

# Reduced-Rank-Regression

Seminar Multivariate Statistik, Regression und Klassifikation

Daniel Herbst

10. Mai 2021

- 1 Einleitung
- 2 Klassisches multivariates Regressionsmodell mit zufälliger Inputvariable
- 3 Reduced-Rank-Regressionsmodell

Ausgangssituation:

$$\mathbf{X} := (X_1, \dots, X_r)^\top \quad \text{und} \quad \mathbf{Y} := (Y_1, \dots, Y_s)^\top$$

mit gemeinsamer Verteilung  $\mathbb{P}^{(\mathbf{X}, \mathbf{Y})}$ ,  $\mathbb{P}^{\mathbf{X}} \ll \lambda^r$ ,  $\mathbb{P}^{\mathbf{Y}} \ll \lambda^s$  sowie  $r, s \in \mathbb{N}$  mit  $s \leq r$ ,

$$\mu_{\mathbf{X}} := \mathbb{E}\mathbf{X}, \quad \mu_{\mathbf{Y}} := \mathbb{E}\mathbf{Y} \quad \text{und} \quad \begin{pmatrix} \Sigma_{\mathbf{X}\mathbf{X}} & \Sigma_{\mathbf{X}\mathbf{Y}} \\ \Sigma_{\mathbf{Y}\mathbf{X}} & \Sigma_{\mathbf{Y}\mathbf{Y}} \end{pmatrix} := \Sigma\left(\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}\right).$$

# Das klassische lineare Regressionsmodell

Klassisches multivariates Regressionsmodell mit zufälliger Inputvariable:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Theta}\mathbf{X} + \mathcal{E},$$

wobei  $\mathbf{X} := (X_1, \dots, X_r)^\top$  und  $\mathbf{Y} := (Y_1, \dots, Y_s)^\top$ ,  $s \leq r$ .

- $\boldsymbol{\mu} \in \mathbb{R}^s$  und  $\boldsymbol{\Theta} \in \mathbb{R}^{s \times r}$  unbekannte Parameter
- $\mathcal{E}$  ein (nicht beobachtbarer)  $s$ -dimensionaler zufälliger Fehler mit

$$\mathbb{E}\mathcal{E} = 0 \quad \text{und} \quad \mathbf{X} \text{ und } \mathcal{E} \text{ unkorreliert.}$$

# Das klassische lineare Regressionsmodell

Möchten optimale  $\mu$  und  $\Theta$  finden in dem Sinne, dass

$$W(\mu, \Theta) := \mathbb{E}[(\mathbf{Y} - \mu - \Theta\mathbf{X})(\mathbf{Y} - \mu - \Theta\mathbf{X})^\top] \in \mathbb{R}^{s \times s}$$

„klein“ wird.

## Definition (Löwner-Halbordnung)

Sei  $V := \{\mathbf{A} \in \mathbb{R}^{s \times s} \mid \mathbf{A}^\top = \mathbf{A}\}$  der Raum der symmetrischen  $s \times s$ -Matrizen.

$$\forall \mathbf{A}, \mathbf{B} \in V : \mathbf{A} \leq_L \mathbf{B} :\Leftrightarrow \mathbf{B} - \mathbf{A} \text{ positiv semidefinit.}$$

*Es lässt sich leicht nachprüfen, dass diese Relation eine Halbordnung auf  $V$ , die sogenannte Löwner-Halbordnung, definiert.*

Löwner-Halbordnung:  $\mathbf{A} \leq_L \mathbf{B} :\Leftrightarrow \mathbf{B} - \mathbf{A}$  pos. semidef.

Wollen:  $W(\boldsymbol{\mu}, \boldsymbol{\Theta}) := \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})(\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})^\top] \rightarrow \min$

Falls  $\lambda_j(\mathbf{A})$  der  $j$ -größte Eigenwert von  $\mathbf{A} \in \mathbb{R}^{s \times s}$ , dann

$$\mathbf{A} \leq_L \mathbf{B} \implies \lambda_j(\mathbf{A}) \leq \lambda_j(\mathbf{B}), \quad j = 1, \dots, s.$$

Also:

$$\text{tr}(W(\boldsymbol{\mu}, \boldsymbol{\Theta})) = \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})^\top (\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})],$$

wird ebenso minimiert.

$$W(\boldsymbol{\mu}, \boldsymbol{\Theta}) := \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})(\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})^\top], \quad \mathbf{A} \leq_L \mathbf{B} :\Leftrightarrow \mathbf{B} - \mathbf{A} \text{ pos. semidef.}$$

