

Machine Learning Unplugged

Lineare Klassifizierer

Michael Brinkmeier
Andreas Hüwel

Institut für Informatik
AG Didaktik der Informatik



Definition von KI

Künstliche Intelligenz

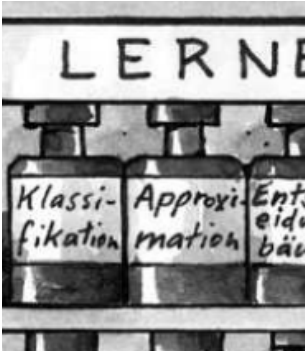
Unter dem Begriff **Künstliche Intelligenz** fasst man Methoden und Techniken der Informatik zur Konstruktion technischer Systeme zusammen, die **zielgerichtet** in Umgebungen oder auf Daten agieren, welche

- **nicht vollständig kontrollierbar**
- und/oder **dynamisch**
- und/oder **vorab ungenau bekannt** sind.

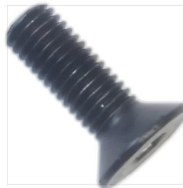
Die KI-Apotheke



Die KI-Apotheke



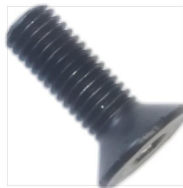
Unser Beispiel



2156x2115

Unser Beispiel

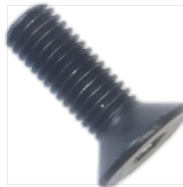
- Zwei Schraubenarten unterschiedlicher Länge



2156x2115

Unser Beispiel

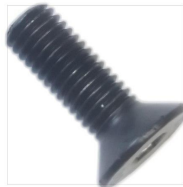
- Zwei Schraubenarten unterschiedlicher Länge
- Einzeln fotografiert auf weißem Hintergrund



2156x2115

Unser Beispiel

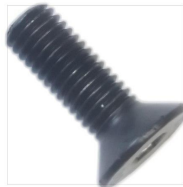
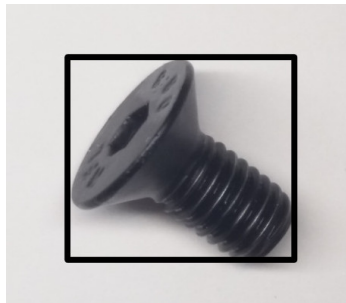
- Zwei Schraubenarten unterschiedlicher Länge
- Einzeln fotografiert auf weißem Hintergrund
- **Bounding Box** \Rightarrow Implementierbar!



2156x2115

Unser Beispiel

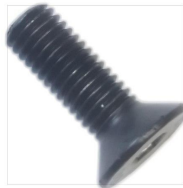
- Zwei Schraubenarten unterschiedlicher Länge
- Einzeln fotografiert auf weißem Hintergrund
- **Bounding Box** \Rightarrow Implementierbar!
- Fehlerquellen



2156x2115

Unser Beispiel

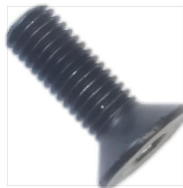
- Zwei Schraubenarten unterschiedlicher Länge
- Einzeln fotografiert auf weißem Hintergrund
- **Bounding Box** \Rightarrow Implementierbar!
- Fehlerquellen
 - Schrauben liegen schräg



2156x2115

Unser Beispiel

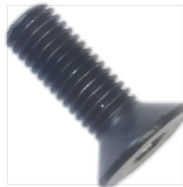
- Zwei Schraubenarten unterschiedlicher Länge
- Einzeln fotografiert auf weißem Hintergrund
- **Bounding Box** \Rightarrow Implementierbar!
- Fehlerquellen
 - Schrauben liegen schräg
 - Schatten



2156x2115

Unser Beispiel

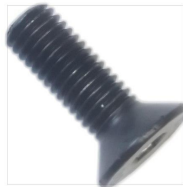
- Zwei Schraubenarten unterschiedlicher Länge
- Einzeln fotografiert auf weißem Hintergrund
- **Bounding Box** \Rightarrow Implementierbar!
- Fehlerquellen
 - Schrauben liegen schräg
 - Schatten
 - Diskretisierung in Pixeln



2156x2115

Unser Beispiel

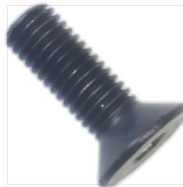
- Zwei Schraubenarten unterschiedlicher Länge
- Einzeln fotografiert auf weißem Hintergrund
- **Bounding Box** \Rightarrow Implementierbar!
- Fehlerquellen
 - Schrauben liegen schräg
 - Schatten
 - Diskretisierung in Pixeln
 - Eingeschränkte Abbildung der Wirklichkeit durch Daten



2156x2115

Unser Beispiel

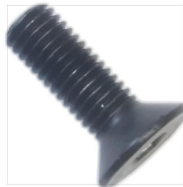
- Zwei Schraubenarten unterschiedlicher Länge
- Einzeln fotografiert auf weißem Hintergrund
- **Bounding Box** \Rightarrow Implementierbar!
- Fehlerquellen
 - Schrauben liegen schräg
 - Schatten
 - Diskretisierung in Pixeln
 - Eingeschränkte Abbildung der Wirklichkeit durch Daten
- Kantenlängen als Punkt im \mathbb{R}^2



