

Übungsblatt 5

Alexander Mattick Kennung: qi69dube

Kapitel 2

24. April 2020

Fragen:

1 Streuung $\frac{1}{n}$ vs empirische Streuung $\frac{1}{n-1}$

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \text{ vs } \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Wahrscheinlichkeitsverteilungen

→ Erwartungswert

→ Varianz

Zum ermitteln dieser muss i.a. Stichproben aus der Echten Verteilung gezogen werden.

Für mehr Stichproben n konvergiert die empirische Varianz besser gegen die echte Varianz!

Die $\frac{1}{n-1}$ ist immer größer als $\frac{1}{n}$!

Das ist eine **pessimistischere Schätzung der Varianz**, woraus die ungewissenheit besser gehandelt werden kann.

Stichprobe → Bild von der Grundverteilung!

→ Wie gut ist das Bild das wir gesammelt haben?

2 Rangwert vs. Ordnungsstatistik.

Rangliste schmeißt alle doppelten Werte raus, Ordnungsstatistik behält diese.

Ordnungsstatistik: $\{1, 2, 2, 2, 3, 4\}$ Rangwert $\{1, 2, 3, 4\}$

Der Rangwert kommt aus der Ordnungsstatistik mit $r + \frac{s-1}{2}$, wobei s =Anzahl der Werte

3 Mittelwerte

Arithmetisches mittel oder Durchschnitt $\bar{x} = \frac{1}{n} \sum x_i$

Harmonisches mittel $\frac{1}{n} (\sum \frac{1}{x_i})^{-1}$ für z.B. geschwindigkeiten und verhältnisse dieser. Wenn man durchschnittliche Raten haben. Es betrachtet die geschwindigkeit der Veränderung.

geometrisches Mittel $\sqrt[n]{\prod_{i=1}^n x_i}$ Für exponentielle zunahmen von z.B. Kontostand und prozentuale Zunahme. Verhältnisse beleiben vorhanden.

4 Kovarianz

Wenn die Kovarianz null ist, dann sind die Werte vollkommen unabhängig.

Die umkehrung gilt nicht, außer bei multinormalen Verteilungen

5 Wie sollte man Klassen aufteilen?

Gibt kein allheilmittel (mein Vorschlag: aufteilen, dass die Varianz innerhalb jeder Klasse grad kleiner ist, als die der Gesamtvarianz)

6 Korrelation

Wenn der korrelationskoeff $r_{xy} = \frac{s_{xy}}{s_x s_y}$ nahe an 1 ist, dann ist die korrelation optimal, bei -1 ist sie invers-optimal.

Bestimmtheitsgrad ist das Verhältniss des zweiten moments normalisiert um den mittelwert der Echten Daten:

$$B = \frac{\sum(\hat{y} - \bar{y})}{\sum(y - \bar{y})}$$

7 Lineares Ausgleichsproblem

Anpassen einer minimalen funktion (z.b. polynom von grad-m $m \ll n$)

$$\min \sum_{j=1}^n (y_j - (p_m x_j^m + \dots + p_0))$$

Dies liefert eine

$$\begin{pmatrix} x_1 & x_1^2 & \dots & x_1^m \\ \dots & \dots & \dots & \dots \\ x_n & x_n^2 & \dots & x_n^m \end{pmatrix}$$

Also eine $N \times M$ matrix.

Dies liefert $\nabla_p z(p) = A^T A p - A_y^T \stackrel{=}{!}$ Für optimalität zweiter Ordnung: $Hf(x)$ muss positiv definit sein, für ein absolutes min!

Das $H_z(p) = A^T A$

$A^T A$ ist immer symmetrisch! Also auch positiv (semi)-definit!

Für den einfachen Fall von 1d-(x,y):

$$\min \frac{1}{2} \sum_{k=1}^n (ax_i + n - y_i)^2 \text{ minimiert } a = \frac{s_{xy}}{(s_x)^2} \text{ und } b = \bar{y} - a\bar{x}$$

Diese Funktion.