## Satz

*In der Situation von vorher ist*

$$\underset{(\boldsymbol{\mu}, \boldsymbol{\Theta})}{\operatorname{argmin}} W(\boldsymbol{\mu}, \boldsymbol{\Theta}) = (\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}_{\min} \boldsymbol{\mu}_{\mathbf{X}}, \Sigma_{\mathbf{YX}} \Sigma_{\mathbf{XX}}^{-1}) =: (\boldsymbol{\mu}_{\min}, \boldsymbol{\Theta}_{\min}) \quad (\text{bezüglich } \leq_L),$$

$$W(\boldsymbol{\mu}_{\min}, \boldsymbol{\Theta}_{\min}) = \Sigma_{\mathbf{YY}} - \Sigma_{\mathbf{YX}} \Sigma_{\mathbf{XX}}^{-1} \Sigma_{\mathbf{XY}}.$$

*Ferner gilt für  $\mathcal{E} := \mathbf{Y} - \boldsymbol{\mu}_{\min} - \boldsymbol{\Theta}_{\min} \mathbf{X}$ , dass*

$$\mathbb{E}\mathcal{E} = \mathbf{0} \quad \text{und} \quad \mathbb{E}[\mathcal{E}(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})^\top] = \mathbf{0}.$$

*Beweis.* Mit  $\mathbf{X}_c := \mathbf{X} - \mu_{\mathbf{X}}$ ,  $\mathbf{Y}_c := \mathbf{Y} - \mu_{\mathbf{Y}}$  gilt für alle  $\Theta \in \mathbb{R}^{s \times r}$ ,  $\mu \in \mathbb{R}^s$ :

$$\begin{aligned}
 W(\mu, \Theta) &= \mathbb{E}[(\mathbf{Y} - \mu - \Theta \mathbf{X})(\mathbf{Y} - \mu - \Theta \mathbf{X})^\top] \\
 &= \mathbb{E}[(\mathbf{Y}_c - \Theta \mathbf{X}_c + (\mu_{\mathbf{Y}} - \Theta \mu_{\mathbf{X}} - \mu))(\mathbf{Y}_c - \Theta \mathbf{X}_c + (\mu_{\mathbf{Y}} - \Theta \mu_{\mathbf{X}} - \mu))^\top] \\
 &= \Sigma_{\mathbf{Y}\mathbf{Y}} - \Sigma_{\mathbf{Y}\mathbf{X}}\Theta^\top - \Theta\Sigma_{\mathbf{X}\mathbf{Y}} + \Theta\Sigma_{\mathbf{X}\mathbf{X}}\Theta^\top \\
 &\quad + (\mu_{\mathbf{Y}} - \Theta\mu_{\mathbf{X}} - \mu)(\mu_{\mathbf{Y}} - \Theta\mu_{\mathbf{X}} - \mu)^\top \\
 &= \Sigma_{\mathbf{Y}\mathbf{Y}} - \Sigma_{\mathbf{Y}\mathbf{X}}\Sigma_{\mathbf{X}\mathbf{X}}^{-1}\Sigma_{\mathbf{X}\mathbf{Y}} \\
 &\quad + \underbrace{(\Sigma_{\mathbf{Y}\mathbf{X}}\Sigma_{\mathbf{X}\mathbf{X}}^{-1/2} - \Theta\Sigma_{\mathbf{X}\mathbf{X}}^{1/2})(\Sigma_{\mathbf{Y}\mathbf{X}}\Sigma_{\mathbf{X}\mathbf{X}}^{-1/2} - \Theta\Sigma_{\mathbf{X}\mathbf{X}}^{1/2})^\top}_{=:A(\Theta) \quad \text{positiv semidefinit}} \\
 &\quad + \underbrace{(\mu_{\mathbf{Y}} - \Theta\mu_{\mathbf{X}} - \mu)(\mu_{\mathbf{Y}} - \Theta\mu_{\mathbf{X}} - \mu)^\top}_{=:B(\mu, \Theta) \quad \text{positiv semidefinit}} \\
 &\geq \Sigma_{\mathbf{Y}\mathbf{Y}} - \Sigma_{\mathbf{Y}\mathbf{X}}\Sigma_{\mathbf{X}\mathbf{X}}^{-1}\Sigma_{\mathbf{X}\mathbf{Y}}.
 \end{aligned}$$



$$W(\boldsymbol{\mu}, \boldsymbol{\Theta}) = \Sigma_{\mathbf{Y}\mathbf{Y}} - \Sigma_{\mathbf{Y}\mathbf{X}}\Sigma_{\mathbf{X}\mathbf{X}}^{-1}\Sigma_{\mathbf{X}\mathbf{Y}} + A(\boldsymbol{\Theta}) + B(\boldsymbol{\mu}, \boldsymbol{\Theta})$$