2156x2115

Unser Beispiel

- Zwei Schraubenarten unterschiedlicher Länge
- Einzeln fotografiert auf weißem Hintergrund
- **Bounding Box** \Rightarrow Implementierbar!
- Fehlerquellen
 - Schrauben liegen schräg
 - Schatten
 - Diskretisierung in Pixeln
 - Eingeschränkte Abbildung der Wirklichkeit durch Daten
- Kantenlängen als Punkt im \mathbb{R}^2
 - Symmetrie: Maximum als x -Koordinate



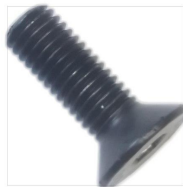
2156x2115

Unser Beispiel

- Zwei Schraubenarten unterschiedlicher Länge
- Einzel fotografiert auf weißem Hintergrund
- **Bounding Box** \Rightarrow Implementierbar!
- Fehlerquellen
 - Schrauben liegen schräg
 - Schatten
 - Diskretisierung in Pixeln
 - Eingeschränkte Abbildung der Wirklichkeit durch Daten
- Kantenlängen als Punkt im \mathbb{R}^2
 - Symmetrie: Maximum als x -Koordinate

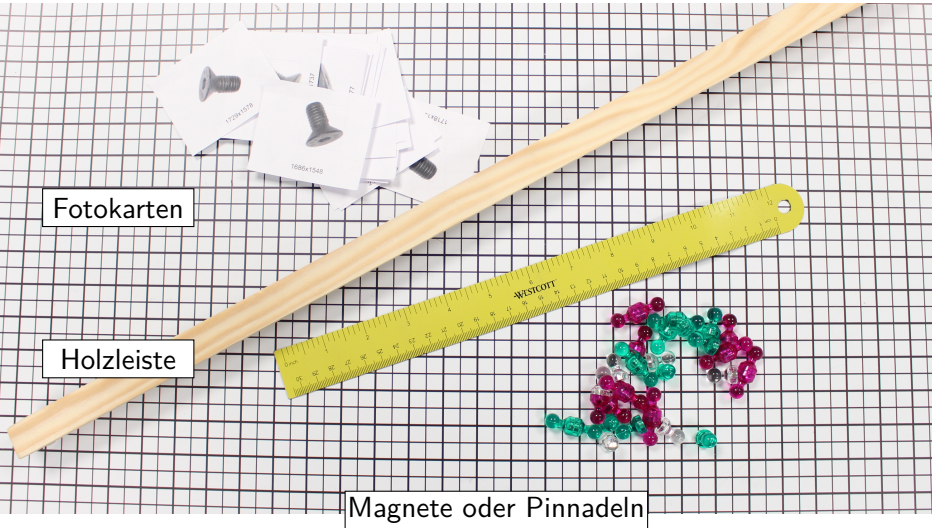
Ziel

Konstruiere einen **Linearen Klassifizierer**, d.h. finde eine **Gerade**, die die beiden Schraubenarten separiert.

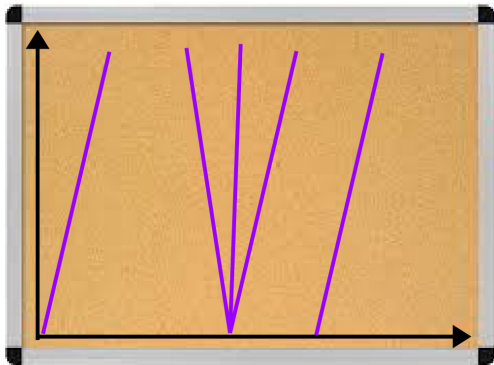


2156x2115

Material

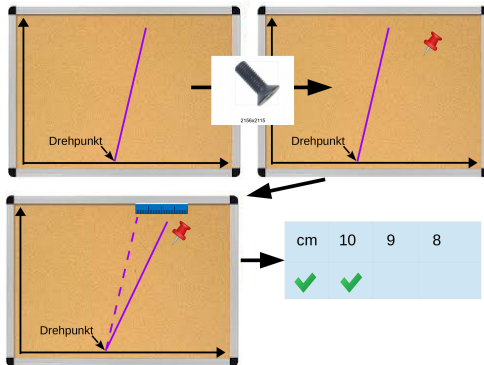


Festlegung der initialen Geraden



- Analyse von Beispieldaten
- Festlegung eines Fixpunktes der Geraden
- Festlegung der Steigung der Geraden

Der Lernprozess



- Analyse von Beispieldaten
- Festlegung eines Fixpunktes der Geraden
- Festlegung der Steigung der Geraden

Wichtiger Aspekt

Reduktion der Verschiebung im Laufe des Verfahrens. Dadurch wird die **Konvergenz** gesichert. (Simulated Annealing)

Der Lernprozess

Lernverfahren

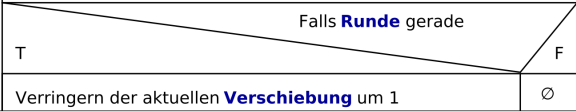
Wählen einer initialen Lage für Holzleiste: Drehpunkt (auf x-Achse) und Steigung

Aktuelle **Verschiebung** = 10 cm

for **Runde** ← 1 to 20

Einfügen eines Datenpunkts

Verschieben des dem Drehpunkt gegenüber liegenden Endes der Holzleiste um aktuelle **Verschiebung** in Richtung des eingefügten Punktes



Zeit für die **Praktische Durchführung**

Anschlussdiskussion

Anschlussdiskussion

- Welche Rolle spielt der Drehpunkt und die anfängliche Steigung?

