

2. Lineare Regression

o) lineare Regressionsmodelle

$$\hookrightarrow \text{Es gilt: } y(k) = \Theta^T \varphi(k) + \xi(k)$$

mit $y(k)$ der Ausgang

$\varphi(k)$ Regressionsvektor an der k -te Messung

Θ^T Parametervektor (unabhängig von φ oder y)

ξ Messrauschen

b) lineare Regressionsmodelle

a) statisches System

$$y(k) = a_1 u_1(k) + a_2 u_2(k) + a_3 u_3(k) + \xi(k)$$

$$\Rightarrow \Theta^T = (a_1 \ a_2 \ a_3) ; \varphi^T = (u_1(k) \ u_2(k) \ u_3(k))$$

b) dynamisches System

$$y(k) + a_1 y(k-1) + \dots + a_{n_A} y(k-n_A) = b_0 u(k-d) + \dots + b_{n_B} u(k-d-n_B) + \xi(k)$$

$$\Rightarrow \text{Parameter } \Theta^T = (a_1 \dots a_{n_A} \ b_0 \dots b_{n_B})$$

$$\text{Regressionsvektor } \varphi^T = (-y(k-1) \dots -y(k-n_A) \ u(k-d) \dots u(k-n_B))$$

Bei der Regression kann der Ausgang durch eine Vorhersage / Prädiktor \hat{y} geschätzt.

\hat{y} , die am kleinsten von y abweicht.

\hookrightarrow der beste Prädiktor $\hat{y} = \Theta^T \varphi(k) = \arg \min_{\hat{y}} E[(y-\hat{y})^2]$

$$\Rightarrow E[(y-\hat{y})^2] = E[(\Theta^T \varphi(k) + \xi - \hat{y})^2]$$

$$= E[(\Theta^T \varphi(k) - \hat{y})^2] - 2E[\Theta^T \varphi(k) - \hat{y}] + E[\xi^2]$$

$$= E[(\Theta^T \varphi(k) - \hat{y})^2] + E[\xi^2]$$

$\underbrace{\quad}_{=0}$, da Varianz des Rauschens = 0

\rightarrow damit $E[(\Theta^\top \varphi - \hat{y})^2]$ minimiert (gleich 0), muss gelten: $\hat{y} = \Theta^\top \varphi \Rightarrow \hat{y}(k) = \Theta^\top \varphi(k)$

o) Methode der kleinsten Fehlerquadrate

↪ mit Modell $M(\Theta, \varphi)$: $y(k) = \Theta^\top \varphi(k) + \xi(k)$

\rightarrow gewählte Gütekriterium:

$$J(\Theta) := \frac{1}{K} \sum_{k=1}^K \alpha_k [y(k) - \Theta^\top \varphi(k)]^2$$

\rightarrow optimaler $\Theta = \hat{\Theta} = \arg \min_{\Theta} J(\Theta)$

$$\hookrightarrow 1. \text{ Ableitung}: \frac{\partial J}{\partial \Theta_i} = \frac{2}{K} \left[\Theta_i^\top \sum_{k=1}^K \alpha_k \varphi(k) \varphi_i(k) - \sum_{k=1}^K \alpha_k y(k) \varphi_i(k) \right]$$

\Rightarrow setze gleich 0:

$$\hat{\Theta} = \left[\sum_{k=1}^K \alpha_k \varphi(k) \varphi^\top(k) \right]^{-1} \sum_{k=1}^K \alpha_k y(k) \varphi(k)$$

damit $\hat{\Theta}$ ein Minimum ist, muss $\frac{\partial^2 J}{\partial \Theta_i \partial \Theta_j}$ positiv definit sein.