$\implies W(\boldsymbol{\mu}, \boldsymbol{\Theta})$  minimal für  $A(\boldsymbol{\Theta}) = B(\boldsymbol{\mu}, \boldsymbol{\Theta}) = \mathbf{0}$ , was der Fall ist, wenn

$$\boldsymbol{\Theta} = \Sigma_{\mathbf{Y}\mathbf{X}}\Sigma_{\mathbf{X}\mathbf{X}}^{-1} =: \boldsymbol{\Theta}_{\min} \quad \text{und} \quad \boldsymbol{\mu} = \boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}_{\min}\boldsymbol{\mu}_{\mathbf{X}} =: \boldsymbol{\mu}_{\min}.$$

Man erhält nun

$$\begin{aligned} \mathcal{E} &= \mathbf{Y} - \boldsymbol{\mu}_{\min} - \boldsymbol{\Theta}_{\min}\mathbf{X} \\ &= \mathbf{Y} - (\boldsymbol{\mu}_{\mathbf{Y}} - (\Sigma_{\mathbf{Y}\mathbf{X}}\Sigma_{\mathbf{X}\mathbf{X}}^{-1})\boldsymbol{\mu}_{\mathbf{X}}) - (\Sigma_{\mathbf{Y}\mathbf{X}}\Sigma_{\mathbf{X}\mathbf{X}}^{-1})\mathbf{X} \\ &= \mathbf{Y}_c - \Sigma_{\mathbf{Y}\mathbf{X}}\Sigma_{\mathbf{X}\mathbf{X}}^{-1}\mathbf{X}_c \end{aligned}$$

und es gilt offensichtlich  $\mathbb{E}\mathcal{E} = \mathbf{0}$  sowie

$$\mathbb{E}[\mathcal{E}\mathbf{X}_c^{\top}] = \mathbb{E}[(\mathbf{Y}_c - \Sigma_{\mathbf{Y}\mathbf{X}}\Sigma_{\mathbf{X}\mathbf{X}}^{-1}\mathbf{X}_c)\mathbf{X}_c^{\top}] = \Sigma_{\mathbf{Y}\mathbf{X}} - \Sigma_{\mathbf{Y}\mathbf{X}}\Sigma_{\mathbf{X}\mathbf{X}}^{-1}\Sigma_{\mathbf{X}\mathbf{X}} = \mathbf{0}.$$

# Das Reduced-Rank-Regressionsmodell

Nun ein etwas anderes Regressionsmodell:

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{C}\mathbf{X} + \mathcal{E},$$

wobei

- $\boldsymbol{\mu} \in \mathbb{R}^s$  und  $\mathbf{C} \in \mathbb{R}^{s \times r}$  unbekannte Parameter
- $\mathcal{E}$  nicht beobachtbarer, zufälliger Fehler mit  $\mathbb{E}\mathcal{E} = 0$  und Kovarianzmatrix  $\Sigma_{\mathcal{E}\mathcal{E}}$
- $\mathbf{X}$  und  $\mathcal{E}$  unkorreliert
- es gibt ein  $t \in \{0, \dots, s\}$  mit  $\text{rk } \mathbf{C} \leq t \leq s$ .

In diesem Fall gibt es Matrizen  $\mathbf{A} \in \mathbb{R}^{s \times t}$  und  $\mathbf{B} \in \mathbb{R}^{t \times r}$  mit  $\mathbf{C} = \mathbf{A}\mathbf{B}$ , d.h. das Modell lässt sich auch schreiben als

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{B}\mathbf{X} + \mathcal{E}.$$

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{C}\mathbf{X} + \mathcal{E}, \quad \mathbf{C} = \mathbf{A}\mathbf{B} \text{ mit } \mathbf{C} \in \mathbb{R}^{s \times r}, \mathbf{A} \in \mathbb{R}^{s \times t}, \mathbf{B} \in \mathbb{R}^{t \times r}$$

$$W(\boldsymbol{\mu}, \boldsymbol{\Theta}) := \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})(\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})^\top], \quad \mathbf{A} \leq_L \mathbf{B} :\Leftrightarrow \mathbf{B} - \mathbf{A} \text{ pos. semidef.}$$

Möchten auch hier wieder  $\boldsymbol{\mu} \in \mathbb{R}^s$ ,  $\mathbf{A} \in \mathbb{R}^{s \times t}$  und  $\mathbf{B} \in \mathbb{R}^{t \times r}$  finden, die  $\mathbf{Y} - \boldsymbol{\mu} - \mathbf{A}\mathbf{B}\mathbf{X}$  auf eine gewisse Art und Weise minimieren, gehen nun aber etwas anders vor:  
 $\boldsymbol{\Gamma} \in \mathbb{R}^{s \times s}$  symmetrisch positiv definite Gewichtsmatrix, minimiere

$$W_{t,\boldsymbol{\Gamma}}(\boldsymbol{\mu}, \mathbf{A}, \mathbf{B}) := \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \mathbf{A}\mathbf{B}\mathbf{X})^\top \boldsymbol{\Gamma} (\mathbf{Y} - \boldsymbol{\mu} - \mathbf{A}\mathbf{B}\mathbf{X})]$$

über  $\boldsymbol{\mu}$ ,  $\mathbf{A}$  und  $\mathbf{B}$ .

