Statistical Data Analysis

Dr. Jana de Wiljes

24. November 2021

Universität Potsdam

Causal relations

Model for simple linear regression

NN(w_i)

Model:

$$Y_i = f(X_i, \beta) + \epsilon_i, \quad i = 1, \ldots, n \tag{1}$$

where ϵ_i are iid with $\mathbb{E}[\epsilon_i] = 0$ and $Var(\epsilon_i) = \sigma^2$

Data: it is possible to observe realisations

$$(y_i, x_i) \quad i = 1, \ldots, n \tag{2}$$

targets

inputs

Goal: estimate parameters β of the function to obtain approximative $f(x, \hat{\beta})$

Note: note that f approximates $\boxed{\mathbb{E}[Y_i|X_i]}$

Linear regression

Model for simple linear regression

Model:

NN

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, \ldots, n \tag{3}$$

where ϵ_i are iid with $\mathbb{E}[\epsilon_i] = 0$ and $Var(\epsilon_i) = \sigma^2$

Data:

$$(y_i, x_i) \quad i = 1, \ldots, n \tag{4}$$

Goal: estimate $f(x, \hat{\beta}) = \hat{\beta}_0 + \hat{\beta}_1 x$

The Ordinary Multiple Linear Regression Model

Model:

$$Y_i = \beta_0 + \beta_1 X_i + \overset{\overset{0}{\shortparallel}}{\beta_2} X_{i,2} + \overset{\overset{0}{\shortparallel}}{\beta_3} X_{i,3} + \cdots + \overset{\overset{0}{\shortparallel}}{\beta_p} X_{i,p} + \epsilon_i, \quad i = 1, \ldots, n \tag{5}$$

where ϵ_i are iid with $\mathbb{E}[\epsilon_i] = 0$ and $Var(\epsilon_i) = \sigma^2$

Data:

$$(y_i, x_i) \quad i = 1, \ldots, n \tag{6}$$

Goal: estimate $\hat{f}(\underbrace{x_1, \ldots, x_p, \hat{\beta}_1, \ldots, \hat{\beta}_p}) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots, \hat{\beta}_p x_p$

Multivariate Random Variables

Def: Let $\mathbf{X}$ be a vector of (univariate) random variables, i.e.,
X $\mathbf{X} = (X_1, \ldots, X_p)^\top$ with $\mathbb{E}[X_i] = \mu_i$. $\mathbf{X}$ is called a multivariate random variable and we denote $\mathbb{E}[\mathbf{X}] = \mu$

Note:

- Variance
$$\mathrm{Var}(X_i) = \mathbb{E}[(X_i - \mathbb{E}(X_i))^2] = \mathbb{E}[(X_i - \mathbb{E}(X_i))(X_i - \mathbb{E}(X_i))]$$
- Covariance $\mathrm{Cov}(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))]$

Covariance

Def: The covariance of the multivariate random variable $\mathbf{X}$ is defined by

$$\Sigma := \mathrm{Cov}(\mathbf{X}) = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^\top] \tag{7}$$

Example:

$\mathcal{N}(\mu, \Sigma)$

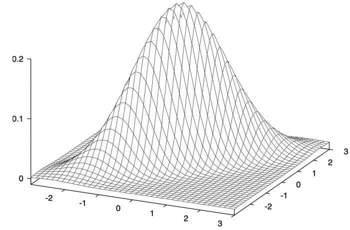
$$\Sigma = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) \end{pmatrix} \tag{8}$$

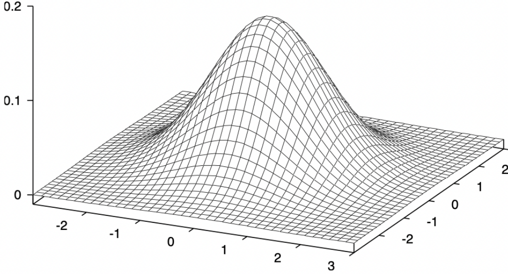
Properties of Σ:

