

Training

Den Bereich Training hatten wir bis jetzt ausgespart.

Das Netzwerk wird für das Training mit Beispieldaten gefüttert.

```
["f", 60, 165] # Alice  
["m", 73, 183] # Bob  
["m", 69, 178] # Charlie  
["f", 54, 152] # Diana
```

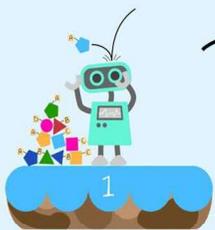
Die Daten sind jeweils mit dem Zusatz "m" (male) oder "f" (female) versehen.

Man spricht hier von überwachtem Lernen (supervised).

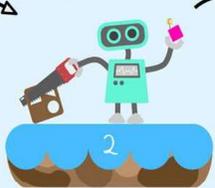
Insgesamt unterscheidet man:

Überwachtes Lernen

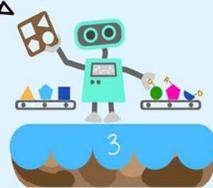
Beschriftete Eingaben erhalten



Regeln finden, die bekannte Eingaben richtig beschriften

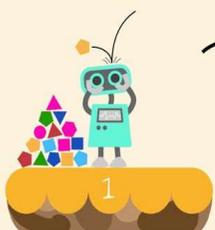


Neue Eingaben entsprechend der gefundenen Regeln beschriften



Unüberwachtes Lernen

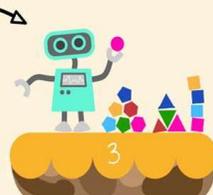
Unbeschriftete Eingaben erhalten



Ähnlichkeiten in den Eingaben erkennen und Muster finden



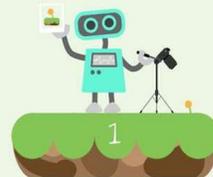
In der Eingabe Gruppen und Ausreißer identifizieren



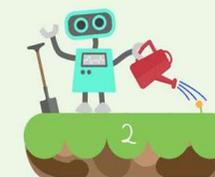
Paradigmen des maschinellen Lernens

Verstärkendes Lernen

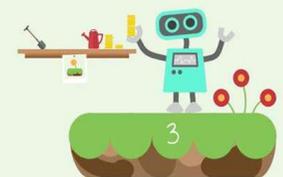
Zustand erfassen



Aktion wählen und durchführen



Belohnung oder Bestrafung erhalten



Strategie anpassen



Stefan Seegerer
Tilman Michaeli
Sven Jatzlau



Durchführung

Die technische Durchführung, wollen wir hier nur ganz kurz und nicht im Detail erklären:

Man beginnt mit zufällig gewählten Werten für die Gewichte und Biases.

Jeder solchen Kombination kann ein Fehlerwert zugeordnet werden.

Berechnung Fehlerwert

Oft nimmt man dafür die Mittlere quadratische Abweichung (MSE - Mean Squared Error).