Seien  $\mathbf{X}$  und  $\mathbf{Y}$  wie vorher und  $0 \leq t \leq s$ . Dann ist

$$\operatorname{argmin}_{(\boldsymbol{\mu}, \mathbf{A}, \mathbf{B})} W_{t, \Gamma}(\boldsymbol{\mu}, \mathbf{A}, \mathbf{B}) = (\boldsymbol{\mu}_{\min}^{(t)}, \mathbf{A}_{\min}^{(t)}, \mathbf{B}_{\min}^{(t)}) \quad \text{mit}$$

$$\mathbf{A}_{\min}^{(t)} := \Gamma^{-1/2} \mathbf{U}_t \quad \in \mathbb{R}^{s \times t}$$

$$\mathbf{B}_{\min}^{(t)} := \mathbf{U}_t^\top \Gamma^{1/2} \Sigma_{\mathbf{YX}} \Sigma_{\mathbf{XX}}^{-1} \quad \in \mathbb{R}^{t \times r}$$

$$\boldsymbol{\mu}_{\min}^{(t)} := \boldsymbol{\mu}_{\mathbf{Y}} - \mathbf{A}_{\min}^{(t)} \mathbf{B}_{\min}^{(t)} \boldsymbol{\mu}_{\mathbf{X}} \quad \in \mathbb{R}^s,$$

wobei  $\mathbf{U}_t := (\mathbf{u}_1, \dots, \mathbf{u}_t) \in \mathbb{R}^{s \times t}$  und  $\mathbf{u}_1, \dots, \mathbf{u}_t$  orthonormale Eigenvektoren zu den Eigenwerten  $\lambda_1 \geq \dots \geq \lambda_t \geq 0$  von  $\Gamma^{1/2} \Sigma_{\mathbf{YX}} \Sigma_{\mathbf{XX}}^{-1} \Sigma_{\mathbf{XY}} \Gamma^{1/2}$  seien. Ferner gilt

$$W_{t, \Gamma}(\boldsymbol{\mu}_{\min}^{(t)}, \mathbf{A}_{\min}^{(t)}, \mathbf{B}_{\min}^{(t)}) = \operatorname{tr}(\Sigma_{\mathbf{YY}}) \operatorname{tr}(\Gamma) - \sum_{i=1}^t \lambda_i.$$

Wollen:  $W_{t,\Gamma}(\boldsymbol{\mu}, \mathbf{C}) := \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \mathbf{C}\mathbf{X})^\top \Gamma (\mathbf{Y} - \boldsymbol{\mu} - \mathbf{C}\mathbf{X})] \rightarrow \min$

*Beweis.* Für  $\boldsymbol{\mu} \in \mathbb{R}^s$ ,  $\mathbf{C} \in \mathbb{R}^{s \times r}$  und  $\mathbf{X}_c := \mathbf{X} - \boldsymbol{\mu}_X$ ,  $\mathbf{Y}_c := \mathbf{Y} - \boldsymbol{\mu}_Y$  gilt:

$$\begin{aligned} W_{t,\Gamma}(\boldsymbol{\mu}, \mathbf{C}) &= \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \mathbf{C}\mathbf{X})^\top \Gamma (\mathbf{Y} - \boldsymbol{\mu} - \mathbf{C}\mathbf{X})] \\ &= \mathbb{E}[(\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c + (\boldsymbol{\mu}_Y - \mathbf{C}\boldsymbol{\mu}_X - \boldsymbol{\mu}))^\top \Gamma (\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c + (\boldsymbol{\mu}_Y - \mathbf{C}\boldsymbol{\mu}_X - \boldsymbol{\mu}))] \\ &= \mathbb{E}[(\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c)^\top \Gamma (\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c)] \\ &\quad + \underbrace{(\boldsymbol{\mu}_Y - \mathbf{C}\boldsymbol{\mu}_X - \boldsymbol{\mu})^\top \Gamma (\boldsymbol{\mu}_Y - \mathbf{C}\boldsymbol{\mu}_X - \boldsymbol{\mu})}_{\geq 0 \quad (\text{da } \Gamma \text{ symmetrisch positiv definit})} \end{aligned}$$