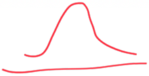
- quadractic
- symmetric
- positive-semidefinite

Recall: a matrix A is positive-semidefinite

$\Leftrightarrow \boxed{x^T A x \geq 0} \quad \forall x \neq 0$







$$\mathbf{X} \sim \mathcal{N}_p(\mu, \Sigma) \tag{9}$$

LS-estimation

Least Squares Estimation: minimize the sum of squared errors

$$LS(\beta) = \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon^T \varepsilon$$

with respect to $\beta \in \mathbb{R}^{p+1}$

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon^T \varepsilon = (y - X\beta)^T (y - X\beta)$$

$$\underbrace{Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}}_{\in \mathbb{R}^{n \times 1}}, \quad \underbrace{X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & & & \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}}_{\in \mathbb{R}^{n \times p+1}}, \quad \underbrace{\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}}_{\in \mathbb{R}^{p+1 \times 1}}, \quad \underbrace{\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{\in \mathbb{R}^{n \times 1}}$$

Positive semi-definite

Lemma: Let $\mathbf{B}$ be an $n \times (p+1)$ matrix. Then the matrix $\mathbf{B}^\top \mathbf{B}$ is symmetric and postive semi-definite. It is positive definite, if $\mathbf{B}$ has full column rank. Then, besides $\mathbf{B}^\top \mathbf{B}$ also $\mathbf{B}\mathbf{B}^\top$ is postive semi-definite.

LS-estimator

Theorem: The LS-estimator of the unknown parameters β is

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \tag{10}$$

if $\mathbf{X}$ has full column rank p+1.

Proof

$$LS(\beta) = \varepsilon^T \varepsilon = \underbrace{(y - X\beta)^T (y - X\beta)}_{} = y^T y - 2\beta^T X^T y + \beta^T X^T X \beta$$

$$\frac{\partial LS(\beta)}{\partial \beta} = -2X^T y + 2X^T X \beta \overset{!}{=} 0$$

$$\frac{\partial LS(\hat\beta)}{\partial \beta} = 0 \iff X^T X \hat\beta = X^T y$$

$$\iff \hat\beta = (X^T X)^{-1} X^T y$$

$$\frac{\partial^2 LS(\beta)}{\partial \beta \, \partial \beta^T} = 2X^T X$$

because we know that $X^T X$ is positive-semi definite

$\rightsquigarrow$ $\hat\beta$ is in fact a minimum □

Proof

$$\mathbb{E}[\hat{Y} \mid X] = \hat{y} = X \underline{\hat{\beta}}$$

$$\hat{y} = \underbrace{X (X^T X)^{-1} X^T}_{H :=} y = H y$$

„prediction matrix"

„hat matrix"

Positive semi-definite

Proposition: The hat-matrix $\mathbf{H} = (h_{ij})_{1 \leq i,j, \leq b}$ has the following properties:

1. $\mathbf{H}$ is symmetric
2. $\mathbf{H}$ is idempotent, i.e., $\mathbf{HH} = \mathbf{H}$
3. $rk(\mathbf{H}) = tr(\mathbf{H}) = p + 1$
4. $0 \leq h_{ii} \leq 1, \quad \forall i = 1, \dots, n$
5. the matrix $\mathbf{I}_n - \mathbf{H}$ is also symmetric and idempotent with $rk(\mathbf{I}_n - \mathbf{H}) = n - p - 1$

Theorem: The ML-estimator of the unknown parameters σ^2 is $\hat{\sigma}^2_{ML} = \frac{\hat{\epsilon}\hat{\epsilon}}{n}$ with $\hat{\epsilon} = \mathbf{y} - \mathbf{X}\underbrace{\hat{\beta}}_{\hat{y}}$.

