Reduced-Rank-Regression

Daniel Herbst

10. Mai 2021

Zusammenfassung

Nachdem in der vorigen Präsentation das multivariate Regressionsmodell mit festen Inputvariablen vorgestellt wurde und bereits Modelle betrachtet wurden, bei denen die Regressionskoeffizienten gewisse (lineare) Nebenbedingungen erfüllen, wollen wir nun eine ähnliche Situation betrachten, bei der wir die Inputvariablen allerdings als zufällig annehmen und den Rang der Regressionskoeffizientenmatrix einschränken. Hierbei handelt es sich dann um das Reduced-Rank-Regressionsmodell, das unter anderem auch einige Techniken der multivariaten Statistik, etwa zur Dimensionsreduktion, verallgemeinert. Im Wesentlichen folgt der Vortrag [Izenman, Kapitel 6.3].

1 Einleitung

Wir betrachten die folgende Situation:

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

seien Zufallsvektoren mit gemeinsamer Verteilung $\mathbb{P}^{(\mathbf{X},\mathbf{Y})}$, wobei $r, s \in \mathbb{N}$ mit $s \leq r$, und ferner definieren wir folgende Schreibweisen für die entsprechenden Erwartungwerte und Kovarianzmatrizen:

$$\boldsymbol{\mu}_{\mathbf{X}} := \mathbb{E}\mathbf{X}, \quad \boldsymbol{\mu}_{\mathbf{Y}} := \mathbb{E}\mathbf{Y} \quad \text{und} \quad \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{XX}} & \boldsymbol{\Sigma}_{\mathbf{XY}} \\ \boldsymbol{\Sigma}_{\mathbf{YX}} & \boldsymbol{\Sigma}_{\mathbf{YY}} \end{pmatrix} := \Sigma\left(\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}\right).$$

Abgesehen von $\mathbb{P}^{\mathbf{X}} \ll \lambda^r$ und $\mathbb{P}^{\mathbf{Y}} \ll \lambda^s$, d.h. der Stetigkeit der Zufallsvektoren $\mathbf{X}$ und $\mathbf{Y}$, möchten wir zunächst keine weiteren Annahmen über die Verteilungen von $\mathbf{X}$ und $\mathbf{Y}$ treffen.

2 Klassisches multivariates Regressionsmodell mit zufälliger Inputvariable

Das klassische multivariate Regressionsmodell mit zufälliger Inputvariable ist von folgender Form: Für $\mathbf{X}$ und $\mathbf{Y}$ gelte

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

wobei $\boldsymbol{\mu} \in \mathbb{R}^s$ und $\boldsymbol{\Theta} \in \mathbb{R}^{s \times r}$ unbekannte Parameter seien sowie $\mathcal{E}$ ein (nicht beobachtbarer) s-dimensionaler zufälliger Fehler ist mit Erwartungswert $\mathbb{E}\mathcal{E} = 0$ und Kovarianzmatrix $\boldsymbol{\Sigma}_{\mathcal{E}\mathcal{E}}$. Zudem werden $\mathbf{X}$ und $\mathcal{E}$ als unabhängig vorausgesetzt.

Ausgehend von der Situation aus der Einleitung wollen wir nun optimale $\boldsymbol{\mu}$ und $\boldsymbol{\Theta}$ finden in dem Sinne, dass

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

in der Spektralnorm

$$\|\mathbf{A}\|_2 := \max_{\|\mathbf{v}\|_2=1} \|\mathbf{A}\mathbf{v}\|_2, \quad \mathbf{A} \in \mathbb{R}^{s\times s}$$

minimiert wird. Für symmetrische, positiv semidefinite Matrizen $\mathbf{A}$ lässt sich (etwa mit dem Spektralsatz) zeigen, dass

$$\|\mathbf{A}\|_2 = \max_{\|\mathbf{v}\|_2=1} \mathbf{v}^\top \mathbf{A}\mathbf{v} = \lambda_{\max}(\mathbf{A}),$$

wobei $\lambda_{\max}(\mathbf{A})$ der größte Eigenvektor von $\mathbf{A}$ sei. Da $W(\boldsymbol{\mu}, \boldsymbol{\Theta})$ für alle möglichen $\boldsymbol{\mu}$, $\boldsymbol{\Theta}$ symmetrisch positiv semidefinit ist, werden wir diese Darstellung der Spektralnorm auch zur Minimierung von $\|W(\boldsymbol{\mu}, \boldsymbol{\Theta})\|_2$ benutzen.