Wollen:  $\mathbb{E}[(\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c)^\top \Gamma (\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c)] \rightarrow \min$

Mit  $\Sigma_{\mathbf{X}\mathbf{X}}^* := \Sigma_{\mathbf{X}\mathbf{X}}$ ,  $\Sigma_{\mathbf{Y}\mathbf{Y}}^* := \Gamma^{1/2} \Sigma_{\mathbf{Y}\mathbf{Y}} \Gamma^{1/2}$ ,  $\Sigma_{\mathbf{Y}\mathbf{X}}^* := \Gamma^{1/2} \Sigma_{\mathbf{Y}\mathbf{X}}$ ,  $\Sigma_{\mathbf{X}\mathbf{Y}}^* := \Sigma_{\mathbf{X}\mathbf{Y}} \Gamma^{1/2}$  und  $\mathbf{C}^* := \Gamma^{1/2} \mathbf{C}$  ist

$$\begin{aligned}\mathbb{E}[(\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c)^\top \Gamma (\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c)] &= \mathbb{E}[\mathbf{Y}_c^\top \Gamma \mathbf{Y}_c - \mathbf{X}_c^\top \mathbf{C}^\top \Gamma \mathbf{Y}_c - \mathbf{Y}_c^\top \Gamma \mathbf{C} \mathbf{X}_c + \mathbf{X}_c^\top \mathbf{C}^\top \Gamma \mathbf{C} \mathbf{X}_c] \\ &= \text{tr}(\Sigma_{\mathbf{Y}\mathbf{Y}}^* - \mathbf{C}^* \Sigma_{\mathbf{X}\mathbf{Y}}^* - \Sigma_{\mathbf{Y}\mathbf{X}}^* \mathbf{C}^{*\top} + \mathbf{C}^* \Sigma_{\mathbf{X}\mathbf{X}}^* \mathbf{C}^{*\top}) \\ &= \text{tr}(\Sigma_{\mathbf{Y}\mathbf{Y}}^* - \Sigma_{\mathbf{Y}\mathbf{X}}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*-1} \Sigma_{\mathbf{X}\mathbf{Y}}^*) \\ &\quad + \text{tr}((\mathbf{C}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*1/2} - \Sigma_{\mathbf{Y}\mathbf{X}}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*-1/2}) \\ &\quad \cdot (\mathbf{C}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*1/2} - \Sigma_{\mathbf{Y}\mathbf{X}}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*-1/2})^\top ).\end{aligned}$$

$$\mathbb{E}[(\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c)^\top \Gamma (\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c)] = \text{Konst.} + \text{tr} \left( (\mathbf{C}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*1/2} - \Sigma_{\mathbf{Y}\mathbf{X}}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*-1/2}) \cdot (\mathbf{C}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*1/2} - \Sigma_{\mathbf{Y}\mathbf{X}}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*-1/2})^\top \right)$$

## Lemma (Satz von Eckart-Young)

Seien  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$  und  $b := \text{rk } \mathbf{B} \leq \text{rk } \mathbf{A} =: r$ , und sei  $\lambda_j(\mathbf{C})$  für eine reelle symmetrische Matrix  $\mathbf{C}$  deren  $j$ -größter Eigenwert. Dann gilt:

$$\lambda_j((\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})^\top) \geq \lambda_{j+b}(\mathbf{A}\mathbf{A}^\top)$$

mit Gleichheit für

$$\mathbf{A}_b := \sum_{i=1}^b \lambda_i^{1/2} \mathbf{u}_i \mathbf{v}_i^\top,$$

wobei  $\lambda_i := \lambda_i(\mathbf{A}\mathbf{A}^\top)$  und  $\mathbf{u}_i$  bzw.  $\mathbf{v}_i$  jeweils orthonormale Eigenvektoren von  $\mathbf{A}\mathbf{A}^\top$  bzw.  $\mathbf{A}^\top \mathbf{A}$  zu  $\lambda_i$  seien.

$$\mathbb{E}[(\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c)^\top \Gamma (\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c)] = \text{Konst.} + \text{tr} \left( (\mathbf{C}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*1/2} - \Sigma_{\mathbf{Y}\mathbf{X}}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*-1/2}) \cdot (\mathbf{C}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*1/2} - \Sigma_{\mathbf{Y}\mathbf{X}}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*-1/2})^\top \right)$$

