

Weitere ML Methoden

Caroline Sporleder

Computational Linguistics
Universität des Saarlandes

Sommersemester 2011

14.07.2011

Separierbarkeit

Die **Separierbarkeit** (engl. **separability**) der Ausgabeklassen hängt ab von:

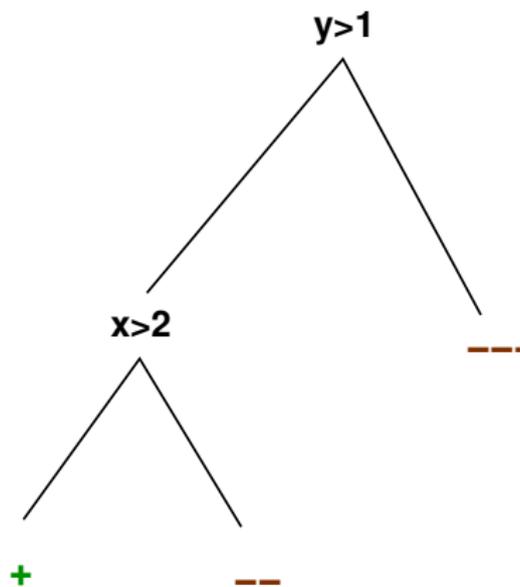
- dem Attributraum
- welche Art von Entscheidungsgrenzen, der Lerner ziehen kann

Ein Zielkonzept läßt sich nicht vollständig lernen, wenn die Ausgabeklassen im Attributraum für einen gegebenen Lerner nicht vollständig separierbar sind.

Beispiel: Separierbarkeit und Entscheidungsgrenzen

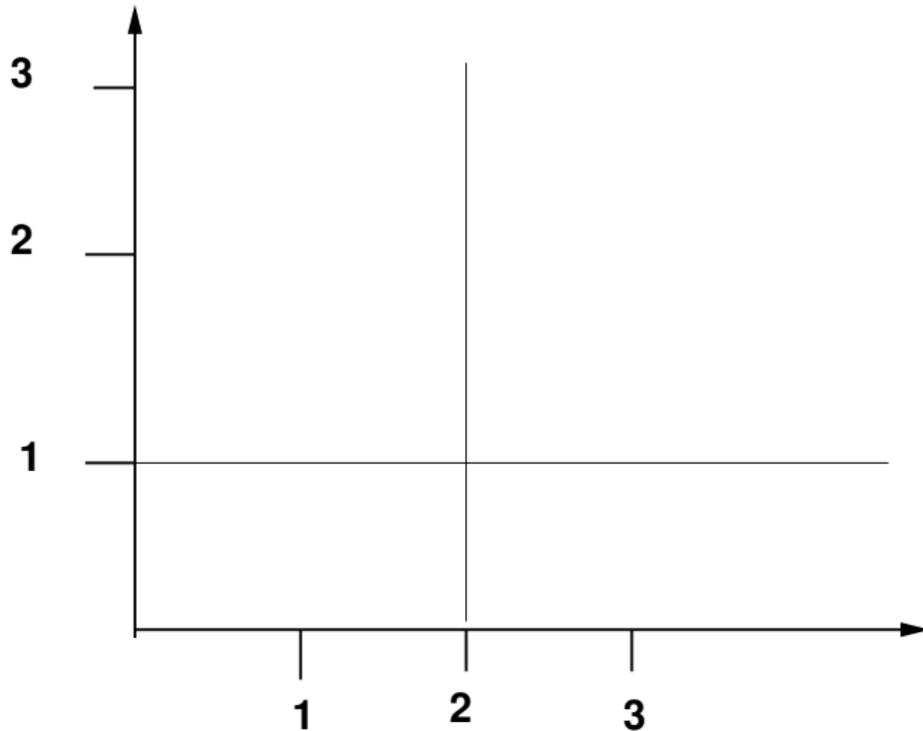
Entscheidungsbaum

- Ausgabeklasse: +, -
- Attribute: $x \rightarrow \mathbb{N}$, $y \rightarrow \mathbb{N}$



