Vorlesung 7b

Korrelationskoeffizient und Regressionsgerade

Teil 3

Beste affin lineare Vorhersage:

die Regressionsgerade

(Buch S. 62-64)

Wie am Ende von Teil 2 angekündigt, wollen wir einsehen:

Das Quadrat des Korrelationskoeffizienten ist ein Maß dafür,

um wieviel besser man Y

durch eine affin lineare Funktion von X vorhersagen kann:

$$Y = \beta_1 X + \beta_0 + \text{“Fehler”},$$

als durch eine Konstante:

$$Y = c + \text{“Fehler”}.$$

Dazu fragen wir erst einmal:

Durch welche Konstante wird die Zufallsvariable Y

(im Sinn des erwarteten quadratischen Fehlers)

am besten vorhergesagt?

Die Antwort ist:

Durch ihren Erwartungswert $\mathbf{E}[Y]$!

Denn:

$$\mathbf{E}[(Y-c)^2] = \stackrel{!}{=} \min$$

$$= \mathbb{E}\left[(Y-\mu_Y+\mu_Y-c)^2\right]$$

$$= \mathbb{E}\left[\underbrace{(Y-\mu_Y)^2 + (\mu_Y-c)^2 + 2(Y-\mu_Y)(\mu_Y-c)}\right]$$

$$\stackrel{\uparrow}{=} \underbrace{\mathbb{E}\left[(Y-\mu_Y)^2\right]}_{=\sigma_Y^2} + \underbrace{\mathbb{E}\left[(\mu_Y-c)^2\right]}_{=(\mu_Y-c)^2} + 2(\mu_Y-c)\underbrace{\mathbb{E}[Y-\mu_Y]}_{0}$$

Lin. des EW

$$= \sigma_Y^2 + (\mu_Y-c)^2$$

wird minimal für $c = \mu_Y$

mit Minimalwert σ_Y^2

$$\mathbf{E}[(Y-c)^2] = \mathbf{E}[(Y-\mu_Y+\mu_Y-c)^2]$$

$$= \mathbf{E}[(Y-\mu_Y)^2] + 2\mathbf{E}[(Y-\mu_Y)(\mu_Y-c)] + (\mu_Y-c)^2$$

$$= \sigma_Y^2 + 0 + (\mu_Y-c)^2.$$

Das wird minimiert von

$$c = \mu_Y$$

und hat den Minimalwert

$$\sigma_Y^2.$$

Jetzt fragen wir:

Durch welche **affin lineare Funktion von** X,

$\beta_1 X + \beta_0$,

wird die Zufallsvariable Y

(wieder im Sinn des erwarteten quadratischen Fehlers)

am besten vorhergesagt?

Genauer:

Für welche Zahlen β_1, β_0 wird

$\mathbf{E}[(Y - \beta_1 X - \beta_0)^2]$ minimal?

Wie wir gleich sehen werden, ist die Lösung:

$$\boxed{\beta_1 := \frac{\sigma_Y}{\sigma_X} \kappa_{XY}}$$

Regressions-
Koeffizient

und β_0 so, dass $\mu_Y = \beta_1 \mu_X + \beta_0$.

M. a. W.: β_0 so, dass der Punkt (μ_X, μ_Y)

auf der Geraden $y = \beta_1 x + \beta_0$ liegt.

Wir nennen diese Gerade

die **Regressionsgerade** für Y auf der Basis von X.

Wir nennen diese Gerade

die **Regressionsgerade** für Y auf der Basis von X.

Wir begründen jetzt die Behauptung über β_0 und β_1:

$\mathbb{E}[H^2] = \mathrm{Var}[H] + \mathbb{E}[H]^2$

Wir begründen jetzt die Behauptung über β_0 und β_1:

$$\mathbf{E}[\overbrace{(Y - \beta_1 X - \beta_0)}^{=:H}{}^2]$$

$$= \mathbf{Var}[Y - \beta_1 X - \beta_0] + (\mathbf{E}[Y - \beta_1 X - \beta_0])^2$$

$$= \mathbf{Var}[Y - \beta_1 X] + (\mu_Y - \beta_1 \mu_X - \beta_0)^2$$

Wir begründen jetzt die Behauptung über β_0 und β_1:

$$\begin{aligned} & \mathbf{E}[(Y - \beta_1 X - \beta_0)^2] \\ & = \mathbf{Var}[Y - \beta_1 X - \beta_0] + (\mathbf{E}[Y - \beta_1 X - \beta_0])^2 \\ & = \mathbf{Var}[Y - \beta_1 X] + (\mu_Y - \beta_1 \mu_X - \beta_0)^2 \end{aligned}$$

Der zweite Summand ist Null für $\beta_0 = \mu_Y - \beta_1 \mu_X$.