$k := \text{rk } \Sigma_{\mathbf{Y}\mathbf{X}} = \text{rk } \Sigma_{\mathbf{Y}\mathbf{X}}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*-1/2}$  und seien  $\mathbf{u}_i$  orthonormale Eigenvektoren zu  $\lambda_i$  von

$$(\Sigma_{\mathbf{Y}\mathbf{X}}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*-1/2})(\Sigma_{\mathbf{Y}\mathbf{X}}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*-1/2})^\top = \Sigma_{\mathbf{Y}\mathbf{X}}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*-1} \Sigma_{\mathbf{X}\mathbf{Y}}^* = \Gamma^{1/2} \Sigma_{\mathbf{Y}\mathbf{X}} \Sigma_{\mathbf{X}\mathbf{X}}^{-1} \Sigma_{\mathbf{X}\mathbf{Y}} \Gamma^{1/2}, \quad 1 \leq i \leq s,$$

$$\mathbf{v}_i := \lambda_i^{-1/2} \Sigma_{\mathbf{X}\mathbf{X}}^{*-1/2} \Sigma_{\mathbf{X}\mathbf{Y}}^* \mathbf{u}_i = \lambda_i^{-1/2} \Sigma_{\mathbf{X}\mathbf{X}}^{-1/2} \Sigma_{\mathbf{X}\mathbf{Y}} \Gamma^{1/2} \mathbf{u}_i, \quad 1 \leq i \leq k$$

$\xrightarrow{\text{Eckart-Young}}$  Zweiter Term von (\*) minimal für

$$\mathbf{C}^* \Sigma_{\mathbf{X}\mathbf{X}}^{*1/2} = \Gamma^{1/2} \mathbf{C} \Sigma_{\mathbf{X}\mathbf{X}}^{1/2} = \sum_{i=1}^{\min\{k,t\}} \lambda_i^{1/2} \mathbf{u}_i \mathbf{v}_i^\top \quad \text{also}$$

$$\mathbf{C}_{\min}^{(t)} := \Gamma^{-1/2} \left( \sum_{i=1}^{\min\{k,t\}} \lambda_i^{1/2} \mathbf{u}_i \mathbf{v}_i^\top \right) \Sigma_{\mathbf{X}\mathbf{X}}^{-1/2} = \Gamma^{-1/2} \left( \sum_{i=1}^t \mathbf{u}_i \mathbf{u}_i^\top \right) \Gamma^{1/2} \Sigma_{\mathbf{Y}\mathbf{X}} \Sigma_{\mathbf{X}\mathbf{X}}^{-1}.$$



$$\begin{aligned}
W_{t,\Gamma}(\boldsymbol{\mu}, \mathbf{C}) &:= \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \mathbf{C}\mathbf{X})^\top \Gamma (\mathbf{Y} - \boldsymbol{\mu} - \mathbf{C}\mathbf{X})] \\
&= \text{Konst.} + \text{tr}\left( (\mathbf{C}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*1/2} - \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*-1/2}) \right. \\
&\quad \cdot (\mathbf{C}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*1/2} - \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*-1/2})^\top \Big) \\
&\quad + (\boldsymbol{\mu}_{\mathbf{Y}} - \mathbf{C}\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu})^\top \Gamma (\boldsymbol{\mu}_{\mathbf{Y}} - \mathbf{C}\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu})
\end{aligned}$$

$$\mathbf{C}_{\min}^{(t)} = \Gamma^{-1/2} \left( \sum_{i=1}^t \mathbf{u}_i \mathbf{u}_i^\top \right) \Gamma^{1/2} \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}$$

$\Rightarrow W_{t,\Gamma}(\boldsymbol{\mu}, \mathbf{A}, \mathbf{B})$  minimal für

$$\mathbf{A}_{\min}^{(t)} = \Gamma^{-1/2} \mathbf{U}_t \quad \in \mathbb{R}^{s \times t}$$

$$\mathbf{B}_{\min}^{(t)} = \mathbf{U}_t^\top \Gamma^{1/2} \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \quad \in \mathbb{R}^{t \times r}$$

$$\boldsymbol{\mu}_{\min}^{(t)} = \boldsymbol{\mu}_{\mathbf{Y}} - \mathbf{A}_{\min}^{(t)} \mathbf{B}_{\min}^{(t)} \boldsymbol{\mu}_{\mathbf{X}} \quad \in \mathbb{R}^s,$$

wobei  $\mathbf{U}_t := (\mathbf{u}_1, \dots, \mathbf{u}_t) \in \mathbb{R}^{s \times t}$ .