Anschlussdiskussion

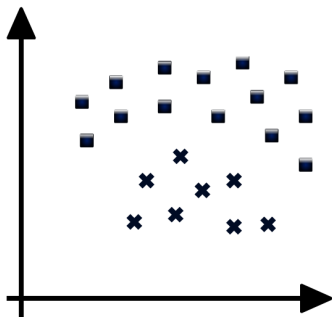
- Welche Rolle spielt der Drehpunkt und die anfängliche Steigung?
- Wozu die Verringerung?

Anschlussdiskussion

- Welche Rolle spielt der Drehpunkt und die anfängliche Steigung?
- Wozu die Verringerung?
- Machen mehr Daten Sinn?

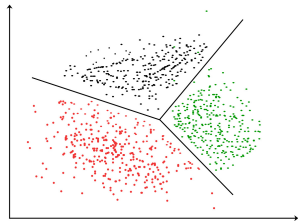
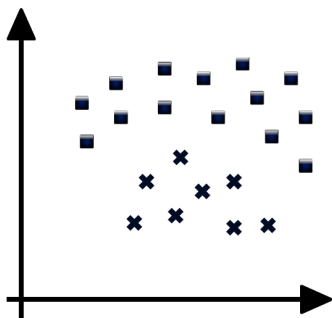
Anschlussdiskussion

- Welche Rolle spielt der Drehpunkt und die anfängliche Steigung?
- Wozu die Verringerung?
- Machen mehr Daten Sinn?
- Andere Form der Cluster?



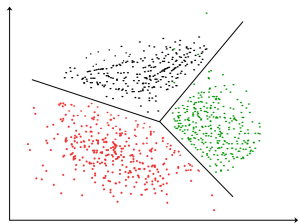
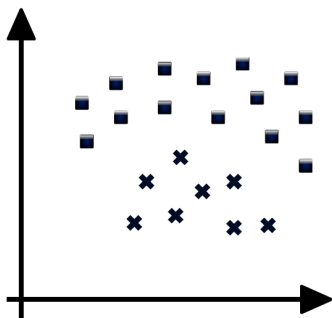
Anschlussdiskussion

- Welche Rolle spielt der Drehpunkt und die anfängliche Steigung?
- Wozu die Verringerung?
- Machen mehr Daten Sinn?
- Andere Form der Cluster?
- Mehr Cluster?



Anschlussdiskussion

- Welche Rolle spielt der Drehpunkt und die anfängliche Steigung?
- Wozu die Verringerung?
- Machen mehr Daten Sinn?
- Andere Form der Cluster?
- Mehr Cluster?
- Geringere Längenunterschiede?

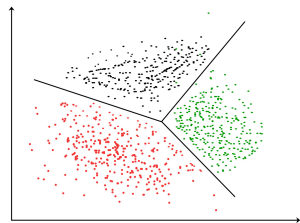
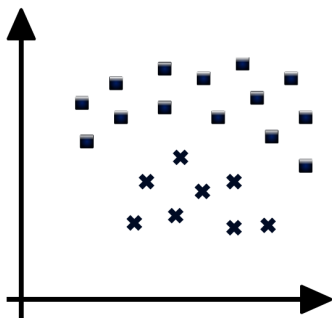


Anschlussdiskussion

- Welche Rolle spielt der Drehpunkt und die anfängliche Steigung?
- Wozu die Verringerung?
- Machen mehr Daten Sinn?
- Andere Form der Cluster?
- Mehr Cluster?
- Geringere Längenunterschiede?

Beobachtung

Das Experiment ermöglicht eine reflektierte Auseinandersetzung mit spezifischen Problemen des Verfahrens aber auch mit grundsätzlichen Problemen des Maschinellen Lernens.



Die theoretische Einordnung

Das Problem

Daten

Es sind Datensätze gegeben, die als Punkte im \mathbb{R}^n interpretiert werden können:

$$\mathbf{x}_i = (x_i^1, \dots, x_i^n)$$

Jeder **Datenpunkt** ist einer von zwei Klassen $y_i \in \{0, 1\}$ zugeordnet.

Das Problem

Daten

Es sind Datensätze gegeben, die als Punkte im \mathbb{R}^n interpretiert werden können:

$$\mathbf{x}_i = (x_i^1, \dots, x_i^n)$$

Jeder **Datenpunkt** ist einer von zwei Klassen $y_i \in \{0, 1\}$ zugeordnet.