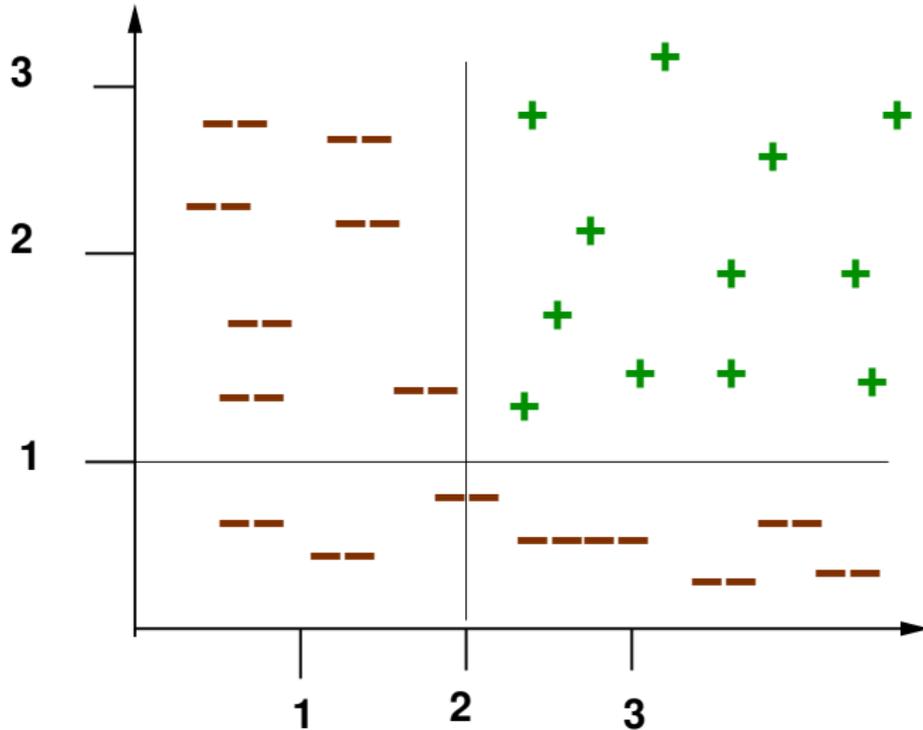
Beispiel: Separierbarkeit und Entscheidungsgrenzen

Entscheidungsbaum



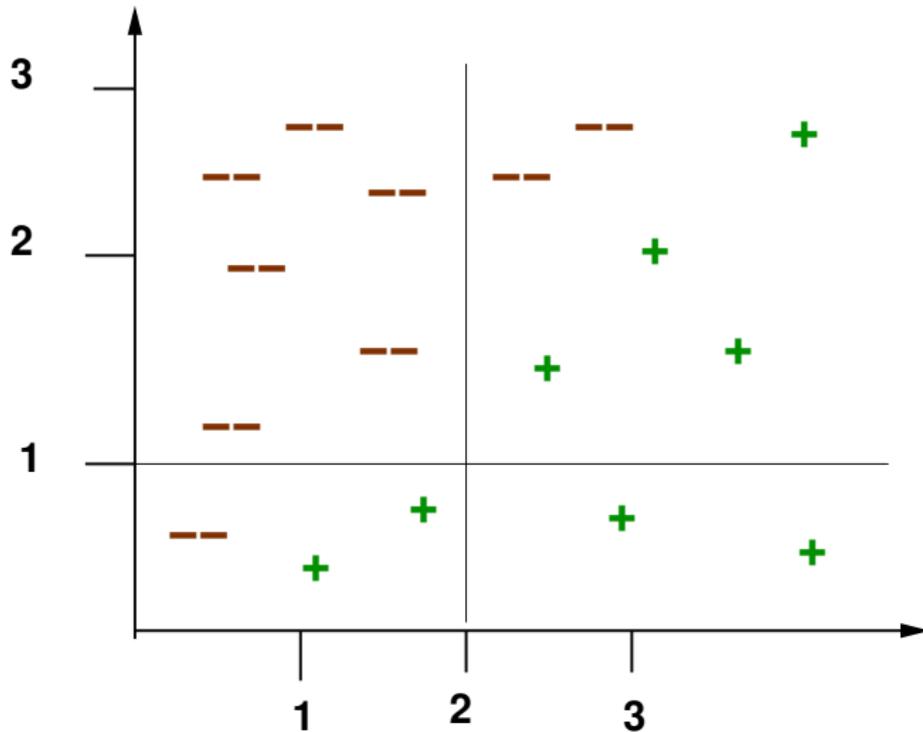
Beispiel: Separierbarkeit und Entscheidungsgrenzen

