ML Begriffe Regression

Die absolute Baseline «Constant»	Bevor man komplizierte Modelle baut, startet man mit einer ganz einfachen „Baseline“. Diese sagt **immer denselben Wert** vorher – egal, was die Eingaben sind. Man nimmt den **Durchschnitt** (arithmetisches Mittel) oder **Median** der Zielwerte aus den Trainingsdaten → und benutzt diesen Wert als Vorhersage **für alle neuen Datenpunkte.**
Parametrische Methoden	- Man nimmt an, dass es eine bestimmte Formel gibt, mit der man den Zusammenhang zwischen den Eingaben (Features) und dem Zielwert beschreiben kann. Meist ist das eine einfache Gleichung wie eine lineare Funktion - Ziel: Man will herausfinden, welche Werte die Gewichte w haben müssen, damit die Gleichung möglichst gut die echten Zielwerte vorhersagt
Einfache Lineare Regression	- Modell mit nur einem Merkmal z.B. Hausgrösse - Modell lernt eine Gerade, die den Zusammenhang zwischen diesem Merkmal und dem Preis beschreibt - Zur Bewertung nutzt man die Fehlersumme Sum of Square Errors
Sum of Square Errors	- SSE misst, **wie falsch** ein Vorhersagemodell ist – also wie stark die Vorhersagen vom echten Wert abweichen. - SSE so klein wie möglich, bedeutet ist nah an den echten Werten - Formel: $SSE = \sum_i (\underbrace{Preis_i}_{\text{Echter Preis (blaue Punkte)}} - \underbrace{[w_0 + w_1 \cdot sq.ft._i]}_{\text{Durch Fkt. geschätzter Preis (Punkte auf der Linie)}})^2$

Gradient Descent	- Verfahren, um die besten Werte für die Gewichte w_0 und w_1 zu finden, sodass der Fehler (SSE) möglichst klein wird.
	- Schrittweise Optimierung der Gewichte
	- Wie funktioniert das?
	    o Starte mit irgendeinem Wert für w_0 und w_1
	    o Berechne die Steigung des Fehlers
	    o Gehe in die Richtung, in der der Fehler kleiner wird; wenn Steigung positiv→ w_0 verringern; wenn Steigung negativ → w_0 vergrössern

Multiple lineare Regression	- Mehr Attribute x_i d.h. mehr Gewichte w_i zum lernen - Formel-Beispiel: Preis = w_0 + w_1 * Grösse + w_2 * Zimmeranzahl - Jedes Merkmal bekommt ein eigenes Gewicht

<table>
<tr><td></td><td>- Probleme:
o Manchmal sind Zusammenhänge nicht linear
o Manchmal spielen Interaktionen zwischen Attributen eine Rolle (Bspw. Eiscreme)</td></tr>
<tr><td>Multiple lineare Regression – Interpretation</td><td>- Der Wert bedeutet, dass pro zusätzlichem Schlafzimmer, der geschätzte Hauspreis im Durchschnitt um 4708.79 Einheiten sinkt, wenn alle anderen Faktoren gleich bleiben
- Warum?: Korrelation mit anderen Variablen
<table>
<tr><th></th><th>name</th><th>coef</th><th></th></tr>
<tr><td>1</td><td>intercept</td><td>92544.2</td><td></td></tr>
<tr><td>2</td><td>houseSize</td><td>34.5512</td><td></td></tr>
<tr><td>3</td><td>lotSize</td><td>3.43763</td><td></td></tr>
<tr><td>4</td><td>bedrooms</td><td>-4708.79</td><td>→ Erklärung?</td></tr>
<tr><td>5</td><td>granite=0</td><td>3472.01</td><td></td></tr>
<tr><td>6</td><td>granite=1</td><td>-3472.01</td><td></td></tr>
<tr><td>7</td><td>bathroom=0</td><td>-14054.5</td><td></td></tr>
<tr><td>8</td><td>bathroom=1</td><td>14054.5</td><td></td></tr>
</table></td></tr>
<tr><td>Nicht-lineare Zusammenhänge</td><td>- Die Daten folgen nicht einer einfachen geraden Linie
- Beispiel: Eisverkauf steigt mit Temperatur, aber ab 40 Grad sinkt er wieder. Das ist nicht linear, sondern gekrümmt.</td></tr>
<tr><td>Gemischte Terme / Interaktionen</td><td>- Zwei Merkmale wirken zusammen auf das Ergebnis</td></tr>
<tr><td>Nichtparametrische Verfahren</td><td>- Ein nichtparametrisches Verfahren macht keine Annahmen über die Form der zugrunde liegenden Funktion (z. B. ob sie linear, quadratisch etc. ist), Stattdessen „lernt“ es direkt aus den vorhandenen Datenpunkten.
- Vorteil: kann gut sein, wenn die Funktion nicht bekannt ist
- Nachteil: schlecht bei dünn verteilten Daten</td></tr>
<tr><td>MAE – Mean absolut error</td><td>- Misst wie «falsch» eine Vorhersage im Durchschnitt ist.
- Je kleiner der MAE desto besser das Modell</td></tr>
<tr><td>Entscheidungsbäume</td><td>- Ein Modell, das Daten Schritt für Schritt aufteilt
- Vorteile:
o Leicht verständlich
o Kann auch nicht-lineare Zusammenhänge abbilden
o Funktioniert gut mit gemischten Daten</td></tr>
<tr><td>Gradient Boosting</td><td>- Eine Technik, bei der viele schwache Modelle nacheinander trainiert werden- jedes Modell verbessert die Fehler des vorherigen
- Vorteile:
o Sehr genau
o Guter Ansatz für Vorhersageprobleme</td></tr>
<tr><td>Regularisierung</td><td>- Eine Methode um zu verhindern, dass ein Modell zu stark auf die Trainingsdaten passt (Overfitting)</td></tr>
</table>

	- Beim Lernen werden grosse Gewichte bestraft, so wird das Modell einfacher gehalten
	- Zwei Arten:
	    o Ridge(L2): Bestraft grosse Gewichte insgesamt
	    o Lasso(L1): Kann unwichtige Gewichte auf 0 setzten, also Merkmale ignorieren
	- Vorteile:
	    o Macht Modelle robuster
	    o Verhindert, dass das Modell sich zu sehr an Ausreisser klammert