Ziel

Konstruktion einer Möglichkeit die Punkte der beiden Klassen voneinander zu unterscheiden, bzw. zu **trennen**.

Das Problem

Daten

Es sind Datensätze gegeben, die als Punkte im \mathbb{R}^n interpretiert werden können:

$$\mathbf{x}_i = (x_i^1, \dots, x_i^n)$$

Jeder **Datenpunkt** ist einer von zwei Klassen $y_i \in \{0, 1\}$ zugeordnet.

Ziel

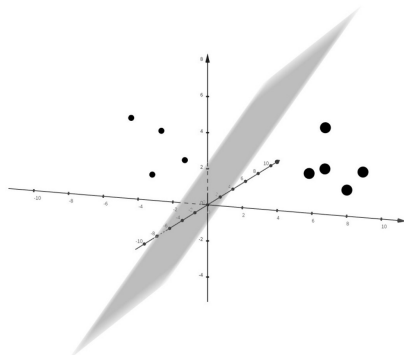
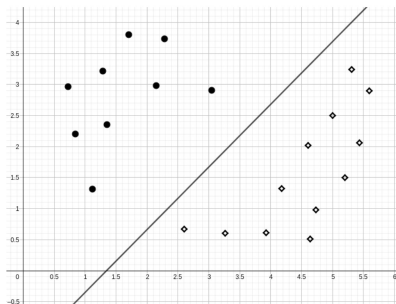
Konstruktion einer Möglichkeit die Punkte der beiden Klassen voneinander zu unterscheiden, bzw. zu **trennen**.

Konkret

Entscheide für zusätzliche Datenpunkte $\mathbf{x} \in \mathbb{R}^n$, zu welcher der beiden Klassen sie gehören.

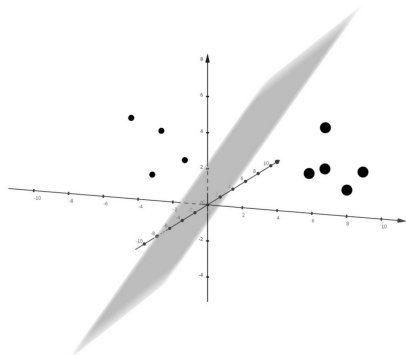
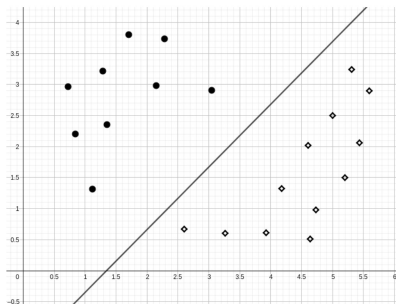
Ansatz

Finde eine $(n - 1)$ -dimensionale Hyperebene $E \in \mathbb{R}^n$, die die beiden Klassen voneinander trennt.



Ansatz

Finde eine $(n - 1)$ -dimensionale Hyperebene $E \in \mathbb{R}^n$, die die beiden Klassen voneinander trennt.



Definition: Linear separabel

Klassen von Punkten im \mathbb{R}^n , die man durch eine Hyperebene trennen kann, nennt man **linear separabel**.

Unsupervised Learning

Unsupervised Verfahren

Wird die Klasse der Trainingsdaten beim Lernen **nicht** benötigt, spricht man von **unsupervised Learning**.

Unsupervised Learning

Unsupervised Verfahren

Wird die Klasse der Trainingsdaten beim Lernen **nicht** benötigt, spricht man von **unsupervised Learning**.

Beispiel

Unser Verfahren ist **unsupervised**, da die Verschiebung der Gerade immer in Richtung des neuen Datenpunktes erfolgt, unabhängig von seiner Klasse.

Supervised Learning

Supervised Learning

Wird beim Lernen die tatsächliche Klasse der Trainingsdaten genutzt, spricht man vom **supervised Learning**.

Supervised Learning

Supervised Learning

Wird beim Lernen die tatsächliche Klasse der Trainingsdaten genutzt, spricht man vom **supervised Learning**.

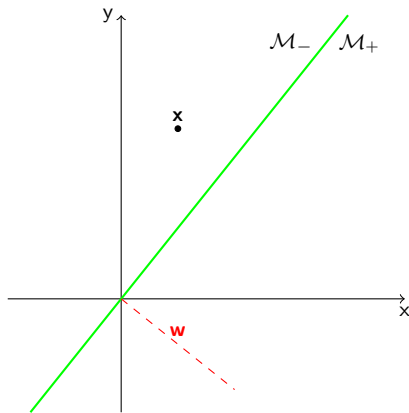
Beispiel

Wird im beschriebenen Verfahren die Gerade nur angepasst, wenn der neue Datenpunkt auf der **falschen** Seite der Gerade liegt, würde es sich um **supervised learning** handeln.

