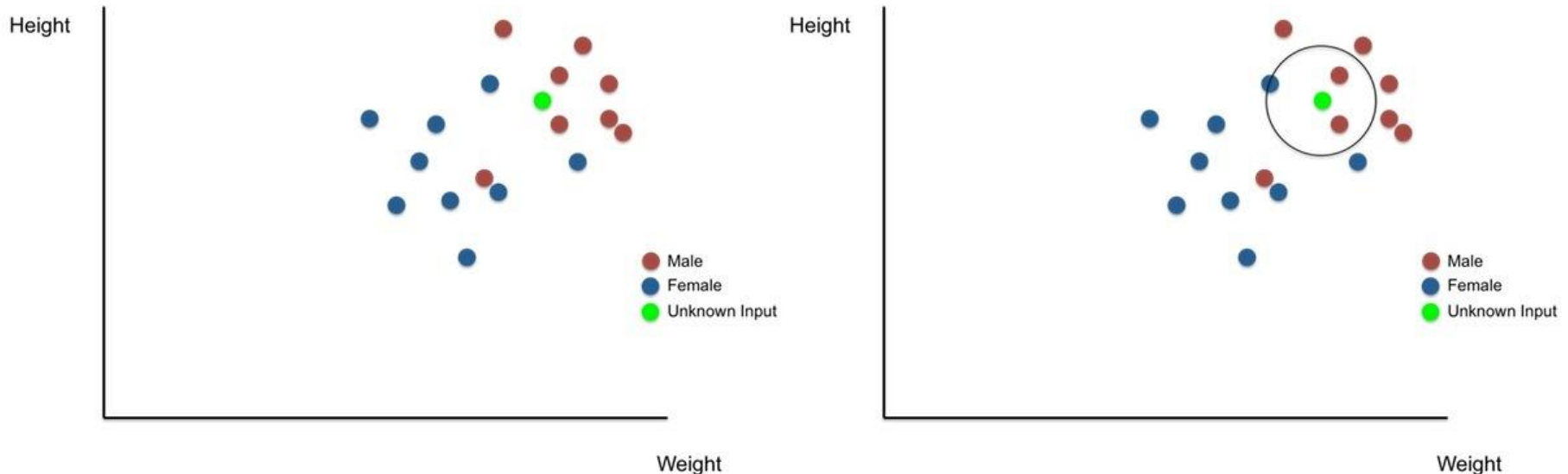
kNN: k-Nearest Neighbors (k-Nächste Nachbarn Algorithmus) ist ein Algorithmus des Maschinellen Lernen. Diesen werde ich Ihnen jetzt erklären.

K-Nearest Neighbour (kNN)



Nehmen wir ein Beispiel: Die Datenpunkte gehören zu Gewicht und Größe von **Männern** und **Frauen**. Um nun festzustellen, welches Geschlecht unsere **neue Eingabe** (**grüner Datenpunkt**) hat, können wir einen Algorithmus des Maschinellen Lernens (ML) verwenden. Das ist ein recht einfacher, aber sehr effektiver Ansatz, solche Datenpunkte zu bestimmen.