Unter der zusätzlichen Annahme, dass $\boldsymbol{\Sigma}_{\mathbf{XX}}$ nichtsingulär ist, und mit $\mathbf{X}_c := \mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}$, $\mathbf{Y}_c := \mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}}$ gilt für alle $\boldsymbol{\Theta} \in \mathbb{R}^{s\times r}$, $\boldsymbol{\mu} \in \mathbb{R}^s$:

$$\begin{aligned}
W(\boldsymbol{\mu}, \boldsymbol{\Theta}) &= \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})(\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})^\top] \\
&= \mathbb{E}[(\mathbf{Y}_c - \boldsymbol{\Theta}\mathbf{X}_c + (\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu}))(\mathbf{Y}_c - \boldsymbol{\Theta}\mathbf{X}_c + (\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu}))^\top] \\
&= \mathbb{E}[\mathbf{Y}_c\mathbf{Y}_c^\top - \mathbf{Y}_c\mathbf{X}_c^\top\boldsymbol{\Theta}^\top + \mathbf{Y}_c(\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu})^\top \\
&\qquad - \boldsymbol{\Theta}\mathbf{X}_c\mathbf{Y}_c^\top + \boldsymbol{\Theta}\mathbf{X}_c\mathbf{X}_c^\top\boldsymbol{\Theta}^\top - \boldsymbol{\Theta}\mathbf{X}_c(\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu})^\top \\
&\qquad + (\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu})\mathbf{Y}_c^\top - (\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu})\mathbf{X}_c^\top\boldsymbol{\Theta}^\top \\
&\qquad + (\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu})(\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu})^\top] \\
&= \boldsymbol{\Sigma}_{\mathbf{YY}} - \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Theta}^\top - \boldsymbol{\Theta}\boldsymbol{\Sigma}_{\mathbf{XY}} + \boldsymbol{\Theta}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\Theta}^\top \\
&\qquad + (\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu})(\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu})^\top \\
&= \boldsymbol{\Sigma}_{\mathbf{YY}} - \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\Sigma}_{\mathbf{XY}} \\
&\qquad + (\boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1/2} - \boldsymbol{\Theta}\boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2})(\boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1/2} - \boldsymbol{\Theta}\boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2})^\top \\
&\qquad + (\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu})(\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu})^\top,
\end{aligned}$$

und mit

$$A(\boldsymbol{\Theta}) := \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1/2} - \boldsymbol{\Theta}\boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2} \quad , \quad B(\boldsymbol{\mu}, \boldsymbol{\Theta}) := \boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu}$$

gilt für $\mathbf{v} \in \mathbb{R}^s$, $\|\mathbf{v}\|_2 = 1$:

$$\begin{aligned}
\mathbf{v}^\top W(\boldsymbol{\mu}, \boldsymbol{\Theta})\mathbf{v} &= \mathbf{v}^\top(\boldsymbol{\Sigma}_{\mathbf{YY}} - \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\Sigma}_{\mathbf{XY}})\mathbf{v} + \mathbf{v}^\top A(\boldsymbol{\Theta})^\top A(\boldsymbol{\Theta})\mathbf{v} + \mathbf{v}^\top B(\boldsymbol{\mu}, \boldsymbol{\Theta})^\top B(\boldsymbol{\mu}, \boldsymbol{\Theta})\mathbf{v} \\
&= \mathbf{v}^\top(\boldsymbol{\Sigma}_{\mathbf{YY}} - \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\Sigma}_{\mathbf{XY}})\mathbf{v} + \|A(\boldsymbol{\Theta})\mathbf{v}\|_2 + \|B(\boldsymbol{\mu}, \boldsymbol{\Theta})\mathbf{v}\|_2 \\
&\geq \mathbf{v}^\top(\boldsymbol{\Sigma}_{\mathbf{YY}} - \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\Sigma}_{\mathbf{XY}})\mathbf{v}
\end{aligned}$$

und daher ist

$$\|W(\boldsymbol{\mu}, \boldsymbol{\Theta})\|_2 \geq \|\boldsymbol{\Sigma}_{\mathbf{YY}} - \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\Sigma}_{\mathbf{XY}}\|_2,$$

mit Gleichheit falls $A(\boldsymbol{\Theta}) = B(\boldsymbol{\mu}, \boldsymbol{\Theta}) = 0$, was der Fall ist, wenn

$$\boldsymbol{\mu} = \boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}\boldsymbol{\mu}_{\mathbf{X}} \quad \text{und} \quad \boldsymbol{\Theta} = \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}.$$

Dann ist

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

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

$$\mathbb{E}[\mathcal{E}\mathbf{X}_c^\top] = \mathbb{E}[(\mathbf{Y}_c - \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\mathbf{X}_c)\mathbf{X}_c^\top] = \boldsymbol{\Sigma}_{\mathbf{YX}} - \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\Sigma}_{\mathbf{XX}} = 0,$$

d.h. $\mathbf{X}$ und $\mathcal{E}$ sind unkorreliert.

3 Reduced-Rank-Regressionsmodell

Literatur

[Izenman] Izenman, Alan Julian. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. Springer Texts in Statistics, Springer-Verlag New York, 2008.