Entscheidungsbaum



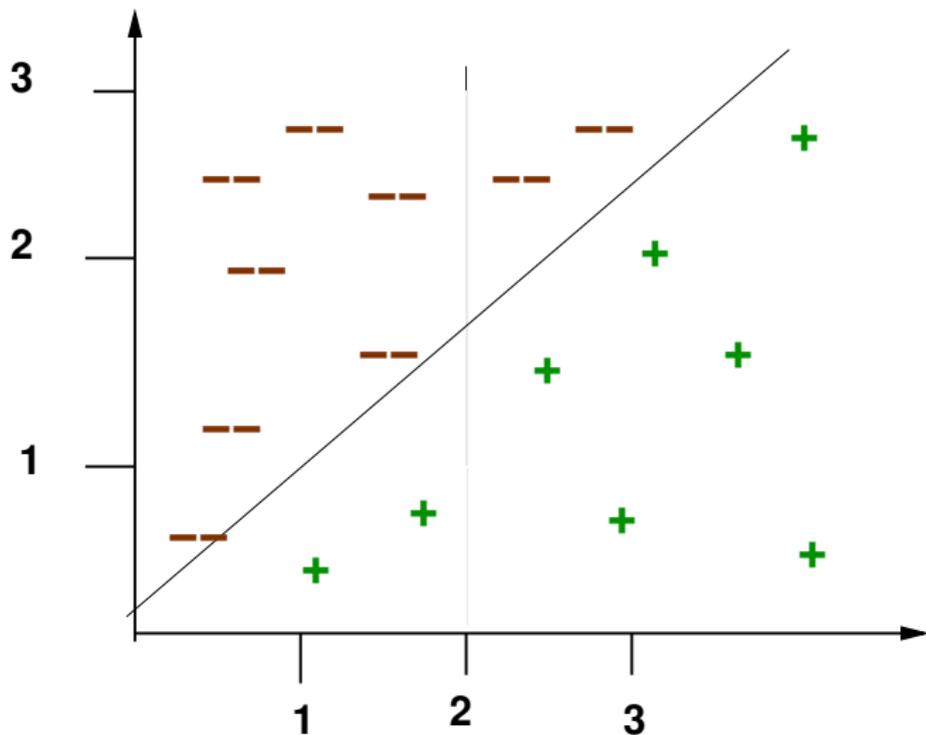
Beispiel: Separierbarkeit und Entscheidungsgrenzen

Entscheidungsbaum



Beispiel: Separierbarkeit und Entscheidungsgrenzen

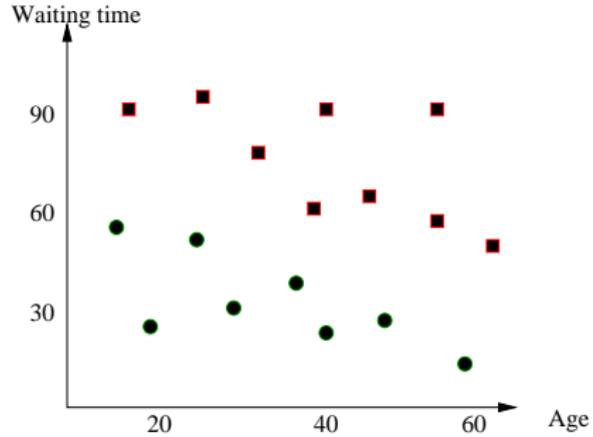
Support Vector Machine (SVM)



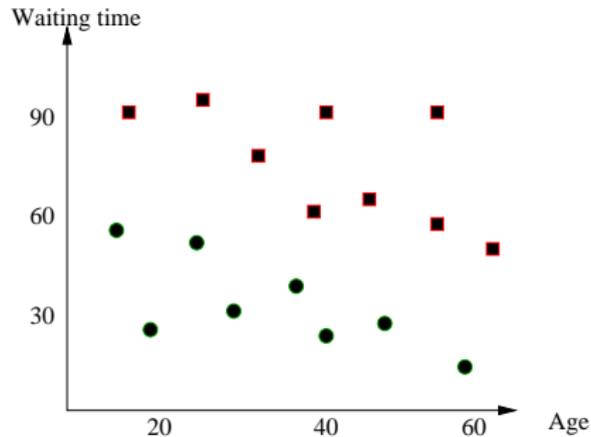
Beispiel: Vorhersage der Kundenzufriedenheit (*happy* vs. *not happy*) eines Call Centers mit den Attributen Alter und Zeit, die sie in der Warteschleife verbringen.

Beispiel: Separierbarkeit und Attributraum

Beispiel: Separierbarkeit und Attributraum

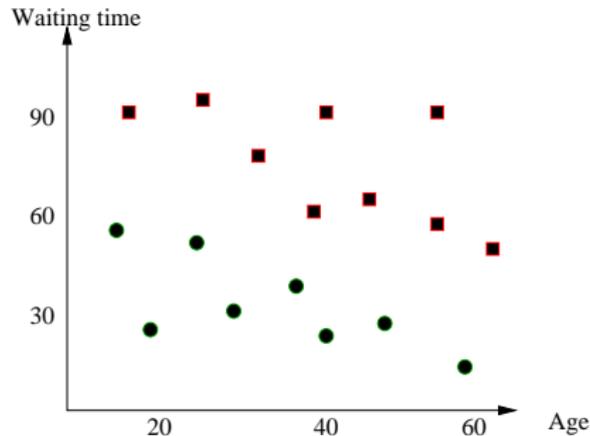


Beispiel: Separierbarkeit und Attributraum



Kann in diesem Attributraum mit Entscheidungsbäumen nicht separiert werden.

