Parameter Estimation

data from experiment $\tilde{y}_i, \underline{x}_i$ $1 \leq i \leq n$ $y = f(\underline{x}, \underline{p})$ known functional form

dependent independent

$$\tilde{y}_i = f(\underline{x}_i, \underline{p}) + \varepsilon_i$$

1.) select function

2.) select statistical modell,
e.g, $\underline{\tilde{x}}_i$ - implicit/explicit
ε_i - error modell
covariance of $(\varepsilon_i, \varepsilon_j)$
estimator

Mostly:

- $\underline{x}_i$: independent, known without errors
- $E(\varepsilon_i) = 0$ unbiased
- ε_i known or can be estimated
- $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$, if $i \neq j$
- Least square estimator

$$Q(\underline{p}) = \sum_{i=1}^{n} (\tilde{y}_i - f(\underline{x}_i, \underline{p}))^2 \cdot w_i \qquad \overset{\text{weight}}{w_i} = 1/\sigma_i^2$$

Theoretically: these + ε_i normally distributed + using maximum likelihood estimation } that is the correct estimator to get unbiased and efficient $\underline{p}$

residual error:

$$r_i = \tilde{y}_i - f(\underline{x}_i, \underline{p})$$

$\underline{p}$ can be described using probality theory.

Univariate linear regression

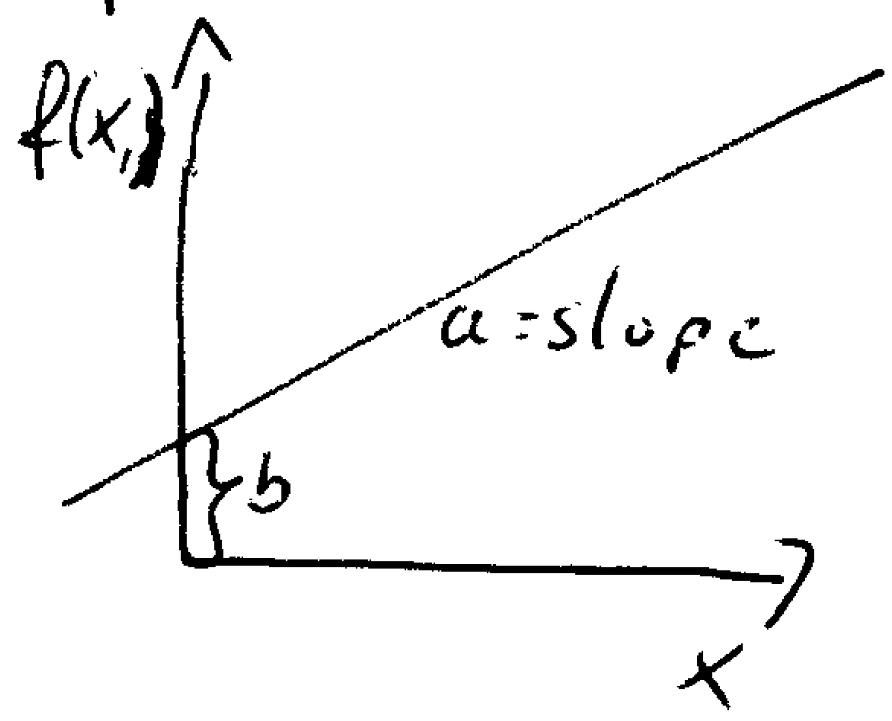
$y = ax + b$, known set of $x_i - \tilde{y}_i$, $D(\varepsilon_i) = \sigma_{\varepsilon_i} = \sigma_\varepsilon$

$$Q(a,b) = \sum_{i=1}^{n} (\tilde{y}_i - a x_i - b)^2$$ $\rightarrow$ for minima first derivates are zero:

