



Überleitung: Vom klassischen Ansatz zum maschinellen Lernen

In der klassischen linearen Regression haben wir die optimalen Parameter (m) (Steigung) und (b) (Achsenabschnitt) mit einer direkten mathematischen Formel berechnet. Das funktioniert gut bei einfachen Modellen, wird aber bei komplexeren Aufgaben unpraktisch oder unmöglich.

Maschinelles Lernen verwendet stattdessen iterative Methoden wie den **Gradientenabstieg**, um die Parameter schrittweise zu optimieren. Damit wird der Prozess „lernend“ und anpassbar.

Die vier Komponenten des maschinellen Lernens

1. Daten:

Unsere Daten bestehen aus Eingabewerten (x) und Zielwerten (y). Diese stellen die Grundlage des Lernprozesses dar.

2. Modell:

Unser Modell ist eine Gleichung der Form:

$$\hat{y} = mx + b$$

Ziel des Modells ist es, die Datenpunkte möglichst gut zu beschreiben, d. h., die Abweichung zwischen den Vorhersagen (\hat{y}) und den tatsächlichen Werten (y) zu minimieren.

3. Zielfunktion:

Die Zielfunktion bewertet, wie gut unser Modell ist. Im Fall der linearen Regression verwenden wir die mittlere quadratische Abweichung (Summe der kleinsten Quadrate):

$$J(m, b) = \frac{1}{N} \sum_{i=1}^N (y_i - (mx_i + b))^2$$

Ziel: Die Zielfunktion so klein wie möglich machen, d. h., die Fehler minimieren.

4. Optimierung:

Der **Gradientenabstieg** ist das Verfahren, mit dem wir m und b so anpassen, dass $J(m, b)$ minimiert wird.

Gradientenabstieg: Eine einfache Erklärung

Grundidee:

Stell dir vor, du stehst auf einem Berg und möchtest ins Tal gelangen. Du kannst nicht sehen, wo das Tal ist, aber du weißt, dass der schnellste Weg nach unten der ist, in dem der Boden am steilsten abfällt. Bei jedem Schritt überprüfst du, in welche Richtung der Hang am stärksten abfällt, und gehst ein Stück weiter.

Das ist genau das, was der Gradientenabstieg tut:

- Er berechnet den steilsten Abstieg der Zielfunktion (den „Gradienten“) und macht einen kleinen Schritt in diese Richtung.
- Mit jedem Schritt nähert er sich dem Minimum der Zielfunktion.

Schritt-für-Schritt-Erklärung des Gradientenabstiegs

1. Startpunkt wählen:

- Wir beginnen mit zufälligen Werten für m und b (z. B. $m = 0$ und $b = 0$).

2. Fehler berechnen:

- Mit dem Modell $\hat{y} = m \times x + b$ berechnen wir die Vorhersagen und den Fehler (die Abweichung zwischen \hat{y} und y).

3. Gradient berechnen:

- Der Gradient zeigt an, wie die Zielfunktion $J(m, b)$ sich ändert, wenn wir m oder b leicht anpassen.
- Wir berechnen die Ableitungen der Zielfunktion:

$$\frac{\partial J}{\partial m} = -\frac{2}{N} \sum_{i=1}^N x_i (y_i - (mx_i + b))$$

$$\frac{\partial J}{\partial b} = -\frac{2}{N} \sum_{i=1}^N (y_i - (mx_i + b))$$

- $\frac{\partial J}{\partial m}$: Wie stark sich $J(m, b)$ ändert, wenn wir m ändern.
- $\frac{\partial J}{\partial b}$: Wie stark sich $J(m, b)$ ändert, wenn wir b ändern.

4. Parameter anpassen:

- Wir passen m und b in die Richtung des steilsten Abstiegs an. Dabei machen wir kleine Schritte, deren Größe durch die **Lernrate** α bestimmt wird:

$$m_{\text{neu}} = m_{\text{alt}} - \alpha \cdot \frac{\partial J}{\partial m}$$

$$b_{\text{neu}} = b_{\text{alt}} - \alpha \cdot \frac{\partial J}{\partial b}$$

5. Wiederholen:

- Die Schritte 2–4 werden wiederholt, bis sich m und b kaum noch ändern oder der Fehler $J(m, b)$ ein Minimum erreicht.

Intuition: Warum funktioniert das?

- Der Gradient ist wie eine „Landkarte“, die uns zeigt, in welche Richtung wir gehen müssen, um den tiefsten Punkt (Minimum) zu erreichen.
- Indem wir m und b iterativ anpassen, nähern wir uns Schritt für Schritt den optimalen Werten, die die Zielfunktion minimieren.

Eigenschaften des Gradientenabstiegs

1. Lernrate α :

- Wenn die Lernrate zu groß ist, machen wir riesige Schritte und können das Minimum „verpassen“.
- Wenn die Lernrate zu klein ist, dauern die Schritte sehr lange.

2. Lokales Minimum:

- Der Gradientenabstieg garantiert, dass wir ein Minimum erreichen – aber ob es das globale Minimum ist, hängt von der Form der Zielfunktion ab. Bei der linearen Regression ist die Zielfunktion jedoch immer konvex, d. h., es gibt nur ein Minimum.

3. Iterativ statt analytisch:

- Gradientenabstieg ist ein iteratives Verfahren, das besonders nützlich ist, wenn die Zielfunktion zu kompliziert für eine direkte Lösung ist.

Beispiel: Gradientenabstieg bei der linearen Regression

Ziel:

Wir möchten eine Linie finden, die die Punkte möglichst gut beschreibt. Dazu verwenden wir Gradientenabstieg, um die Parameter (m) und (b) so zu optimieren, dass der Fehler minimiert wird.

1. Start:

- Wir beginnen mit ($m = 0$) und ($b = 0$).

2. Fehlerberechnung:

- Für jeden Punkt berechnen wir die Abweichung zwischen der vorhergesagten Linie und dem tatsächlichen Wert.

3. Gradientenberechnung:

- Wir berechnen, wie wir (m) und (b) anpassen müssen, um den Fehler zu reduzieren.

4. Parameteranpassung:

- Wir bewegen (m) und (b) in Richtung des steilsten Abstiegs.

5. Wiederholung:

- Dieser Prozess wird so lange wiederholt, bis der Fehler ein Minimum erreicht.