Das Perzeptron

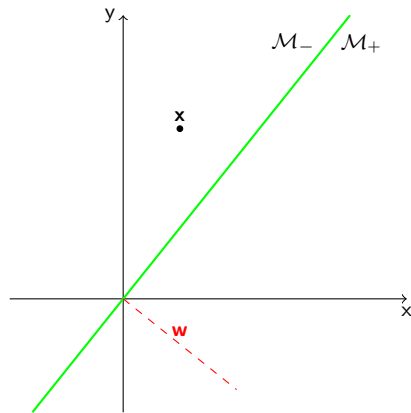
Das Perzeptron - Der historische Hintergrund

- Zwei Klassen \mathcal{M}_+ und \mathcal{M}_-



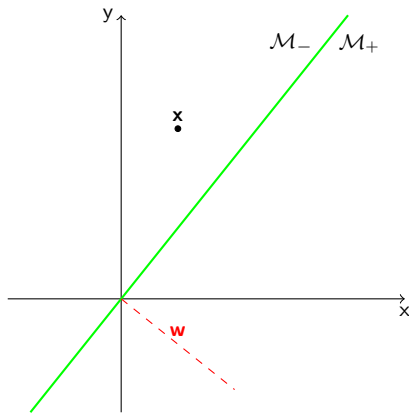
Das Perzeptron - Der historische Hintergrund

- Zwei Klassen \mathcal{M}_+ und \mathcal{M}_-
- **Ursprungsgerade** ist Klassifizierer



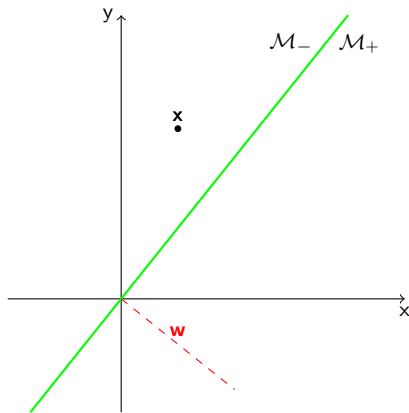
Das Perzeptron - Der historische Hintergrund

- Zwei Klassen \mathcal{M}_+ und \mathcal{M}_-
- **Ursprungsgerade** ist Klassifizierer
- beschrieben durch **Normale**
 $\mathbf{w} = (w_1, \dots, w_n)$



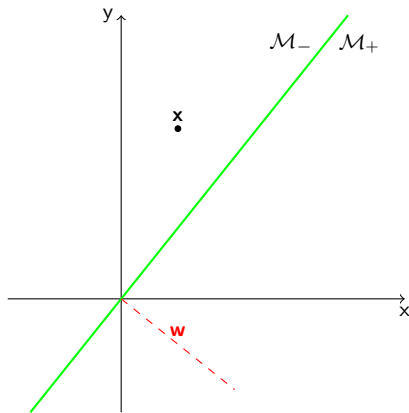
Das Perzeptron - Der historische Hintergrund

- Zwei Klassen \mathcal{M}_+ und \mathcal{M}_-
- **Ursprungsgerade** ist Klassifizierer
- beschrieben durch **Normale**
 $\mathbf{w} = (w_1, \dots, w_n)$
- Klasse von \mathbf{x} ist Vorzeichen des **Skalarproduktes** $\mathbf{x} \cdot \mathbf{w}$



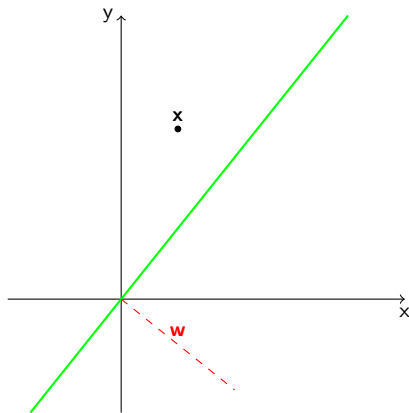
Das Perzeptron - Der historische Hintergrund

- Zwei Klassen \mathcal{M}_+ und \mathcal{M}_-
- **Ursprungsgerade** ist Klassifizierer
- beschrieben durch **Normale**
 $\mathbf{w} = (w_1, \dots, w_n)$
- Klasse von \mathbf{x} ist Vorzeichen des **Skalarproduktes** $\mathbf{x} \cdot \mathbf{w}$
- Es soll gelten:
 - $\mathbf{x} \cdot \mathbf{w} > 0$ für $\mathbf{x} \in \mathcal{M}_+$
 - $\mathbf{x} \cdot \mathbf{w} < 0$ für $\mathbf{x} \in \mathcal{M}_-$



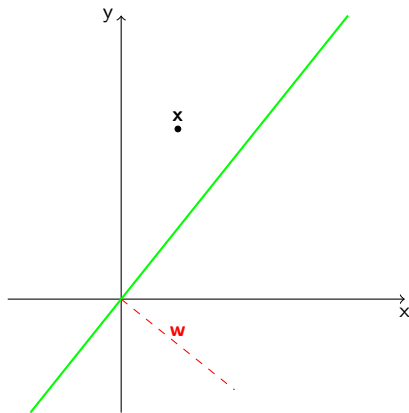
Das Perzeptron - Der Lernschritt

- Alle Trainingspunkte werden durchlaufen



Das Perzeptron - Der Lernschritt

- Alle Trainingspunkte werden durchlaufen
- Wenn aktuelle Klassifizierung **falsch** ist, erfolgt **Korrektur**

