

Unsupervised Learning (Nicht-überwachtes Lernen)

Eigenschaften:

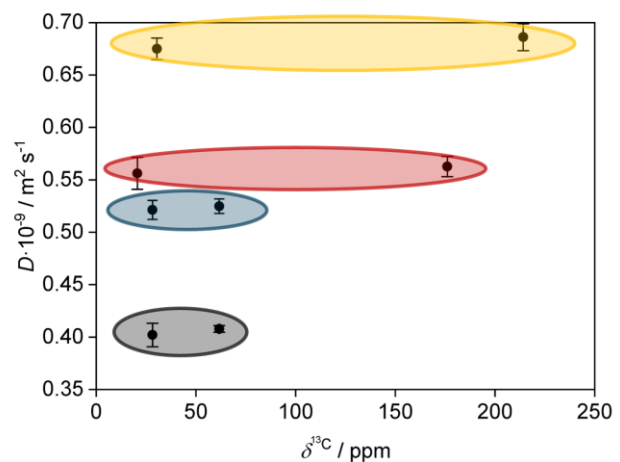
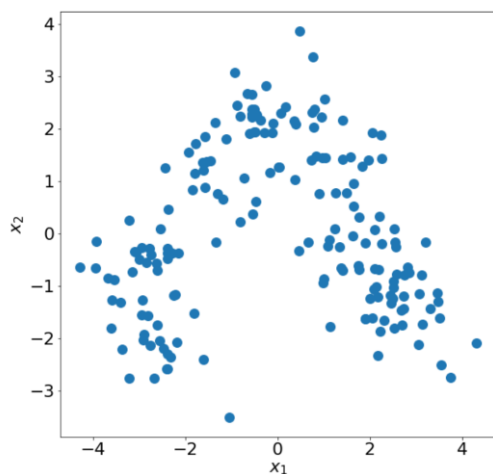
- Die konkreten Zielgrößen sind nicht (notwendigerweise) bekannt.
- Suche nach Strukturen, Mustern oder Erklärungen.

Ziele:

- Entwickeln eines Verständnisses für die Daten.
- Identifikation der zur Beschreibung der Daten notwendigen Variablen.
- Reduktion der Zahl der Variablen → *Dimensionsreduktion*.
- Auffinden von Gruppierungen in den Daten → *Clustering*.

Beispiele:

- Identifikation von Ausreißern in Stoff- oder Prozessdaten, z.B. zur Qualitätskontrolle oder Anomaliedetektion.
- Definition von charakteristischen Stoffgruppen auf Basis von Stoffdaten.
- Auswahl geeigneter Reinstoffdeskriptoren zur Beschreibung von Stoffdaten.
- Definition von Pseudokomponenten.



Quelle: T. Specht, K. Münnemann, H. Hasse, F. Jirasek, PCCP 25 (2023) 10288-10300.

Supervised Learning (Überwachtes Lernen)

Eigenschaften:

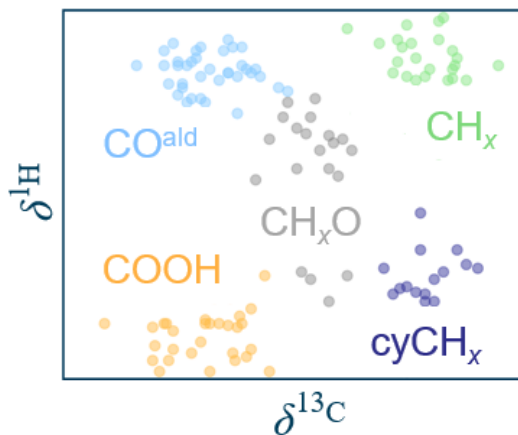
- Die konkreten Zielgrößen sind bekannt („*Ich weiß, nach was ich suche*“).
- Zu jedem Datenpunkt gibt es einen „richtigen Wert“ für die Zielgröße („*Ich weiß, was ich finden will*“).
- Es ist ein Set von *gelabelten* Trainingsdaten vorhanden.