Damit haben wir schon mal die eine Bedingung gefunden.

Für welches β_1 wird der erste Summand minimal?

Im Rest des Abschnittes schreiben wir um der besseren Lesbarkeit willen κ statt κ_{XY}.

$$\mathbf{Var}[Y - \beta_1 X] = \mathbf{Var}[Y] - 2\beta_1 \mathbf{Cov}[X, Y] + \beta_1^2 \mathbf{Var}[X]$$

$$= \sigma_Y^2 - 2\beta_1 \kappa\, \sigma_X \sigma_Y + \beta_1^2 \sigma_X^2$$

$$= \sigma_Y^2 - \sigma_Y^2 \kappa^2 + (\sigma_Y^2 \kappa^2 - 2\beta_1 \kappa\, \sigma_X \sigma_Y + \beta_1^2 \sigma_X^2)$$

$$\mathbf{Var}[Y-\beta_1 X] = \mathbf{Var}[Y] - 2\beta_1 \mathbf{Cov}[X,Y] + \beta_1^2 \mathbf{Var}[X]$$

$$= \sigma_Y^2 - 2\beta_1 \kappa\, \sigma_X \sigma_Y + \beta_1^2 \sigma_X^2$$

$$= \sigma_Y^2 - \sigma_Y^2 \kappa^2 \quad + \quad \underline{(\sigma_Y \kappa - \beta_1 \sigma_X)^2}$$

Der rechte Summand wird Null für

$$\beta_1 = \frac{\sigma_Y}{\sigma_X} \kappa.$$

Und der Minimalwert von $\mathbf{Var}[Y-\beta_1 X]$ ist $\sigma_Y^2(1-\kappa^2)$.

Damit ist auch der Minimalwert von $\mathbf{Var}[Y - \beta_1 X - \beta_0\,1]$

gleich $\sigma_Y^2(1-\kappa^2)$.

Der Minimalwert von $\mathbf{Var}[Y - c\,1]$ war σ_Y^2.

Die Verbesserung der Approximation ("Vorhersage") von Y

im quadratischen Mittel, wenn man zu den Vielfachen von 1

die Vielfachen von X dazunimmt, beträgt

$$\sigma_Y^2 - \sigma_Y^2(1-\kappa^2) = \kappa^2\sigma_Y^2.$$

Also ist der Anteil von σ_Y^2,

der von den Vielfachen von X zusätzlich zu

den Vielfachen von 1 "erklärt" wird, gleich $\kappa^2\sigma_Y^2$.

Wir halten fest: Die Minimierungsaufgabe

$$\mathbf{E}[(Y - \beta_1 X - \beta_0)^2] \stackrel{!}{=} \min$$

für die *beste affin lineare Vorhersage von* Y

auf der Basis von X

(im Sinn des quadratischen Mittels)

hat die Lösung

$$\beta_1 = \frac{\sigma_Y}{\sigma_X}\kappa, \quad \mu_Y = \beta_1 \mu_X + \beta_0$$