# Spezialfälle der Reduced-Rank-Regression

Haben soeben  $W_{t,\Gamma}(\boldsymbol{\mu}, \mathbf{C}) := \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \mathbf{C}\mathbf{X})^\top \Gamma (\mathbf{Y} - \boldsymbol{\mu} - \mathbf{C}\mathbf{X})]$  minimiert

Spezielle Wahlen von  $\Gamma$  und  $\mathbf{X}, \mathbf{Y}$ :

- $\mathbf{X} = \mathbf{Y}$  und  $\Gamma = \mathbf{I}_s$ : *Hauptkomponentenanalyse (principal component analysis, PCA)* (Vortrag 6).
- $\Gamma = \Sigma_{\mathbf{Y}\mathbf{Y}}^{-1}$ : *kanonische Korrelationsanalyse* (Vortrag 7).
- Setup der kanonischen Korrelationsanalyse für einen  $\{0, 1\}$ -wertigen Vektor  $\mathbf{Y}$ , der Gruppenzugehörigkeiten modelliert: *lineare Diskriminanzanalyse* (Vortrag 8).

# Stichprobenschätzung bei der Reduced-Rank-Regression

Zur Minimierung benötigt:  $\mu_{\mathbf{X}}, \mu_{\mathbf{Y}}$  und  $\Sigma_{\mathbf{XX}}, \Sigma_{\mathbf{XY}}, \Sigma_{\mathbf{YX}}, \Sigma_{\mathbf{YY}}$  - typischerweise unbekannt. I.d.R. Stichprobe gegeben:

$$\mathcal{D} := \{(\mathbf{x}_j, \mathbf{y}_j) \mid 1 \leq j \leq n\},$$

unabhängige Realisierungen von  $r$ - bzw.  $s$ -dimensionalen Zufallsvektoren  $\mathbf{X}$  bzw.  $\mathbf{Y}$ .  
Arbeiten stattdessen einfach mit den entsprechenden empirischen Größen:

$$\hat{\mu}_{\mathbf{X}} := \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j = \bar{\mathbf{x}}_n, \quad \hat{\mu}_{\mathbf{Y}} := \frac{1}{n} \sum_{j=1}^n \mathbf{y}_j = \bar{\mathbf{y}}_n$$

$$\mathbf{x}_{cj} := \mathbf{x}_j - \bar{\mathbf{x}}_n, \quad \mathbf{y}_{cj} := \mathbf{y}_j - \bar{\mathbf{y}}_n, \quad \mathcal{X}_c := (\mathbf{x}_{c1}, \dots, \mathbf{x}_{cn}), \quad \mathcal{Y}_c := (\mathbf{y}_{c1}, \dots, \mathbf{y}_{cn})$$

$$\hat{\Sigma}_{\mathbf{XX}} := \frac{1}{n} \mathcal{X}_c \mathcal{X}_c^{\top} \in \mathbb{R}^{r \times r}, \quad \hat{\Sigma}_{\mathbf{YY}} := \frac{1}{n} \mathcal{Y}_c \mathcal{Y}_c^{\top} \in \mathbb{R}^{s \times s},$$

$$\hat{\Sigma}_{\mathbf{YX}} := \frac{1}{n} \mathcal{Y}_c \mathcal{X}_c^{\top} = \hat{\Sigma}_{\mathbf{XY}}^{\top} \in \mathbb{R}^{s \times r}.$$

# Stichprobenschätzung bei der Reduced-Rank-Regression

$$\hat{\mathbf{C}}_{\min}^{(t)} := \mathbf{\Gamma}^{-1/2} \left( \sum_{i=1}^t \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^\top \right) \mathbf{\Gamma}^{1/2} \hat{\Sigma}_{\mathbf{YX}} \hat{\Sigma}_{\mathbf{XX}}^{-1},$$

wobei  $\hat{\mathbf{u}}_i$ ,  $1 \leq i \leq t$  orthonormale Eigenvektoren zu den je  $i$ -größten Eigenwerten  $\lambda_i$ ,  $1 \leq i \leq t$  von  $\mathbf{\Gamma}^{1/2} \hat{\Sigma}_{\mathbf{YX}} \hat{\Sigma}_{\mathbf{XX}}^{-1} \hat{\Sigma}_{\mathbf{XY}} \mathbf{\Gamma}^{1/2}$  seien.  $\mathbf{\Theta}_{\min}$  wird analog durch

$$\hat{\mathbf{\Theta}}_{\min} := \hat{\Sigma}_{\mathbf{YX}} \hat{\Sigma}_{\mathbf{XX}}^{-1}$$

geschätzt.