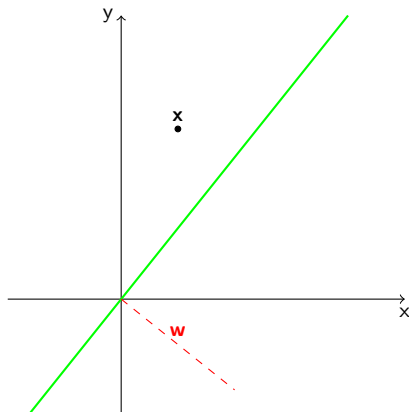


Das Perzeptron - Der Lernschritt

- Alle Trainingspunkte werden durchlaufen
- Wenn aktuelle Klassifizierung **falsch** ist, erfolgt **Korrektur**
- Falls $\mathbf{x} \in \mathcal{M}_+$, dann

$$\mathbf{w} = \mathbf{w} + \alpha \cdot \mathbf{x}$$

mit **Lernrate** $\alpha > 0$.



Das Perzeptron - Der Lernschritt

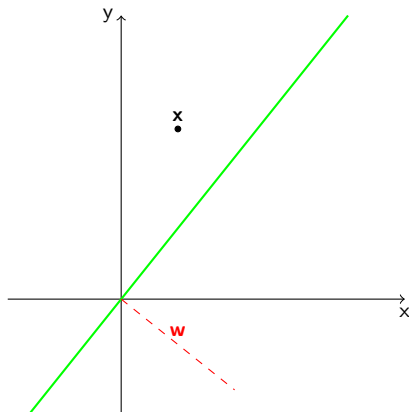
- Alle Trainingspunkte werden durchlaufen
- Wenn aktuelle Klassifizierung **falsch** ist, erfolgt **Korrektur**
- Falls $\mathbf{x} \in \mathcal{M}_+$, dann

$$\mathbf{w} = \mathbf{w} + \alpha \cdot \mathbf{x}$$

mit **Lernrate** $\alpha > 0$.

- Falls $\mathbf{x} \in \mathcal{M}_-$, dann

$$\mathbf{w} = \mathbf{w} - \alpha \cdot \mathbf{x}$$



Das Perzeptron - Der Lernschritt

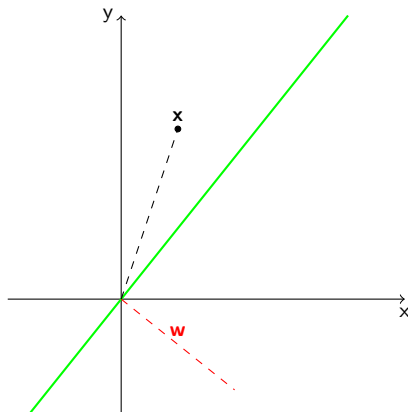
- Alle Trainingspunkte werden durchlaufen
- Wenn aktuelle Klassifizierung **falsch** ist, erfolgt **Korrektur**
- Falls $\mathbf{x} \in \mathcal{M}_+$, dann

$$\mathbf{w} = \mathbf{w} + \alpha \cdot \mathbf{x}$$

mit **Lernrate** $\alpha > 0$.

- Falls $\mathbf{x} \in \mathcal{M}_-$, dann

$$\mathbf{w} = \mathbf{w} - \alpha \cdot \mathbf{x}$$



Das Perzeptron - Der Lernschritt

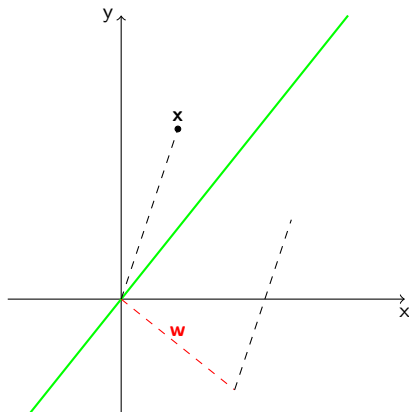
- Alle Trainingspunkte werden durchlaufen
- Wenn aktuelle Klassifizierung **falsch** ist, erfolgt **Korrektur**
- Falls $\mathbf{x} \in \mathcal{M}_+$, dann

$$\mathbf{w} = \mathbf{w} + \alpha \cdot \mathbf{x}$$

mit **Lernrate** $\alpha > 0$.

- Falls $\mathbf{x} \in \mathcal{M}_-$, dann

$$\mathbf{w} = \mathbf{w} - \alpha \cdot \mathbf{x}$$



Das Perzeptron - Der Lernschritt

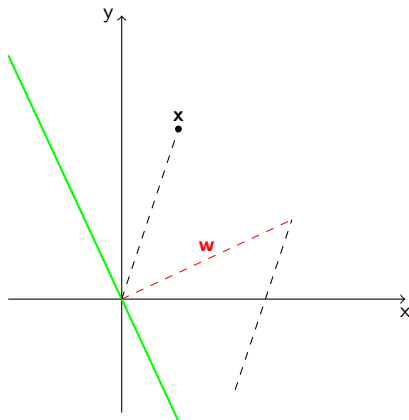
- Alle Trainingspunkte werden durchlaufen
- Wenn aktuelle Klassifizierung **falsch** ist, erfolgt **Korrektur**
- Falls $\mathbf{x} \in \mathcal{M}_+$, dann

$$\mathbf{w} = \mathbf{w} + \alpha \cdot \mathbf{x}$$

mit **Lernrate** $\alpha > 0$.