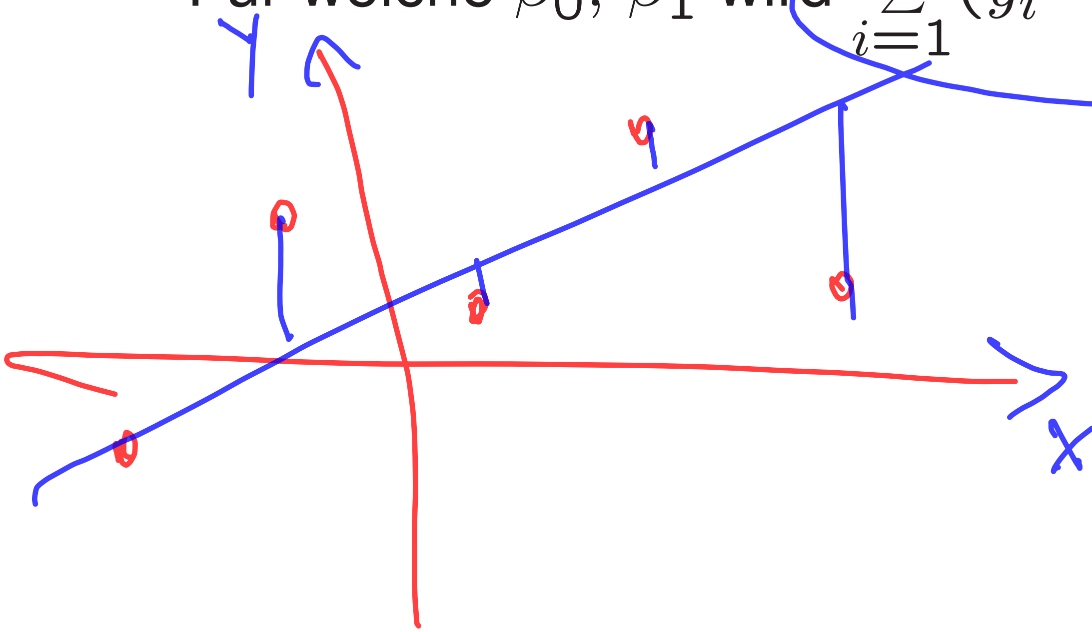
und den Minimalwert $\quad (1 - \kappa^2)\, \sigma_Y^2.$

Beispiel: “Welche Gerade passt am besten?”

Die Methode der kleinsten Quadrate.

$(x_1, y_1), \ldots, (x_n, y_n)$ seien n verschiedene Punkte im $\mathbb{R}^2$.

Für welche β_0, β_1 wird $\sum_{i=1}^{n} (y_i - \beta_1 x_i - \beta_0)^2$ minimal?



Beispiel: “Welche Gerade passt am besten?”

Die Methode der kleinsten Quadrate.

$(x_1, y_1), \ldots, (x_n, y_n)$ seien n verschiedene Punkte im $\mathbb{R}^2$.

Für welche β_0, β_1 wird $\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_1 x_i - \beta_0)^2$ minimal?

Diese Aufgabe interpretieren wir stochastisch. Sei J uniform verteilt auf $\{1, \ldots, n\}$ und $(X, Y) := (x_J, y_J)$, also

$$\mathbf{P}((X,Y) = (x_i, y_i)) = \frac{1}{n}, \quad i = 1, \ldots, n.$$

Diese Aufgabe interpretieren wir stochastisch. Sei J uniform verteilt auf $\{1,\ldots,n\}$ und $(X,Y) := (x_J, y_J)$, also

$$\mathbf{P}((X,Y) = (x_i, y_i)) = \frac{1}{n}, \quad i = 1,\ldots,n.$$

Dann ist

$$\mathbf{E}[X] = \frac{1}{n}\sum x_i =: \bar{x}, \quad \mathbf{E}[Y] = \frac{1}{n}\sum y_i =: \bar{y},$$

$$\sigma_X^2 = \frac{1}{n}\sum (x_i - \bar{x})^2, \quad \sigma_Y^2 = \frac{1}{n}\sum (y_i - \bar{y})^2,$$

$$\mathbf{Cov}[X, Y] = \frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y}),$$

$$\kappa := \kappa_{XY} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2}\sqrt{\sum (y_i - \bar{y})^2}}.$$

$$\mathbf{E}[(Y - \beta_1 X - \beta_0)^2] = \frac{1}{n} \sum_{i=1}^{n} (y_i - \beta_1 x_i - \beta_0)^2$$

wird, wie wir gezeigt haben, minimiert durch

$$\boxed{\beta_1 := \frac{\sigma_Y}{\sigma_X}\kappa = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}}$$

und β_0 so, dass $\bar{y} = \beta_1 \bar{x} + \beta_0$.

Diese Gerade $y = \beta_1 x + \beta_0$ heißt die Regressionsgerade zu den Punkten $(x_i, y_i),\ i = 1, \ldots, n$, oder auch *die mit der Methode der kleinsten Quadrate gefundene Ausgleichsgerade*.