Beispiel: Separierbarkeit und Attributraum



Kann in diesem Attributraum mit Entscheidungsbäumen nicht separiert werden.

Was aber, wenn man noch ein drittes Attribut hinzufügt (z.B. 'gut gelaunt' vs. 'schlecht gelaunt')?

... kann man

- zusätzliche Attribut hinzufügen
- einen anderen Lerner (mit anderen Entscheidungsgrenzen) nehmen

Faustregel

Je größer der Attributraum (=je höher die Dimensionalität):

- desto größer die Chance, dass die Klassen separierbar sind
- **aber**

... kann man

- zusätzliche Attribut hinzufügen
- einen anderen Lerner (mit anderen Entscheidungsgrenzen) nehmen

Faustregel

Je größer der Attributraum (=je höher die Dimensionalität):

- desto größer die Chance, dass die Klassen separierbar sind
- **aber** desto größer auch die Data Sparseness (mehr Trainingsdaten notwendig)

Weitere Lernverfahren

Idee

Mehrere verschiedene Lerner werden auf den Trainingsdaten trainiert und dann auf die Testdaten angewendet. Die Ausgabeklasse wird dann z.B. durch **Voting** über die Ausgabeklassen der verschiedenen Lerner ermittelt.

⇒ Ergebnisse oft besser als für einzelne Modelle

Warum?

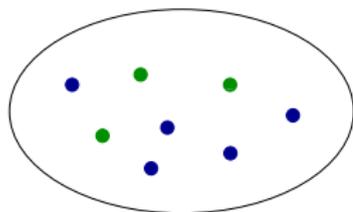
Idee

Mehrere verschiedene Lerner werden auf den Trainingsdaten trainiert und dann auf die Testdaten angewendet. Die Ausgabeklasse wird dann z.B. durch **Voting** über die Ausgabeklassen der verschiedenen Lerner ermittelt.

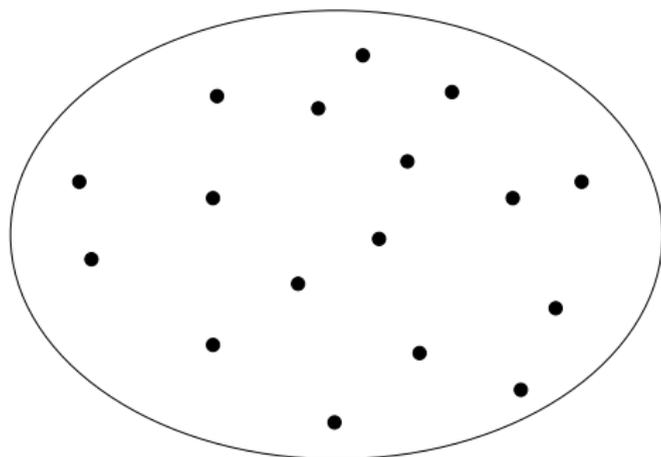
⇒ Ergebnisse oft besser als für einzelne Modelle

Warum?

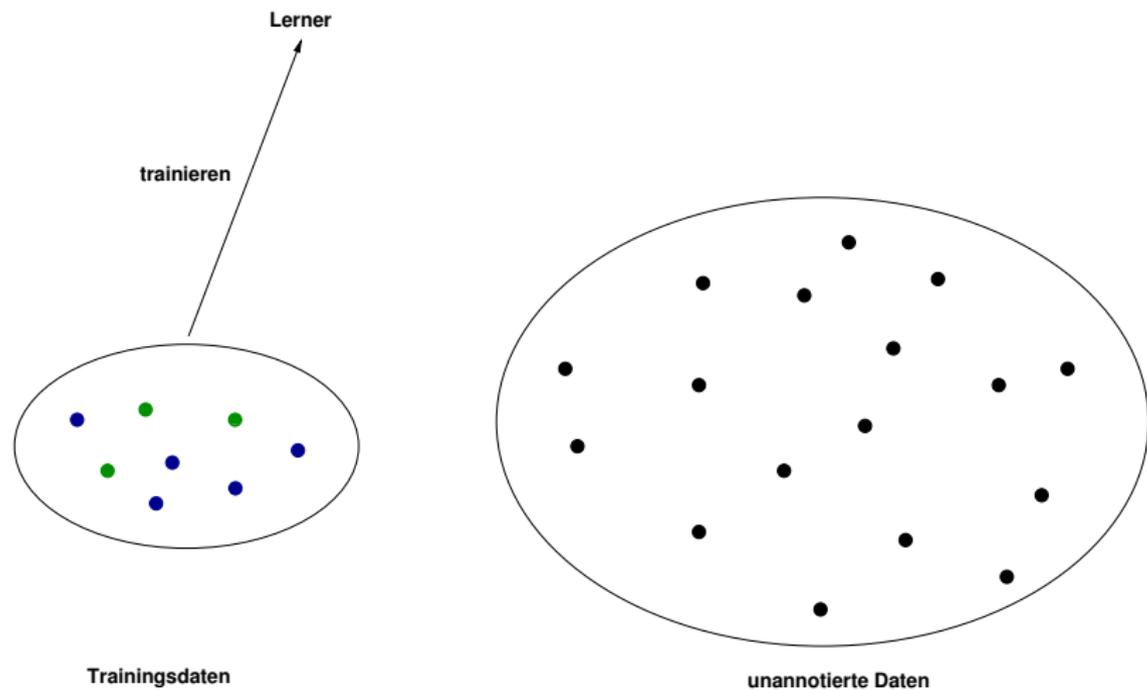
- Verschiedene Lerner haben verschiedene Stärken und Schwächen (Entscheidungsgrenzen, induktiver Bias etc.).
- Einzelne Schwächen können im Ensemble kompensiert werden.
- Dadurch können auch Hypothesen gelernt werden, die nicht in den Hypothesenräumen der einzelnen Modelle vorhanden sind.
- Ideal ist die Kombination von relativ unterschiedlichen Lernverfahren.