- Falls $\mathbf{x} \in \mathcal{M}_-$, dann

$$\mathbf{w} = \mathbf{w} - \alpha \cdot \mathbf{x}$$



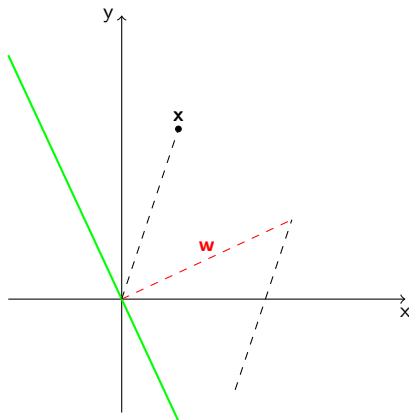
Das Perzeptron - Der Lernschritt

- Alle Trainingspunkte werden durchlaufen
- Wenn aktuelle Klassifizierung **falsch** ist, erfolgt **Korrektur**
- Falls $\mathbf{x} \in \mathcal{M}_+$, dann

$$\mathbf{w} = \mathbf{w} + \alpha \cdot \mathbf{x}$$

mit **Lernrate** $\alpha > 0$.

- Falls $\mathbf{x} \in \mathcal{M}_-$, dann
- $$\mathbf{w} = \mathbf{w} - \alpha \cdot \mathbf{x}$$
- Wiederhole das, bis der Fehler klein genug ist.



Das Perzeptron - Der Lernschritt

- Alle Trainingspunkte werden durchlaufen
- Wenn aktuelle Klassifizierung **falsch** ist, erfolgt **Korrektur**
- Falls $\mathbf{x} \in \mathcal{M}_+$, dann

$$\mathbf{w} = \mathbf{w} + \alpha \cdot \mathbf{x}$$

mit **Lernrate** $\alpha > 0$.

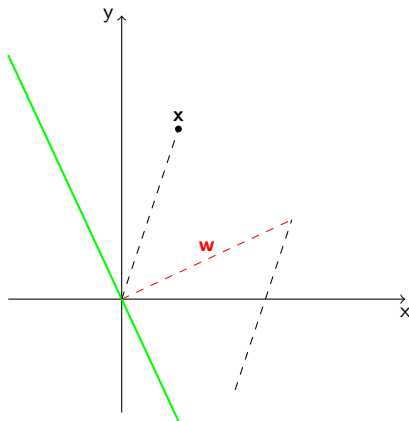
- Falls $\mathbf{x} \in \mathcal{M}_-$, dann

$$\mathbf{w} = \mathbf{w} - \alpha \cdot \mathbf{x}$$

- Wiederhole das, bis der Fehler klein genug ist.

Satz

Sind \mathcal{M}_+ und \mathcal{M}_- linear separabel, dann terminiert das Verfahren.



Das Perzeptron

Supervised Learning

Beim Perzeptron handelt es sich um **Supervised Learning**.

Das Perzeptron

Supervised Learning

Beim Perzeptron handelt es sich um **Supervised Learning**.

Problem

Das Perzeptron versagt bei nicht linear seperablen Klassen.

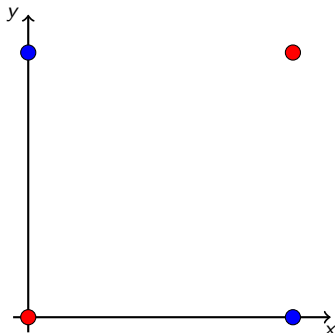
Das Perzeptron

Supervised Learning

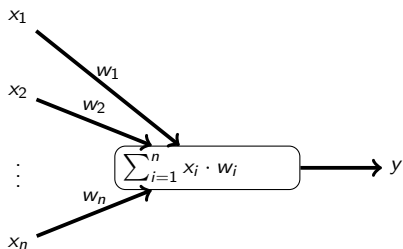
Beim Perzeptron handelt es sich um **Supervised Learning**.

Problem

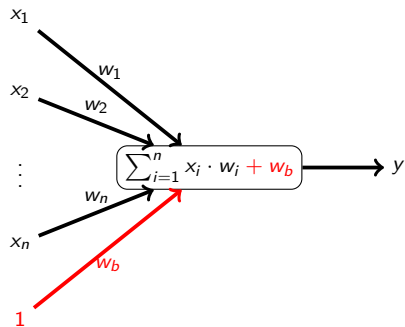
Das Perzeptron versagt bei nicht linear seperablen Klassen.



