Anwendung der partiellen Differentiation in der Statistik: Regressionsgerade

Gesucht: Eine Gerade, die zu allen Punkten den kleinsten Abstand hat;

d.h. die Differenzen e_i müssen insgesamt möglichst klein sein (Optimierungsproblem)

Ansatz: $$\sum_{i=1}^{n}(e_i)^2 \quad \rightarrow \quad \min.$$

Gerade:

$$y = b_0 + b_1 x \xrightarrow{\text{für jeden Wert auf der Geraden}} y_i = b_0 + b_1 x_i$$

$$\rightarrow \quad \sum_{i=1}^{n}(e_i)^2 = \sum_{i=1}^{n}\left(y_i - y_i\right)^2 = \sum_{i=1}^{n}\left(y_i - b_0 - b_1 x_i\right)^2$$

$$E(b_1;b_0) = \sum_{i=1}^{n}\left(y_i^2 - y_i b_0 - y_i b_1 x_i - y_i b_0 + b_0^2 + b_0 b_1 x_i - y_i b_1 x_i + b_0 b_1 x_i + b_1^2 x_i^2\right)$$

$$E(b_1;b_0) = \sum_{i=1}^{n}\left(y_i^2 - 2y_i b_0 - 2y_i b_1 x_i + b_0^2 + 2b_0 b_1 x_i + b_1^2 x_i^2\right)$$

$$E(b_1;b_0) = \sum_{i=1}^{n} y_i^2 - 2b_0\sum_{i=1}^{n} y_i - 2b_1\sum_{i=1}^{n} x_i y_i + b_0^2\sum_{i=1}^{n} 1 + 2b_0 b_1\sum_{i=1}^{n} x_i + b_1^2\sum_{i=1}^{n} x_i^2$$

$$E(b_1;b_0) = \sum_{i=1}^{n} y_i^2 - 2b_0\sum_{i=1}^{n} y_i - 2b_1\sum_{i=1}^{n} x_i y_i + nb_0^2 + 2b_0 b_1\sum_{i=1}^{n} x_i + b_1^2\sum_{i=1}^{n} x_i^2$$

$$\frac{\partial E(b_1;b_0)}{\partial b_1} = -2\sum_{i=1}^{n} x_i y_i + 2b_0\sum_{i=1}^{n} x_i + 2b_1\sum_{i=1}^{n} x_i^2 = 0$$

$$\frac{\partial E(b_1;b_0)}{\partial b_0} = -2\sum_{i=1}^{n} y_i + 2nb_0 + 2b_1\sum_{i=1}^{n} x_i = 0$$

Lösung der Gleichungen der partiellen Ableitungen:

$$\text{I)}\quad -2\sum_{i=1}^{n} x_i y_i + 2b_0\sum_{i=1}^{n} x_i + 2b_1\sum_{i=1}^{n} x_i^{\,2} \;=\; 0$$

$$\text{II)}\quad -2\sum_{i=1}^{n} y_i + 2nb_0 + 2b_1\sum_{i=1}^{n} x_i \;=\; 0$$

Es gilt:

$$\sum_{i=1}^{n} x_i \;=\; n\cdot\overline{x} \qquad und \qquad \sum_{i=1}^{n} y_i \;=\; n\cdot\overline{y} \quad \text{(arithmetisches Mittel)}$$

eingesetzt :

$$\text{I)}\quad -2\sum_{i=1}^{n} x_i y_i + 2b_0 n\cdot\overline{x} + 2b_1\sum_{i=1}^{n} x_i^{\,2} \;=\; 0$$

$$\text{II)}\quad -2n\cdot\overline{y} + 2nb_0 + 2b_1 n\cdot\overline{x} \;=\; 0 \xrightarrow{:(2n)} -\overline{y} + b_0 + b_1\overline{x} \;=\; 0$$

$$\rightarrow\; b_0 \;=\; \overline{y} - b_1\overline{x}$$

$$\xrightarrow{\text{II) } eingesetzt\ in \text{ I)}} \text{I)}\quad -2\sum_{i=1}^{n} x_i y_i + 2\left(\overline{y} - b_1\overline{x}\right)n\cdot\overline{x} + 2b_1\sum_{i=1}^{n} x_i^{\,2} \;=\; 0$$

$$\xrightarrow{:2} \text{I)}\quad -\sum_{i=1}^{n} x_i y_i + n\cdot\overline{x}\cdot\overline{y} - b_1 n\cdot\left(\overline{x}\right)^2 + b_1\sum_{i=1}^{n} x_i^{\,2} \;=\; 0$$

$$\rightarrow\; -b_1 n\cdot\left(\overline{x}\right)^2 + b_1\sum_{i=1}^{n} x_i^{\,2} \;=\; \sum_{i=1}^{n} x_i y_i - n\cdot\overline{x}\cdot\overline{y}$$

$$\rightarrow\; b_1\left[\sum_{i=1}^{n} x_i^{\,2} - n\cdot\left(\overline{x}\right)^2\right] \;=\; \sum_{i=1}^{n} x_i y_i - n\cdot\overline{x}\cdot\overline{y}$$

$$\rightarrow\; b_1 \;=\; \frac{\sum_{i=1}^{n} x_i y_i - n\cdot\overline{x}\cdot\overline{y}}{\sum_{i=1}^{n} x_i^{\,2} - n\cdot\left(\overline{x}\right)^2} \;\overset{\cdot\frac{1}{n}}{\underset{\cdot\frac{1}{n}}{=}}\; \frac{\frac{1}{n}\cdot\sum_{i=1}^{n} x_i y_i - \overline{x}\cdot\overline{y}}{\frac{1}{n}\cdot\sum_{i=1}^{n} x_i^{\,2} - \left(\overline{x}\right)^2} \;=\; \frac{Cov(X,Y)}{V(X)}$$