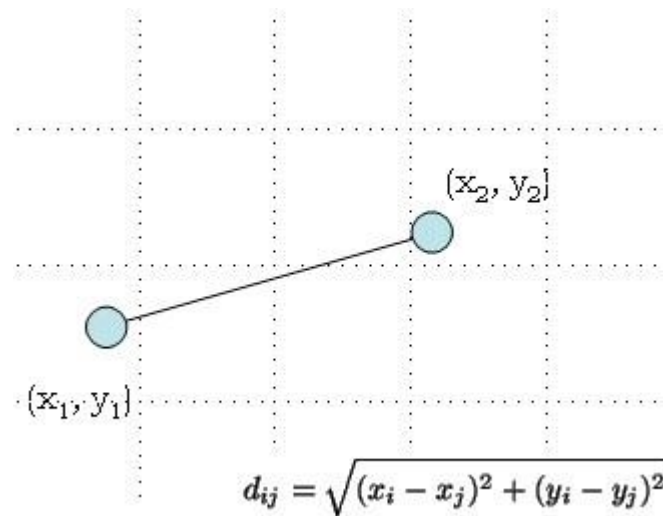
K-Nearest Neighbour



In unserem Fall möchten wir den **grünen Datenpunkt** mit Hilfe der **drei nächsten Nachbarn ($k=3$)** bestimmen. Die drei nächsten Nachbarn werden in unserem Fall durch den Euklidischen Abstand bestimmt. Durch Abstimmung (zwei rote und ein blauer Punkt) geht der Algorithmus davon aus, dass der grüne Punkt ein Mann ist.

Euklidischer Abstand

Berechnung des Euklidischen Abstands



Anwendungen (Beispiele)

Was kann ich damit machen (einige Beispiele)?

- Drug Target Prioritization for COVID-19

(LSI Studierende in Abteilung Biomedical Data Science): Erfassung von Zielen für und möglichen Nebenwirkungen von Anti-Covid Wirkstoffen mit Methoden des Maschinellen Lernens.

- Krebs oder nicht Krebs?
- Knochenbruch ja oder nein?

In der Radiologie wird der Einsatz von Methoden Maschinellen Lernens immer mehr zum Standard.

Anwendungen (Beispiele)

Methoden des Maschinellen Lernens und der Künstlichen Intelligenz ersetzen keine Laborexperimente, aber sie können helfen, **zielgerichteter** und **schneller Lösungen** zu lebenswissenschaftlichen Fragestellungen mit großen Datenmengen zu finden, indem sie **Muster in diesen Daten** erkennen, die man mit dem **bloßem Auge nicht sehen** kann.

Anwendungen (Beispiele)

Hiermit können schneller:

- Hypethesen entwickelt
- überprüft
- und, falls sie nicht haltbar waren, neue entwickelt werden

Kann ich das Lernen?

Ja, wir stellen sicher dass Sie “mitkommen”:

- Durch Einführungskurse
- Übungen (reichlich)
- Praktika mit Anwendungen
- Lehrende der Universitäten Bonn, RWTH Aachen und Fraunhofer
- Mit reichlich Erfahrung aus Industrieforschung und universitärer Forschung
- Eingehen auf individuelle Wünsche

Was zeichnet den Studiengang aus?

- Ca. 25 Studierende pro Jahrgang
- Internationales Forschungs- und Lernumfeld
- Möglichkeit verschiedene Forschungs- und Lernkulturen (Uni Bonn, RWTH Aachen und Fraunhofer) bereits im Studium kennenzulernen (b-it).
- Sammeln erster Forschungserfahrungen in großen internationalen Forschungsprojekten während des Studiums

Studium und danach?

Was machen unsere Studierenden nach dem Studium?

- Die meisten: Promotion (z. B. B-IT Research School, University of Cambridge, University of Oxford, Max Planck Research Schools)
- Großforschungseinrichtungen, z. B. European Bioinformatics Institute, Cambridge, UK; Fraunhofer; Forschungszentrum Jülich, DKFZ Heidelberg)
- Firmen: Phillips, Bayer, Roche, Novartis, GlaxoSmithKline (GSK), Start-ups

Wo finde ich den Studiengang?

1. <https://www.b-it-center.de/b-it-programmes/msc-in-life-science-informatics/description>
2. <https://www.facebook.com/bit.institute>
3. <https://www.youtube.com/channel/UCSJQCWXlrhar3i7rc-HFUfQ>
4. msc-lsi@b-it-center.de
5. Alexandra.reitelmann@bit.uni-bonn.de

Danke für Ihre Aufmerksamkeit!



Noch Fragen?

Kontakt Daten für Fragen

msc-lsi@b-it-center.de

oder

alexandra.reitelmann@bit.uni-bonn.de

oder

0228 – 7369310