> als Matrixdarstellung

$$\Phi = \begin{pmatrix} \varphi^\top(1) \\ \vdots \\ \varphi^\top(K) \end{pmatrix}, \quad y = (y(1) \dots y(K)), \quad \xi = (\xi(1) \dots \xi(K))$$

\rightarrow gilt noch immer $y = \Phi \Theta + \xi$

$$\text{für } \alpha_k = 1: \quad J(\Theta) = \frac{1}{K} (y^\top y - 2\Theta^\top \Phi^\top y + \Theta^\top \Phi^\top \Phi \Theta)$$

$$\hookrightarrow \hat{\Theta} = \arg \min_{\Theta} J(\Theta) = \underbrace{(\Phi^\top \Phi)^{-1}}_{\text{Hessianmatrix}} \Phi^\top y$$

damit $\Phi^\top \Phi$ invertierbar ist $\in \mathbb{R}^{K \times K}$ mit $K \geq I$

↳ Interpretation:

- $\hat{\phi}$ entspricht den gewählten Eingangssignalen
- System muss ausreichend angeregt werden, damit $\hat{\phi}^T \hat{\phi}$ nicht singulär wird.
- Verhältnis zw. dem größten zum kleinsten Wurzel der Eigenwerte von $(\hat{\phi}^T \hat{\phi})^{-1} (\hat{\phi}^T \hat{\xi})$ (**Konditionierungszahl**) bestimmt, wie nah die Matrix von Singularität sein darf.

↳ In Matlab kann $\hat{\Theta}$ berechnet mit „theta_hat = Phi / y“

→ Eigenschaften des Least-Squares-Schätzers

↳ da y durch Zufallsgröße gestört wird, wird auch die Schätzung eine Zufallsgröße sein

↳ Biasfreie Schätzung: Erwartungswert von $\hat{\Theta} = \Theta$ (wahren Param.)

↳ Hinreichende Voraussetzung: $E(\xi) = 0$

• ξ und ϕ statistisch unabhängig

↳ Kovarianz des Schätzfehlers (Streuung von $\hat{\Theta}$ um den Erwart.wert)

$P = (\hat{\phi}^T \hat{\phi})^{-1} \sigma_\xi^2$, σ_ξ^2 die Varianz von ξ

→ P steht als ein Maß für die Genauigkeit des Schätzungen
leder ist σ_ξ^2 unbekannt und kann nur geschätzt werden

$$\hat{\sigma}_\xi^2 = \frac{(y - \hat{\phi} \hat{\Theta})^T (y - \hat{\phi} \hat{\Theta})}{k - I}$$

↳ $\hat{\sigma}_\xi^2$ verkleinert wenn:

- mehr Messungen durchgeführt werden
- Konditionierungszahl verkleinern (Testsignale werden komplexer oder gegen das System stärker an!)

> Interpretation der Varianz und Kovarianzmatrix

↪ Normalverteilte Zufallsgröße mit σ^2 :

- 68% Konfidenzintervall $\rightarrow \pm 1\sigma$
- 95% Konfidenzintervall $\rightarrow \pm 2\sigma$
- 99,7% Konfidenzintervall $\rightarrow \pm 3\sigma$

↪ Kovarianzmatrix $P = \begin{pmatrix} \sigma_{\theta_1}^2 & \sigma_{\theta_1, \theta_2} \\ \sigma_{\theta_2, \theta_1} & \sigma_{\theta_2}^2 \end{pmatrix}$ (Beispiel $P \in \mathbb{R}^{2 \times 2}$)

→ Konfidenzintervall als Ellipse dargestellt wird
 → Konfidenzintervall ergibt sich aus der Chi-Quadrat-Verteilung (für zwei Freiheitsgrade)

$P[\chi^2]$	99,5	99	95,7	95	90
χ_p^2	10,6	9,21	7,38	5,99	4,66

> Konfidenzintervall für den prädizierten Systemausgang

↪ $P_{\hat{y}}$ ist die Kovarianzmatrix des prädizierten Ausgangs mit $\hat{y} = \phi_{\text{neu}} \hat{\Theta}$ und $\hat{\Theta} = (\phi_{\text{alt}}^\top \phi_{\text{alt}})^\top \phi_{\text{alt}}^{-1} y$