$\epsilon \sim N(0, \underbrace{\sigma^2 I_N}_{\Sigma})$

y

$\epsilon = y - X\beta$

$y \sim N(X\beta, \sigma^2 I_n)$

$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)\right)$

log-likelihood

$\ell(\beta, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2)$

$-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)$

Proof

$$\boxed{\frac{\partial \ell(\beta, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - X\beta)^T(y - X\beta) \overset{!}{=} 0}$$

Plugging in $\hat{\beta}$ for β

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - \underbrace{X\hat{\beta}}_{\hat{y}})^T(y - \underbrace{X\hat{\beta}}_{\hat{y}}) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\hat{\varepsilon}^T\hat{\varepsilon} \overset{!}{=} 0$$

$\sigma^2 \neq 0$ we hence obtain $\boxed{\hat{\sigma}^2_{ML} = \frac{\hat{\varepsilon}^T\hat{\varepsilon}}{n}}$

$\square$

ML-estimator

Proposition: For the ML-estimator $\hat{\sigma}^2_{ML}$ of σ^2 the following property holds:

$$\mathbb{E}[\sigma^2_{ML}] = \frac{n-p-1}{n}\sigma^2 \tag{11}$$

$\mathbb{E}[x^2 a]$

Proof: $\mathbb{E}[\hat{\varepsilon}^T\hat{\varepsilon}] = \mathbb{E}[(y - X\hat{\beta})^T(y - X\hat{\beta})]$

$= \mathbb{E}[(y - \underbrace{X(X^TX)^{-1}X^T}_{H}y)(y - X(X^TX)^{-1}X^Ty)]$

$= \mathbb{E}[(y - Hy)^T(y - Hy)]$

$= \mathbb{E}[y^T(I_n - H)^T(I_n - H)y]$

$= tr((I_n - H)\cdot\sigma^2 I_n) + \beta^T\cancel{X}^T(I_n - H)X\beta$

$\mathbb{E}[x^TAx] = tr(A\Sigma) + \boxed{\mu^TA\mu}$

$Cov[x] = \Sigma$

$\mathbb{E}[x] = \mu$

Proof

$$= \sigma^2 (n-p-1) + \beta^T X^T (I_n - X(X^T X)^{-1} \underbrace{X^T) X}\beta$$

$$= \sigma^2 (n-p-1) + \beta^T X^T X \beta - \beta^T X^T X \underbrace{(X^T X)^{-1} X^T X}\beta$$

$$= \sigma^2 (n-p-1) + \underline{\beta^T X^T X \beta} - \underline{\beta^T X^T X \beta}$$

$$= \sigma^2 (n-p-1)$$

$\frac{1}{n}$ $\frac{\varepsilon^T \cdot \varepsilon}{n}$

Adjusted estimator

Proposition: The adjusted estimator

$$\hat{\sigma}^2_{ad} = \frac{\hat{\epsilon}\hat{\epsilon}}{n-p-1} \tag{12}$$

of the unknown parameter σ^2 can be written as

$$\hat{\sigma}^2_{ad} = \frac{\mathbf{y}^\top\mathbf{y} - \hat{\beta}^\top\mathbf{X}^\top\mathbf{y}}{n-p-1} \tag{13}$$

ML estimator

Proposition: The LS-estimator $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}$ is equivalent to the ML-estimator based on maximization of the log-likelihood

$$l(\beta, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\underbrace{(\mathbf{y} - \mathbf{X}\beta)^\top(\mathbf{y} - \mathbf{X}\beta)}_{\text{need to minimize with respect to } \beta} \quad (14)$$

$\hat{\beta}$

$L(\sigma^2) = \int L(\beta, \sigma^2)\, d\beta$

LS estimator

Proposition: The LS-estimator $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ and the REML-estimator $\hat{\sigma}^2 = \frac{1}{n-p-1} \hat{\epsilon}^\top \hat{\epsilon}$ the following properties hold:

1. $\mathbb{E}[\hat{\beta}] = \beta$, $\mathrm{Cov}(\hat{\beta}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$
2. $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$