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

MSE = mean squared error

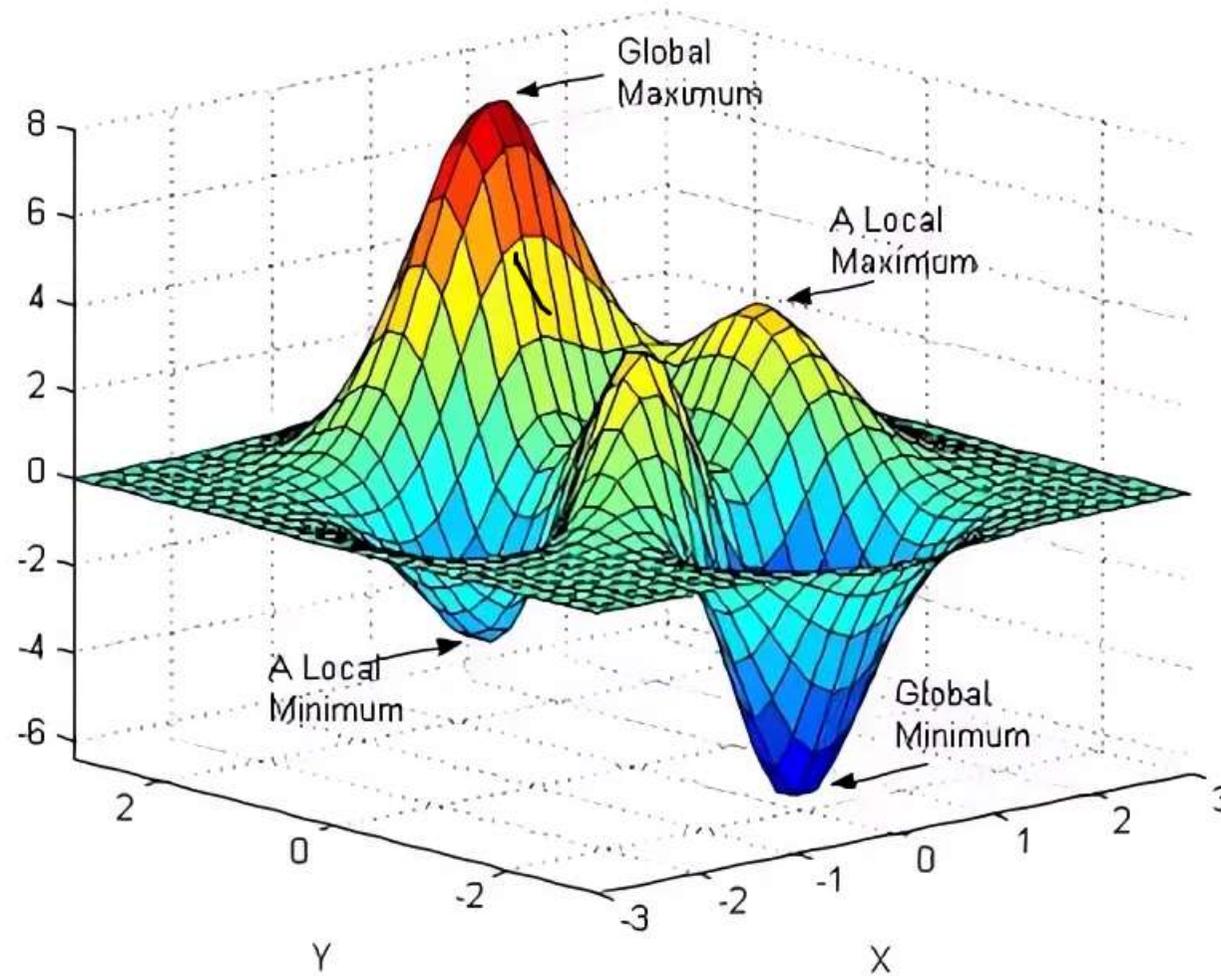
n = number of data points

Y_i = observed values

\hat{Y}_i = predicted values

n Beobachtungen wurden gemacht. Es wird jeweils die Differenz zwischen Erwartungswert (" f "=1, " m "=0) und beobachtetem Wert quadriert. Dann werden alle Werte summiert und die Summe durch n dividiert.

A Non-Convex Combination of Gaussian Distributions



Zachary kaplan, CC BY-SA 4.0, via Wikimedia Commons

Die Parameter spannen zusammen mit dem Fehlerwert eine n -dimensionale Ebene auf. Zur Veranschaulichung hier der Fall für nur 2 Parameter. Die Suche nach einer Lösung entspricht hier der Suche nach (lokalen oder globalen) Minima in dieser Ebene.

Gradient Descent

Sehr häufig, aber nicht ausschließlich, wird hier das sogenannte "Gradient Descent-Verfahren" verwendet. Es ermittelt die Steigung am momentanen Punkt (Gradient) und versucht bergab in Richtung des Minimums zu gehen.

Das Netzwerk wurde auf diese Art trainiert und erfolgreich getestet.

Die so entstandenen Gewichte und Biases können dann ausgelesen werden.

Versuche für die angegebenen 2 Wertepaare selbst die Berechnung durchzuführen.

Handelt es sich um einen Mann oder eine Frau? Du kannst auch selbst einen Wert ausprobieren.

Anmerkungen

Vor der Berechnung werden vom Wertepaar die durchschnittlichen Werte für Gewicht und Größe abgezogen. Auf diese Art wurde das Netzwerk trainiert.

Als Beschränkungsfunktion kommt die Sigmoidfunktion zum Einsatz, siehe die Datei `sigmoid.pdf` unten.