Ergebnis aus alter Schätzung

$$P_{\hat{y}} = E\{(\hat{y} - E\{\hat{y}\})(\hat{y} - E\{\hat{y}\})^\top\} = \phi_{\text{neu}} E\{(\Theta - \hat{\Theta})(\Theta - \hat{\Theta})^\top\} \phi_{\text{neu}}^\top = \phi_{\text{neu}} P_{\text{alt}} \phi_{\text{neu}}^\top = \phi_{\text{neu}} (\phi_{\text{alt}}^\top \phi_{\text{alt}})^{-1} \phi_{\text{neu}}$$

$\Rightarrow y(i) = \hat{y}(i) \pm k \sqrt{\text{diag}(P_{\hat{y}})}$ wobei k vom konf. Intervall abhängt ($k=1$ für 68%, $k=1,96$ für 95%)

o) Rekursive Least Squares (RLS)

↪ Schätzung wird durch neue Beobachtung geupdated
 $\rightarrow \hat{\theta}_{\text{neu}} = \hat{\theta}_{\text{alt}} + K_{\text{faktor}} (\varphi_{\text{neu}} - \hat{\varphi}_{\text{alt}})$

$$\text{Es gilt: } \hat{\theta}(k-1) = \left[\sum_{i=1}^{k-1} \alpha_i \varphi(i) \varphi^T(i) \right] \sum_{i=1}^{k-1} \alpha_i \varphi(i) y(i)$$

$$\text{und } \hat{\theta}(k) = \left[\sum_{i=1}^{k-1} \alpha_i \varphi(i) \varphi^T(i) + \alpha_k \varphi(k) \varphi^T(k) \right] \left(\sum_{i=1}^{k-1} \alpha_i \varphi(i) y(i) + \alpha_k \varphi(k) y(k) \right)$$

$$\text{sei } R(k) = \sum_{i=1}^k \alpha_i \varphi(i) \varphi^T(i) = R(k-1) + \alpha_k \varphi(k) \varphi^T(k)$$

$$\Rightarrow \hat{\theta}(k-1) = R^{-1}(k-1) \sum_{i=1}^{k-1} \alpha_i \varphi(i) y(i)$$

$$\Rightarrow \hat{\theta}(k) = R^{-1}(k) [R(k-1) \hat{\theta}(k-1) + \alpha_k \varphi(k) y(k)]$$

$$\Rightarrow \hat{\theta}(k) = \hat{\theta}(k-1) + R^{-1}(k) \alpha_k \varphi(k) [y(k) - \varphi^T(k) \hat{\theta}(k-1)]$$

↪ $R(k)$ wird vielfach invertiert \Rightarrow problematisch

$$\Rightarrow P(k) = R^{-1}(k) = P^{-1}(k-1) + \alpha_k \varphi(k) \varphi^T(k)$$

$$\text{nach Inversionslemma: } (A + BC)^{-1} = A^{-1} - A^{-1}B[A^{-1}B + C^{-1}]^{-1}A^{-1}$$

$$\Rightarrow P(k) = P(k-1) - \frac{P(k-1) \varphi(k) \varphi^T(k) P(k-1)}{\alpha_k + \varphi^T(k) P(k-1) \varphi(k)}$$

$$\Rightarrow \hat{\theta}(k) = \hat{\theta}(k-1) + \underbrace{P(k) \alpha_k \varphi(k)}_{L(k) = \frac{P(k-1) \varphi(k)}{\alpha_k + \varphi^T(k) P(k-1) \varphi(k)}} [y(k) - \varphi^T(k) \hat{\theta}(k-1)]$$

> Initialwerte

↪ Anfangsparameter $\hat{\theta}(k_0)$ und deren Kovarianzmatrix $P(k_0)$
 wird durch Least-Squares mit den ersten I -Messungen geschätzt
 $\rightarrow \dim(\Theta)$

↪ Andere Möglichkeit: $\hat{\theta}(k_0) = 0$ und $P(k_0) = C I$, $C \in \mathbb{R}$
 und I die Identitätsmatrix. Größe C impliziert ein schlechtes
 Vertrauen \Rightarrow stärkerer Korrektur \Rightarrow schnellerer Konvergenz
 \Rightarrow empfindlich auf Messrauschen.

> Algorithms