# Simulationsbeispiel

Erzeugen je  $n = 1000$  unabhängige Realisierungen

$$\mathbf{x}_j \sim \mathcal{N}_3(\mathbf{0}, 0.3 \cdot \mathbf{I}_3), \quad \boldsymbol{\varepsilon}_j \sim \mathcal{N}_3(\mathbf{0}, 0.15 \cdot \mathbf{I}_3), \quad 1 \leq j \leq n,$$

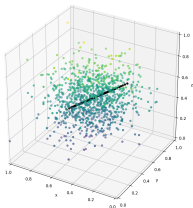
$$\mathbf{C} := \frac{1}{4} \cdot \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \in \mathbb{R}^{3 \times 3}, \quad \boldsymbol{\mu} := \begin{pmatrix} 0.5 \\ 0.5 \\ 0.5 \end{pmatrix}, \quad \mathbf{y}_j := \boldsymbol{\mu} + \mathbf{C}\mathbf{x}_j + \boldsymbol{\varepsilon}_j$$

Stichprobe:

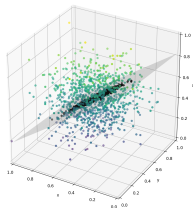
$$\mathcal{D} := \{(\mathbf{x}_j, \mathbf{y}_j) \mid 1 \leq j \leq n\},$$

verwenden  $\boldsymbol{\Gamma} := \mathbf{I}_s$  bei der Reduced-Rank-Regression.

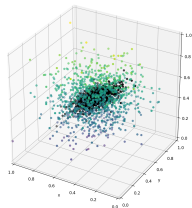
$$\Rightarrow \quad \hat{\boldsymbol{\mu}}_{\min}^{(3)} \approx \begin{pmatrix} 0.504 \\ 0.490 \\ 0.499 \end{pmatrix}, \quad \hat{\mathbf{C}}_{\min}^{(3)} \approx \begin{pmatrix} 0.246 & -0.018 & 0.241 \\ -0.018 & 0.212 & 0.007 \\ -0.017 & -0.017 & -0.279 \end{pmatrix}$$



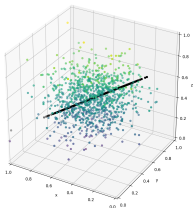
(a)  $x, y; t = 1$



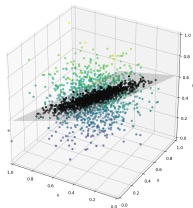
(b)  $x, y; t = 2$



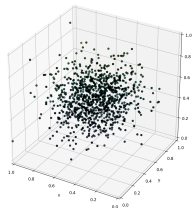
(c)  $x, y; t = 3$



(d)  $y, y; t = 1$



(e)  $y, y; t = 2$



(f)  $y, y; t = 3$

# Effektive Dimension der Regression

Reduced-Rank-Regressionsmodell:  $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{C}\mathbf{X} + \boldsymbol{\varepsilon}, \quad \text{rk } \mathbf{C} \leq t$

Haben zuvor  $W_{t,\Gamma}(\boldsymbol{\mu}, \mathbf{C}) := \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \mathbf{C}\mathbf{X})^\top \Gamma (\mathbf{Y} - \boldsymbol{\mu} - \mathbf{C}\mathbf{X})]$  minimiert

Wie wählt man die *effektive Dimension*  $t$  bei der Reduced-Rank-Regression geeignet?  
Vergrößert man den Rang von  $\mathbf{C}$  von  $t = t_0$  zu  $t = t_1$ ,  $t_0 < t_1$ , so

$$W_{t_0,\Gamma}(\boldsymbol{\mu}_{\min}^{(t_0)}, \mathbf{A}_{\min}^{(t_0)}, \mathbf{B}_{\min}^{(t_0)}) - W_{t_1,\Gamma}(\boldsymbol{\mu}_{\min}^{(t_1)}, \mathbf{A}_{\min}^{(t_1)}, \mathbf{B}_{\min}^{(t_1)}) = \sum_{i=t_0+1}^{t_1} \lambda_i$$

wobei  $\lambda_1 \geq \dots \geq \lambda_t \geq 0$  Eigenwerte von  $\Gamma^{1/2} \Sigma_{\mathbf{YX}} \Sigma_{\mathbf{XX}}^{-1} \Sigma_{\mathbf{XY}} \Gamma^{1/2}$  nach Größe sortiert.

- In der Praxis (mit  $\hat{\lambda}_i$ ) untersuchen, bis wann es sich „lohnt“,  $t$  zu vergrößern
- etwa mit Tests, ob  $\text{rk } \Gamma^{1/2} \Sigma_{\mathbf{YX}} \Sigma_{\mathbf{XX}}^{-1} \Sigma_{\mathbf{XY}} \Gamma^{1/2} = \text{rk } \Sigma_{\mathbf{YX}}$  kleiner als eine festgelegte Zahl ist

Vielen Dank für Ihre Aufmerksamkeit.