Vorlesung am 17.11.2008

Regressionsrechnung:

An eine gegebene Menge von Punkten

$$(x_1, y_1), \ldots, (x_n, y_n)$$

wird eine Gerade

$$y = \mathbf{a} \cdot x + \mathbf{b}$$

durch Minimierung von

$$(y_1 - (a \cdot x_1 + b))^2 + \cdots + (y_n - (a \cdot x_n + b))^2$$

angepasst.

Für die Berechnung dieser Regressionsgeraden existiert eine explizite Formel.

Das Vorzeichen der Steigung der Geraden stimmt mit dem Vorzeichen der **empirischen Kovarianz**

$$s_{x,y} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

überein.

Ebenfalls stimmt das Vorzeichen dieser Geraden mit dem der **empirischen Korrelation**

$$\frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}} \in [-1, 1]$$

überein.

Ist diese gleich $+1$ oder gleich -1, so liegen die Punkte alle auf einer Geraden.

Bisher:

Vorstellung einzelner Verfahren zur Analyse von Daten, ohne dass ein Modell zur Entstehung der Daten zugrundegelegt wurde.

Neu:

Herleitung von Modellen zur Beschreibung “fehlerbehafteter” bzw. “zufälliger” Daten.

Konstruktion von Verfahren zur Analyse solcher Daten, bei denen explizite Schranken für die Fehler der Verfahren angegeben werden können.

Zum Niveau dieser Vorlesung:

Bisher wurde jede Vorlesung thematisch etwas neues behandelt.

Ab sofort bauen die weiteren Vorlesungen aufeinander auf: Wenn man eine Vorlesung nicht versteht, bekommt man auch Schwierigkeiten bei den weiteren Vorlesungen.

Daher meine Bitte: Bitte jede Vorlesung daheim nacharbeiten!

Und nicht zu früh aufgeben: Das Verstehen von Mathematik braucht manchmal auch einfach nur Zeit . . .