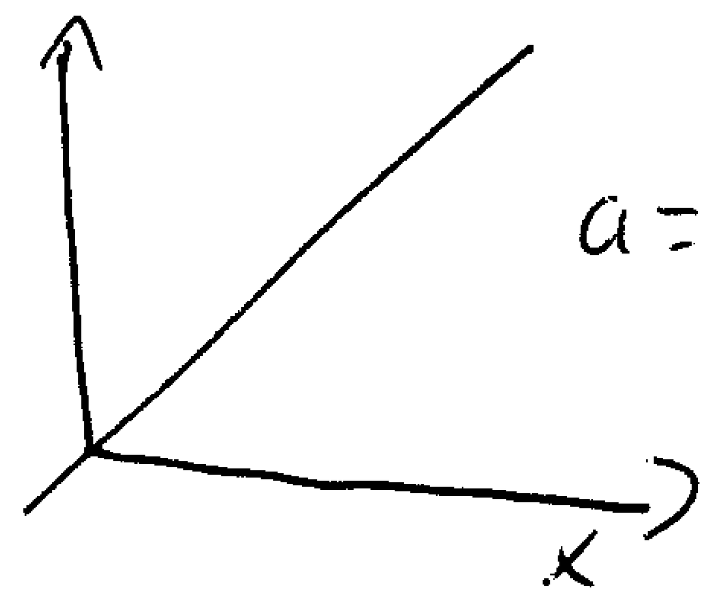
$$\frac{\partial Q(a,b)}{\partial a} = 0 \qquad \frac{\partial Q(a,b)}{\partial b} = 0$$

closed forms: $\bar{x} = \sum_{i=1}^{n} x_i / n \qquad \bar{y} = \sum_{i=1}^{n} \tilde{y}_i / n$

$$a = \frac{\sum x_i \tilde{y}_i - \bar{x}\bar{y} \cdot n}{\sum x_i^2 - \bar{x}\bar{x} \cdot n} \qquad b = \bar{y} - a\bar{x}$$



if $b = 0$



$$\left(a = \frac{\sum x_i y_i}{\sum x_i^2} \right)$$

variance of a-b through residual variances:

$$s_r^2 = \frac{\sum r_i^2}{n-2} \qquad s_a^2 = s_r^2 \frac{n}{n\sum x_i^2 - (\sum x_i)^2}$$

$$s_b^2 = s_r^2 \cdot \frac{\sum x_i^2}{n\sum x_i^2 - (\sum x_i)^2} \qquad a \pm s_a \cdot t^{-1}(\underset{n-2}{\nu})$$

confidence intervals can be given

Multivariate linear regression

$f(\underline{x}, \underline{p}) = p_1 x_1 + p_2 x_2 + p_3 x_3 + \ldots p_m x_m$

n measurements $\tilde{y}_i - \underline{x}_i$

$$\underset{\text{weight matrix}}{W} = \begin{pmatrix} w_1 & & \emptyset \\ & \ddots & \\ \emptyset & & w_n \end{pmatrix} \qquad \underset{\text{predictor matrix}}{X} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nm} \end{pmatrix}$$

$$\underline{\tilde{y}} = \begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ y_3 \end{pmatrix} \text{ result vector} \qquad \underline{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\underline{p} = \begin{pmatrix} p_1 \\ \vdots \\ p_m \end{pmatrix} \text{ parameter vector}$$

$$Q(\underline{p}) = (\underline{y} - X\underline{p})^T W (\underline{y} - X\underline{p})$$

$$\frac{\partial Q(\underline{p})}{\partial p_i} = 0 \ldots \quad \rightarrow \text{„normal equations"} \rightarrow$$

$$\rightarrow \quad \underline{p} = (X^T W X)^{-1} \cdot X^T \cdot W \underline{y} \quad \text{in one step}$$