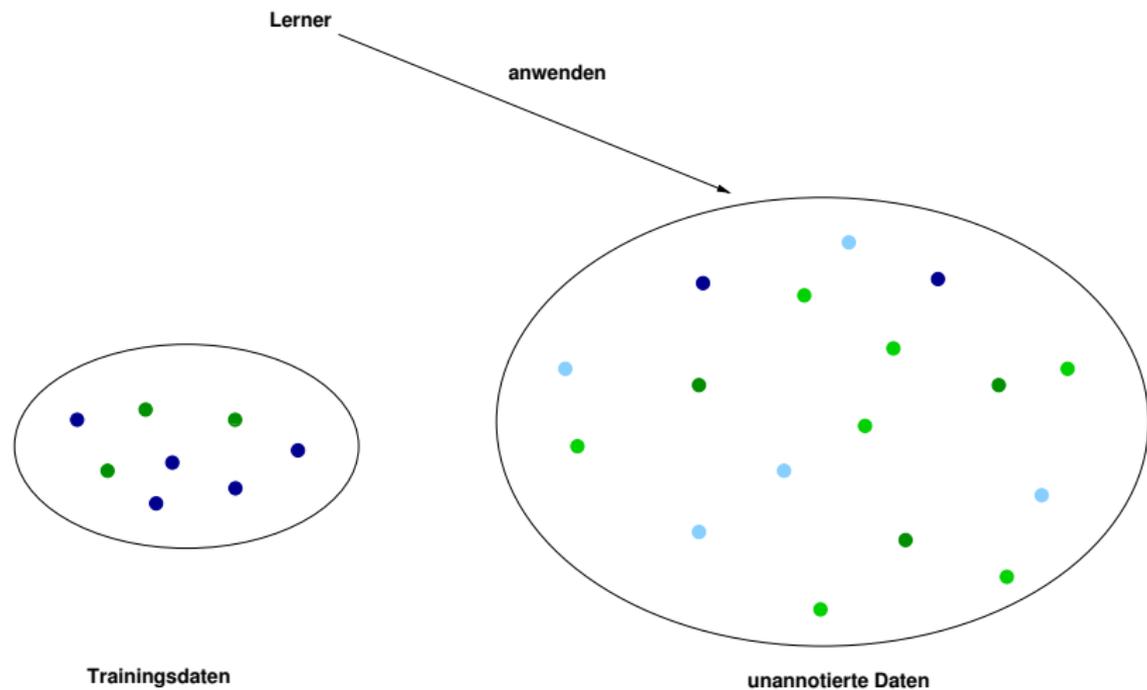


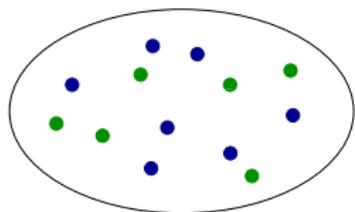
Trainingsdaten



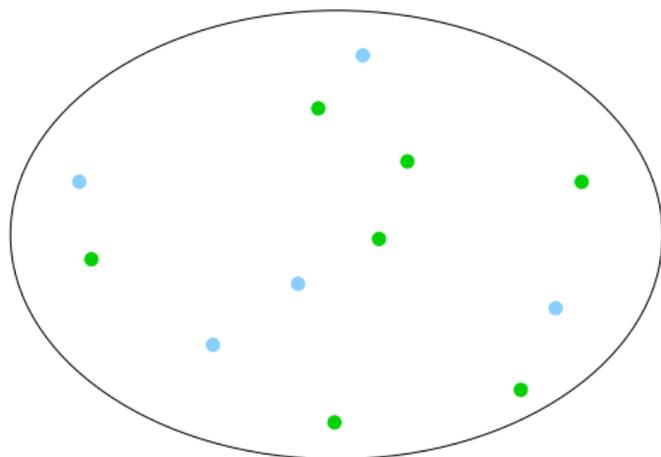
unannotierte Daten



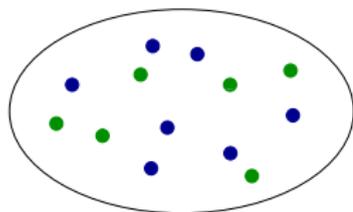




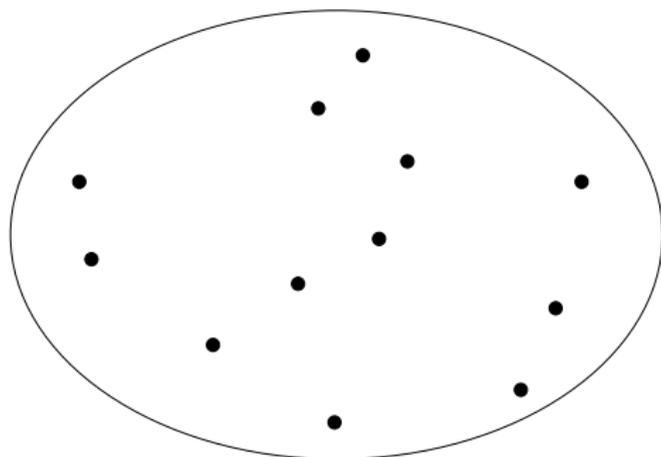
Trainingsdaten



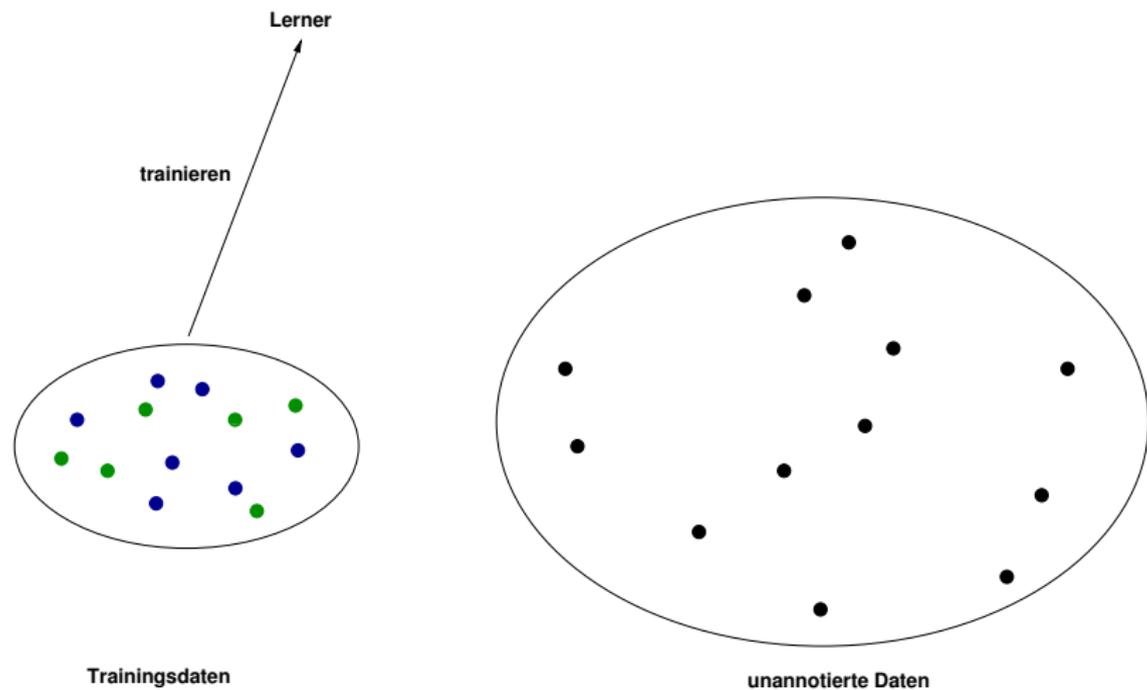
unannotierte Daten

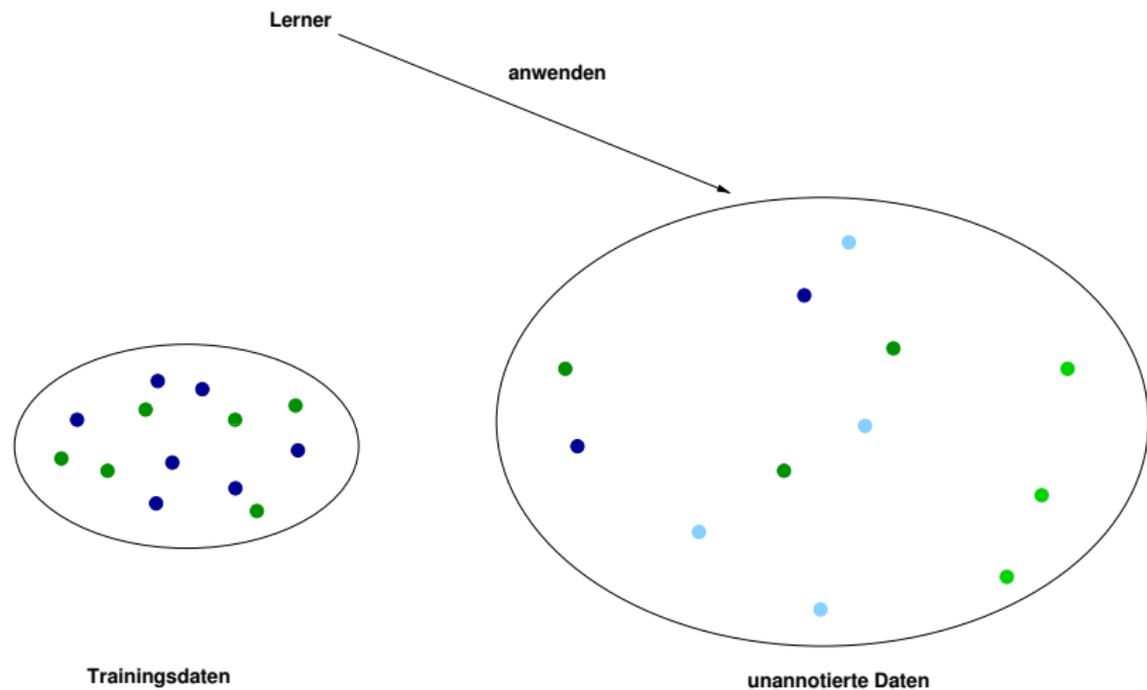


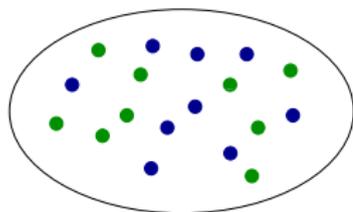
Trainingsdaten



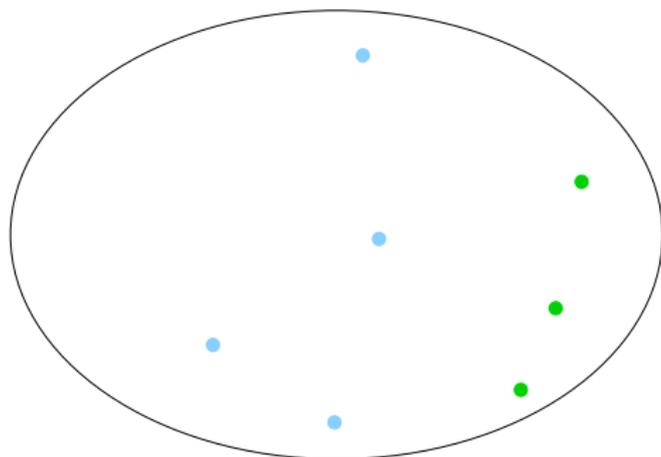
unannotierte Daten



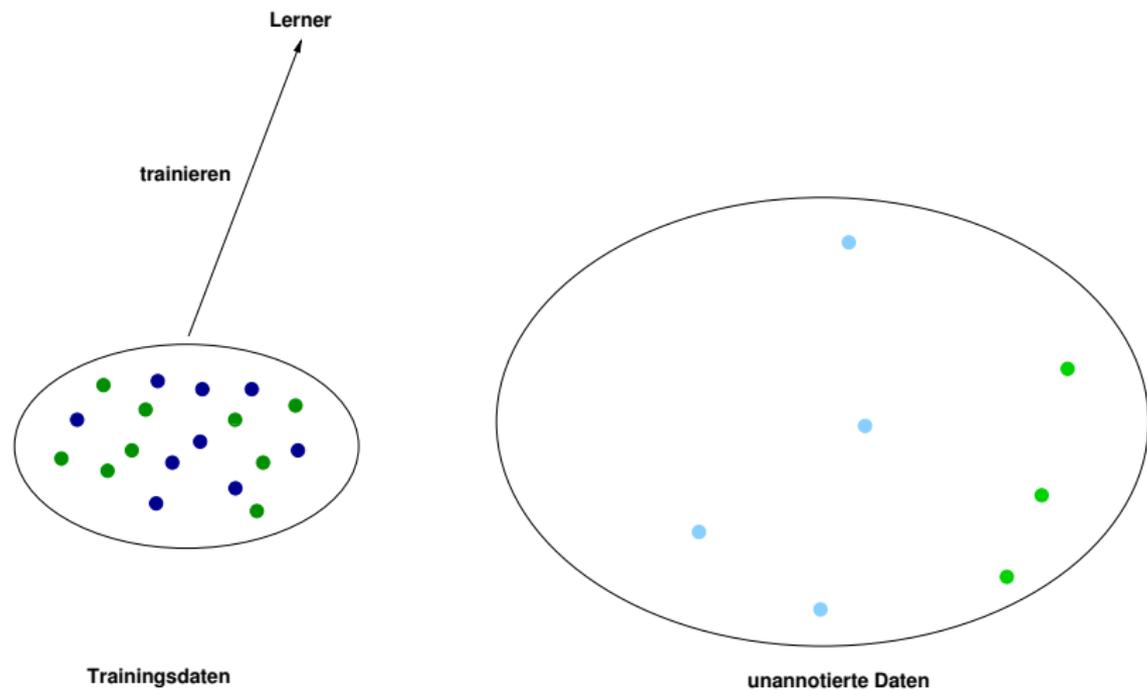




Trainingsdaten



unannotierte Daten



Self learning ist ein **teilüberwachtes Lernverfahren**, in dem ein Lerner von seinem eigenen Output lernt.

Problem

Self learning ist ein **teilüberwachtes Lernverfahren**, in dem ein Lerner von seinem eigenen Output lernt.

Problem

