Lineare Regression

Gegeben: n Paare (x_i, y_i) von reellen Zahlen. Gesucht ist eine Gerade, die diese Punkte möglichst gut approximiert:

$$\hat{y} = a \cdot x + b \tag{1}$$

Wir versuchen nun, a und b so zu wählten, dass der Fehler $\hat{y}_i - y_i$ über alle Werte gleichmäßig minimiert wird. Dazu quadrieren wir ihn und addieren alle Werte auf:

$$\begin{aligned}
f(a, b) &= \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 && (2)\\
&= \sum_{i=1}^{n} (ax_i + b - y_i)^2 && (3)\\
&= \sum_{i=1}^{n} (a^2x_i^2 + b^2 + y_i^2 + 2abx_i - 2ax_iy_i - 2by_i) && (4)
\end{aligned}$$

Wir erhalten eine quadratische Funktion bezüglich a und b. Der Fehler wird minimal, wenn die (partiellen) Ableitungen nach a und b Null sind:

$$\begin{aligned}
\frac{\partial f}{\partial a} &= 0 && (5)\\
\frac{\partial f}{\partial b} &= 0 && (6)\\
& && (7)
\end{aligned}$$

also

$$\begin{aligned}
0 = \frac{\partial f}{\partial a} &= \sum_{i=1}^{n} (2ax_i^2 + 2bx_i - 2x_iy_i) && (8)\\
&= 2a\sum_{i=1}^{n} x_i^2 + 2b\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} x_iy_i && (9)\\
0 &= a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_iy_i && (10)\\
0 = \frac{\partial f}{\partial b} &= \sum_{i=1}^{n} (2b + 2ax_i - 2y_i) && (11)\\
&= 2a\sum_{i=1}^{n} x_i + 2bn - 2\sum_{i=1}^{n} y_i && (12)\\
0 &= a\sum_{i=1}^{n} x_i + bn - \sum_{i=1}^{n} y_i && (13)
\end{aligned}$$

Um die Übersicht zu erhöhen, verwenden wir für die benötigten Summen eine abgekürzte Schreibweise:

$$[X] := \sum_{i=1}^{n} x_i \tag{14}$$

$$[Y] := \sum_{i=1}^{n} y_i \tag{15}$$

$$[X^2] := \sum_{i=1}^{n} x_i^2 \tag{16}$$

$$[XY] := \sum_{i=1}^{n} x_i y_i \tag{17}$$

$$\tag{18}$$

Damit erhalten wir:

$$0 = [X^2]a + [X]b - [XY] \tag{19}$$

$$0 = [X]a + nb - [Y] \tag{20}$$

Wir stellen die zweite Gleichung nach b um:

$$b = \frac{[Y] - [X]a}{n} \tag{21}$$

$$= \frac{[Y]}{n} - \frac{[X]}{n}a \tag{22}$$

Nun setzen wir b in erste Gleichung ein:

$$0 = [X^2]a + [X]\frac{[Y]}{n} - [X]\frac{[X]}{n}a - [XY] \tag{23}$$

$$a = \frac{[XY] - [X]\frac{[Y]}{n}}{[X^2] - [X]\frac{[X]}{n}} \tag{24}$$

$$= \frac{n[XY] - [X] \cdot [Y]}{n[X^2] - [X] \cdot [X]} \tag{25}$$