Connection to overdetermined set of linear equations:
(easy way to memorize and to derive)

$$A \in \mathbb{R}^{n\times m} \quad \underline{b} \in \mathbb{R}^{n} \quad \underline{x} \in \mathbb{R}^{m}, \quad m<n$$

$$A\underline{x} \approx \underline{b}$$

$$A^T A \underline{x} = A^T \underline{b}$$

$$\underbrace{(A^T A)^{-1} A^T A}_{=E} \underline{x} = (A^T A)^{-1} A^T \underline{b} \Rightarrow \underline{x} = (A^T A)^{-1} A^T \underline{b}$$

but this is often singular → Singular Value Decom… → Ridge

Mutivariate + nonlinear parameter estimation

$y = f(\underline{X}, \underline{p})$, but f non-linear function:

$$\tilde{y}_i = f(\underline{x}_i, \underline{p}) + \varepsilon_i$$

$$\underline{f}(\underline{X}, \underline{p}) = \begin{pmatrix} f(\underline{x}_1, \underline{p}) \\ f(\underline{x}_2, \underline{p}) \\ \vdots \\ f(\underline{x}_n, \underline{p}) \end{pmatrix}$$

$$Q(\underline{p}) = (\underline{y} - \underline{f}(X, \underline{p}))^T W \cdot (\underline{y} - \underline{f}(X, \underline{p}))$$

Gauss-Newton method:

Taylor expansion around $\underline{p}^k$, as variable:

$$\underline{f}(\underline{p}^{k+1}) \approx \underline{f}(\underline{p}^k) + J(\underline{p}^k)(\underline{p}^{k+1} - \underline{p}^k)$$

$$J(\underline{p}^k) = \begin{pmatrix} \frac{\partial f(\underline{x}_1, \underline{p}^k)}{\partial p_1} & \cdots & \frac{}{\partial p_m} \\ \vdots & & \\ \frac{\partial f(\underline{x}_n, \underline{p}^k)}{\partial p_1} & \cdots & \frac{\partial f(\underline{x}_n, \underline{p})}{\partial p_m} \end{pmatrix}$$

Jacobian

$$\Downarrow$$

$$Q(\underline{p}) = (\underline{y} - \underline{f}(\underline{p}^k) - J(\underline{p}^k)\cdot(\underline{p}^{k+1} - \underline{p}^k))^T W (\underline{y} - \underline{f}(\underline{p}^k) - J(\underline{p}^k)(\underline{p}^{k+1} - \underline{p}^k))$$

$$\Downarrow$$

$$\underline{p}^{k+1} = \underline{p}^k + (J^T W J)^{-1} \cdot J^T W \cdot (\underline{y} - \underline{f}) \quad \text{iteratively!}$$

Not easy to invert $(J^T W J)$, often quasi-singular!

In physics: Tikhonov regularization

e.g. to minimize

$$\|A\underline{x}-\underline{b}\|^2 \rightarrow \text{add to } \underbrace{\|A\underline{x}-\underline{b}\|^2 + \overbrace{\|\Gamma x\|^2}^{\text{regularization part}}}_{\text{search for this minima}}$$

for overdetermined $A\underline{x} \approx \underline{b}$

$$\underline{x} = (A^T A)^{-1} A^T \underline{b} \Rightarrow \tilde{\underline{x}} = (A^T A + \Gamma^T \Gamma)^{-1} A^T \underline{b}$$

here, Ridge regression, if $\Gamma^T \Gamma = \lambda \cdot I$ $\quad \lambda \geq 0 \quad \lambda \in \mathbb{R}$

In parameter estimation it is called as

Marquardt method.

$$(J^T W J + \lambda \cdot \underline{I})^{-1} \qquad \lambda^k \geq 0$$

how to change λ^k, e.g. Levenberg

if $Q^{k+1} > Q^k \Rightarrow \lambda^{k+1} = 10 \cdot \lambda^k$

$\quad\; < \qquad\quad \lambda^{k+1} = \lambda^k / 10$

"Gauss-Newton-Marquardt-Levenberg" method