Teil A. Mathematische Grundlagen

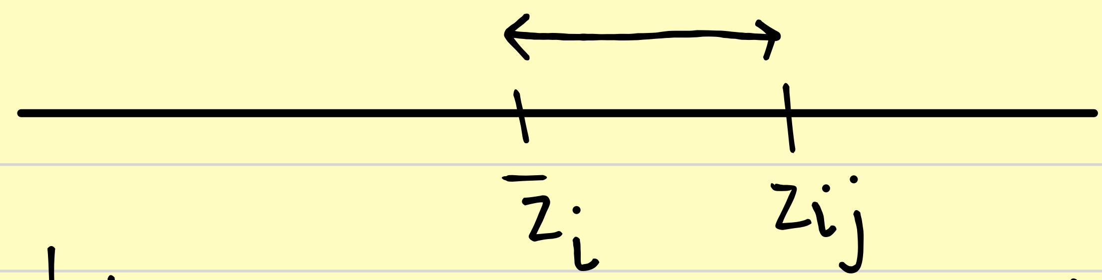
A1. Bausteine & Notation

- Zahl. $a \in \mathbb{R}$ z.B.: $a = 7{,}13$
- Liste (Vektor). eine geordnete Liste von Zahlen
  z.B.: $\vec{v} = [1, 3{,}7, 2{,}3, -1{,}3, \dots]$
- Tabelle (Matrix). eine rechteckige Anordnung von Zahlen
  Jede Zeile = eine Beobachtung
  Jede Spalte = ein Merkmal.

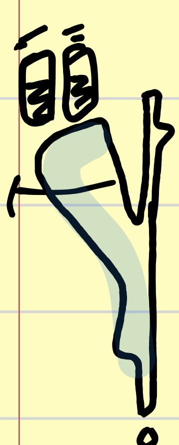
$$A = \begin{bmatrix} a_{11} & a_{21} & a_{31} & \dots & a_{n1} \\ a_{12} & a_{22} & & & \\ a_{13} & & a_{37} & & \\ \vdots & & & \ddots & \\ a_{1n} & & & & a_{nn} \end{bmatrix} \rightarrow \text{Beobachtungen (Zeit)}$$

$\downarrow$
Merkmalen (KPIs)

- Indexierung: das Elemen a_{ij} bedeutet dann die Beobachtung „j“ und die Merkmal „i“.
- Mittelwert (Durchschnitt): $\bar{z}_i = \frac{1}{n} \sum_{j=1}^{n} z_{ij}$
  Mittelwert Merkmal „i“.
- Abweichung vom Durchschnitt: $z_{ij} - \bar{z}_i$
  Abstand zum Mittelwert.

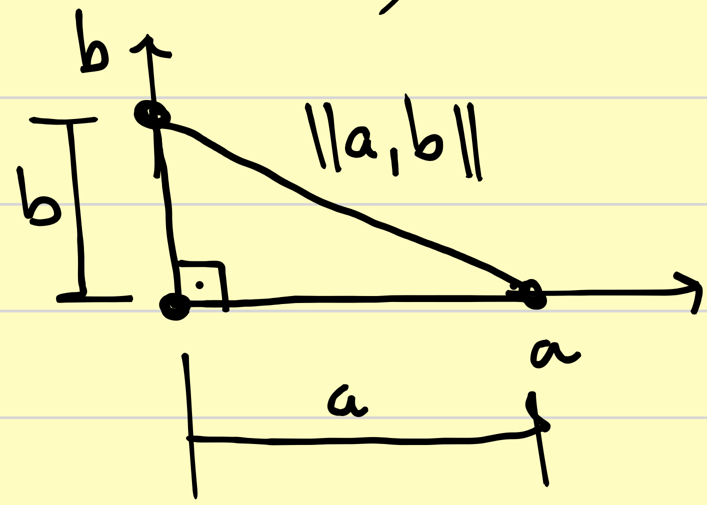


- Stichprobenvarianz. $VAR(z_i) = \frac{1}{n} \sum_{i=1}^{n} (z_{ij} - \bar{z}_i)^2$
- Standard Abweichung. $\sigma(z_i) = \sqrt{VAR(z_i)}$
- Kovarianz. $cov(x,y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$
- Korrelation. $corr(x,y) = \frac{cov(x,y)}{\sigma(x)\sigma(y)} \in [-1,1]$