Ziele:

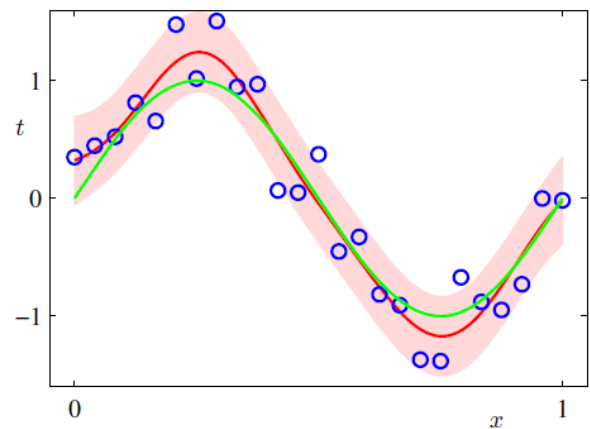
- Entwicklung eines Modells:
 - Zur Reproduzierung der korrekten Zielgrößen im Trainingsset.
 - Zur Vorhersage der korrekten Zielgrößen neuer, ungelabelter Datenpunkte.
- Dabei wird anhand der Art der Zielgrößen unterschieden:
 - Kann diese nur diskrete Werte annehmen, sprechen wir von *Klassifikation*.
 - Kann diese kontinuierliche Werte annehmen, sprechen wir von *Regression*.

Beispiele:

- Beurteilung von Prozessverhalten: fehlerfrei / fehlerhaft.
- Vorhersage von Stoffdaten auf Basis molekularer Deskriptoren.
- Auswertung chemischer Analysen.



Quelle: T. Specht, K. Münnemann, H. Hasse, F. Jirasek, *J. Chem. Inf. Model.* 61 (2021) 143-155.



Quelle: C. Bishop., *M. Pattern Recognition and Machine Learning*; 1st ed.; Springer: Berlin, 2006

Reinforcement Learning (Bestärkendes Lernen)

Eigenschaften:

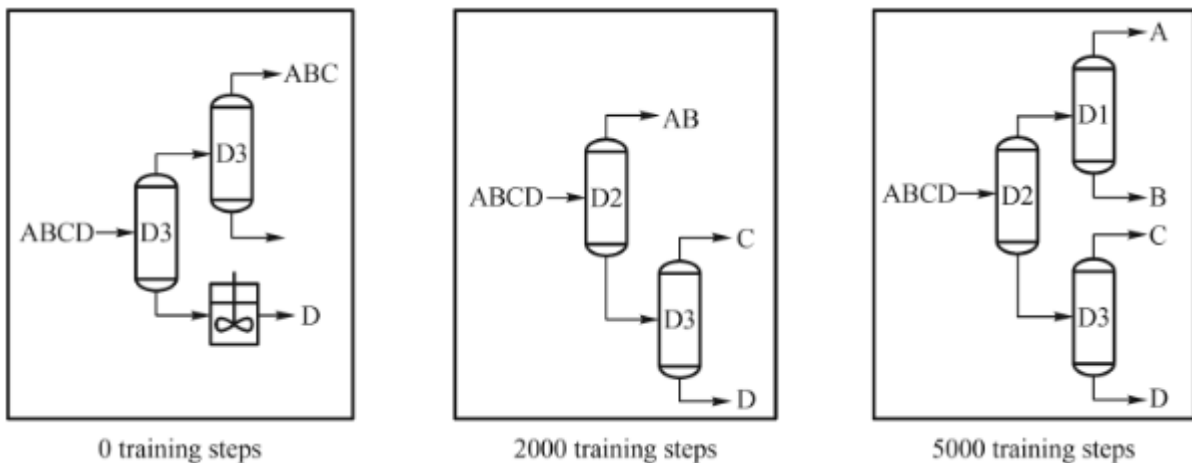
- Die korrekten Zielgrößen sind nicht bekannt.
- Das Programm kann *Entscheidungen* treffen, die *Konsequenzen* haben.
- Statt einer konkreten Zielgröße gibt es eine relative *Qualitätskennzahl*, die uns sagt, ob unser Programm im Trainingsprozess besser oder schlechter wird.
- Das Feedback über die Qualitätskennzahl ist in der Regel verzögert, damit sind die Konsequenzen *nicht direkt einzelnen* Entscheidungen zuzuordnen.

Ziele:

- Entwickeln eines „guten“ Programms zur Lösung eines Problems auf Basis einer definierten Qualitätskennzahl und einem Trial-and-Error Ansatz.

Beispiele:

- Automatisierte Flowsheet Synthesis: Finde eine günstige (optimale?) Anordnung von Apparaten, um einen Prozess mit vorgegebenen Spezifikationen auszulegen.



Quelle: Q. Göttl, D. Grimm, J. Burger, *Comput. Aided Chem. Eng.* 50 (2021), 209-214.