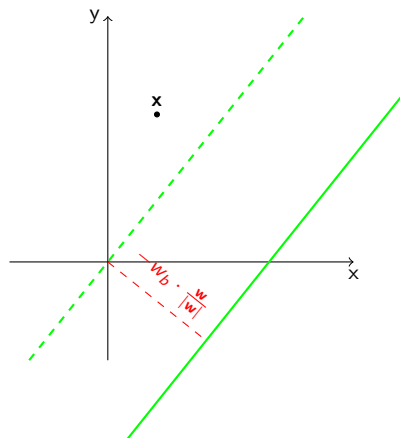
Eine andere Darstellung



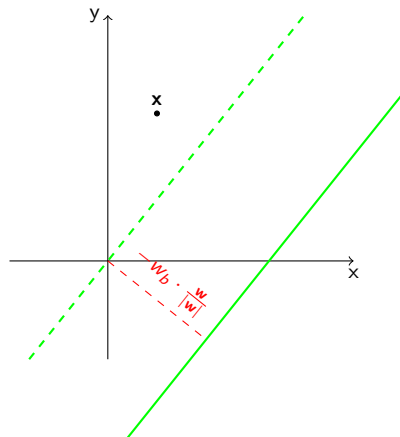
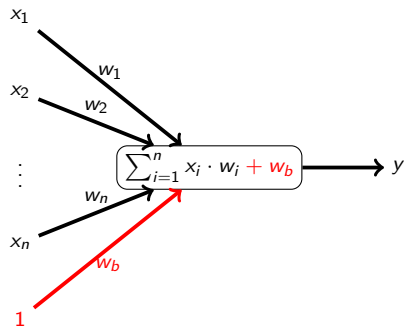
Eine andere Darstellung



- Bias w_b ergänzt zur Verschiebung



Eine andere Darstellung

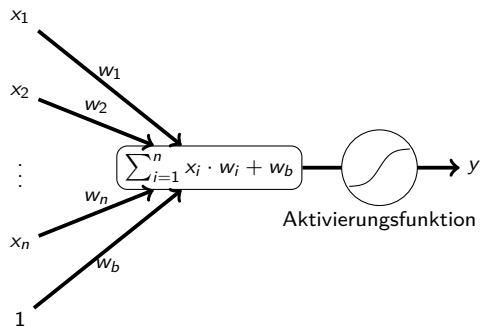


- Bias w_b ergänzt zur Verschiebung

Perspektive

Das entspricht bereits der Idee eines **Neurons**.

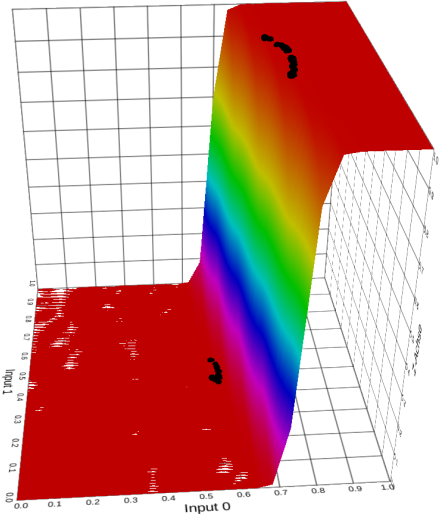
Das Neuron

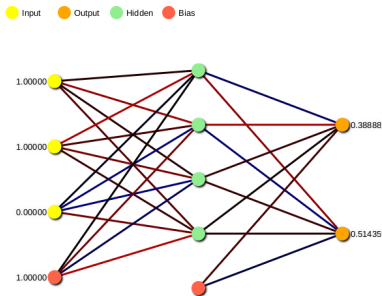


- **Aktivierungsfunktion** berechnet das Ausgangssignal y

- z.B. Sigmoidfunktion
$$\sigma(t) = \frac{1}{1+e^{-t}}$$

Die Schrauben und ein Neuronales Netz





- Kombination von vielen Neuronen in mehreren Schichten
- Ausgangssignale sind Eingangssignale der nächsten Schicht
- Beim Lernen wird der Fehler durch ein **Gradientenverfahren rückwärts** auf die Gewichte übertragen ⇒ **Backpropagation**
- Durch die Kombination vieler Neuronen in mehreren Schichten lassen sich sehr komplexe Funktionen lernen.
- **Deep Learning**

Ein grundsätzliches Problem

Ein grundsätzliches Problem

Die Theorie



Die Wirklichkeit



Die Daten



Ein grundsätzliches Problem

Die Theorie



Die Wirklichkeit



Die Daten



Daten ...

- sind immer fehlerhaft.
- widersprechen sich.
- bilden die Wirklichkeit nur unvollständig ab.

Ein grundsätzliches Problem

Die Theorie



Die Wirklichkeit



Die Daten



Daten ...

- sind immer fehlerhaft.
- widersprechen sich.
- bilden die Wirklichkeit nur unvollständig ab.

Konsequenz

Maschinelles Lernen (und andere datengetriebene Systeme) enthalten **immer** Fehler, die bei ihrer Anwendung berücksichtigt werden müssen.



Michael Brinkmeier
mbrinkmeier@uni-osnabrueck.de

Andrewas Hüwel
ahuwel@uni-osnabrueck.de



<https://abbozza.uni-osnabrueck.de>