Funktioniert, wenn überhaupt, nur eingeschränkt. Lerner hat die Tendenz sich in seinen eigenen Fehlern zu bestärken.

Co-Training ist ebenfalls ein **teilüberwachtes Lernverfahren**.

- **Zwei** (oder mehr) **Lerner** mit verschiedenen Attributräumen (zwei verschiedene **views** auf die Daten).
- Beide Lerner werden auf unannotierte Daten angewendet.
- Für jeden Lerner werden die Instanzen, die der Lerner mit der höchsten Konfidenz labelt, zu den Trainingsdaten zugefügt.
- Dann werden beider Lerner neu trainiert.
- und so weiter . . .

⇒ De facto lernt ein Lerner von dem Output des anderen.

In der Theorie funktioniert Co-training genau dann wenn

- die views bedingt unabhängig voneinander sind
- beide Modelle die Zielhypothese in ihren Hypothesenraum haben

In der Praxis

funktioniert Co-Training auch oft, wenn diese Voraussetzungen nicht gegeben sind.

Idee

Lerner wählt sich selber die Instanzen aus, von denen er am meisten profitieren würde.

Genauer

- Lerner wird auf einer kleinen annotierten Datenmenge trainiert.
- Lerner wird auf unannotierte Daten angewendet.
- Die Instanzen, für die die Konfidenz des Lerners am geringsten ist, werden von einem menschlichen Annotator gelabelt und zum Trainingsset hinzugefügt.
- Lerner wird wieder trainiert.
- und so weiter ...

Active Learning (AL) erreicht oft mit weniger annotierten Daten die gleiche Performanz wie herkömmliche betreute Lernverfahren.

Warum?

Active Learning (AL) erreicht oft mit weniger annotierten Daten die gleiche Performanz wie herkömmliche betreute Lernverfahren.

Warum?

AL erlaubt es dem Lerner gezielt, die Regionen des Hypothesenraums zu erforschen, über die am wenigsten bekannt ist.

Active Learning (AL) erreicht oft mit weniger annotierten Daten die gleiche Performanz wie herkömmliche betreute Lernverfahren.

Warum?

AL erlaubt es dem Lerner gezielt, die Regionen des Hypothesenraums zu erforschen, über die am wenigsten bekannt ist.

Nachteil

AL ist stark vom Lerner abhängig und die annotierten Daten sind für andere Lerner oft nicht optimal.

Ensemble Learning

ist ein voll überwachtes Verfahren, das oft bessere Ergebnisse liefert als einzelne Lerner.

Active Learning

ist ein voll überwachtes Verfahren, das auf optimale Ausnutzung der annotierten Daten abzielt (und dadurch oft mit weniger Annotationen auskommt).

Self-Learning und Co-Training

sind teil-überwachte Verfahren, d.h. sie kommt mit einer reduzierten manuell annotierten Datenmenge aus und nutzen unannotierte Daten zur Performanzverbesserung.