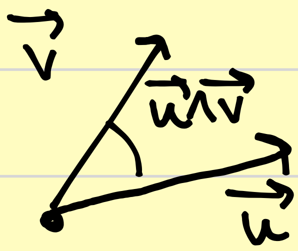
- Standardisierung (Z-Wert): $z_i = \frac{x_i - \bar{x}}{\sigma(x)}$
- Vektor Länge (euklidische Norm)

$$\|a,b\| = \sqrt{a^2 + b^2}$$



- Skalar Produkt (Innenprodukt)

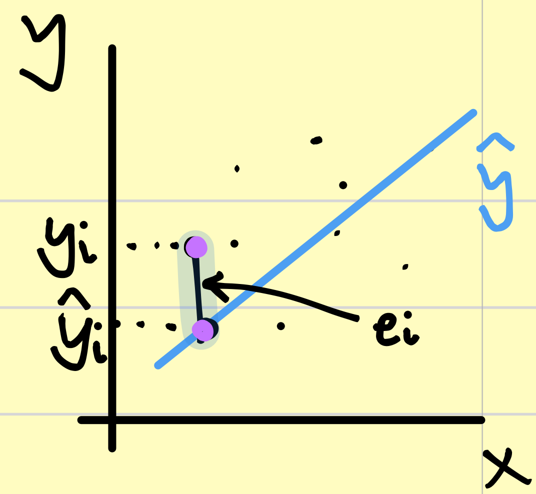
-für gleich lange Listen-



$$\vec{u} \cdot \vec{v} = \sum_j u_j v_j = \|\vec{u}\| \|\vec{v}\| \cdot \cos(\vec{u} \wedge \vec{v})$$

Interpretation: großer Skalarprodukt = ähnliche Richtung

A2. MSE Fehler.



- Lineares 1D-Modell: $\hat{y} = a + bx$
- Vorhersagefehler $e_i = y_i - \hat{y}_i =$

  $= y_i - a - bx_i$
- Quadratischer Fehler eines Punkts: $e_i^2 \geq 0$.
- Mittlerer Quadratischer Fehler
  Mean Squared Error (MSE)

$$\boxed{MSE(a,b) = \frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - a - bx_i)^2}$$

A3. Ableitung/Gradient.

- Ableitung ≡ Steigung einer Funktion. Wenn wir a oder b ein wenig verändern und die MSE dadurch steigt, dann ist die Veränderung nicht in die erwünschte Richtung!
- Elementare Ableitungsregeln:

$$\frac{d}{dz}(z^n) = n z^{n-1}$$

- Gradient (bei zwei Variabeln a, b) ≡ Vektor der Ableitungen $\left[\frac{\partial MSE}{\partial a}, \frac{\partial MSE}{\partial b}\right]$

- Gradient descent (GD): Starten mit a, b Werte (Initialisierung), gehen in kleinen Schritten entgegen dem Gradienten.

η = Lernrate

$$a \to a - \eta \frac{\partial MSE}{\partial a}$$

$$b \to b - \eta \frac{\partial MSE}{\partial b}$$

(Schau Lineare Regression vom 2. Semester).

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \qquad a = \bar{y} - b\bar{x}$$

A4. Mehrere Merkmale (KPIs)

- Wenn wir mehrere Merkmalen $x_1, \dots, x_d$ haben:

$$\hat{y} = a + w_1 x_1 + w_2 x_2 + \dots + w_d x_d$$

- Designmatrix: wir bauen aus den Daten eine Tabelle X mit "n" Zeilen (Beobachtungen) und "d" Spalten (Merkmale) und fügen links eine 1-Spalte für "a" hinzu.

- LSP: Least Square Problem: "Finde $a, w_1, \dots, w_d$ die, die MSE minimieren".

Praktisch: lösen wir das numerisch mit einem eingebauten L.S. Solver

A6. Train / Test Aufteilung.

Daten werden fürs „ML" zufällig gemischt, in Trainings- & Testmengen aufgeteilt (zB 75%, 25%)

1000 Bilder → 750 Training → GD $[a, w_i]$

1000 Bilder → 250 Test → Accuracy (%)

A7. Beispiel:

$x = [30, 40, 50, 60, 70]$ $\quad y = [2, 3, 4, 5, 6]$

- $\bar{x} = 50$; $\bar{y} = 4$
- $x - \bar{x} = [-20, -10, 0, 10, 20]$; $y - \bar{y} = [-2, -1, 0, 1, 2]$
- KOVARIANZ $= \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} = \dfrac{90}{5} = 18$
- VARIANZ $\quad VAR(x) = \dfrac{400 + 100 + 0 + 100 + 400}{5} = 200$
- $b = \dfrac{18}{200} = 0'09$
- $a = \bar{y} - b\bar{x} = -0'5$

$\hat{y} = -0'5 + 0'09\,x$