

# Long · Short · Term · Memory (LSTM).

## und weitere ML · Methoden für Zeitreihenanalyse.

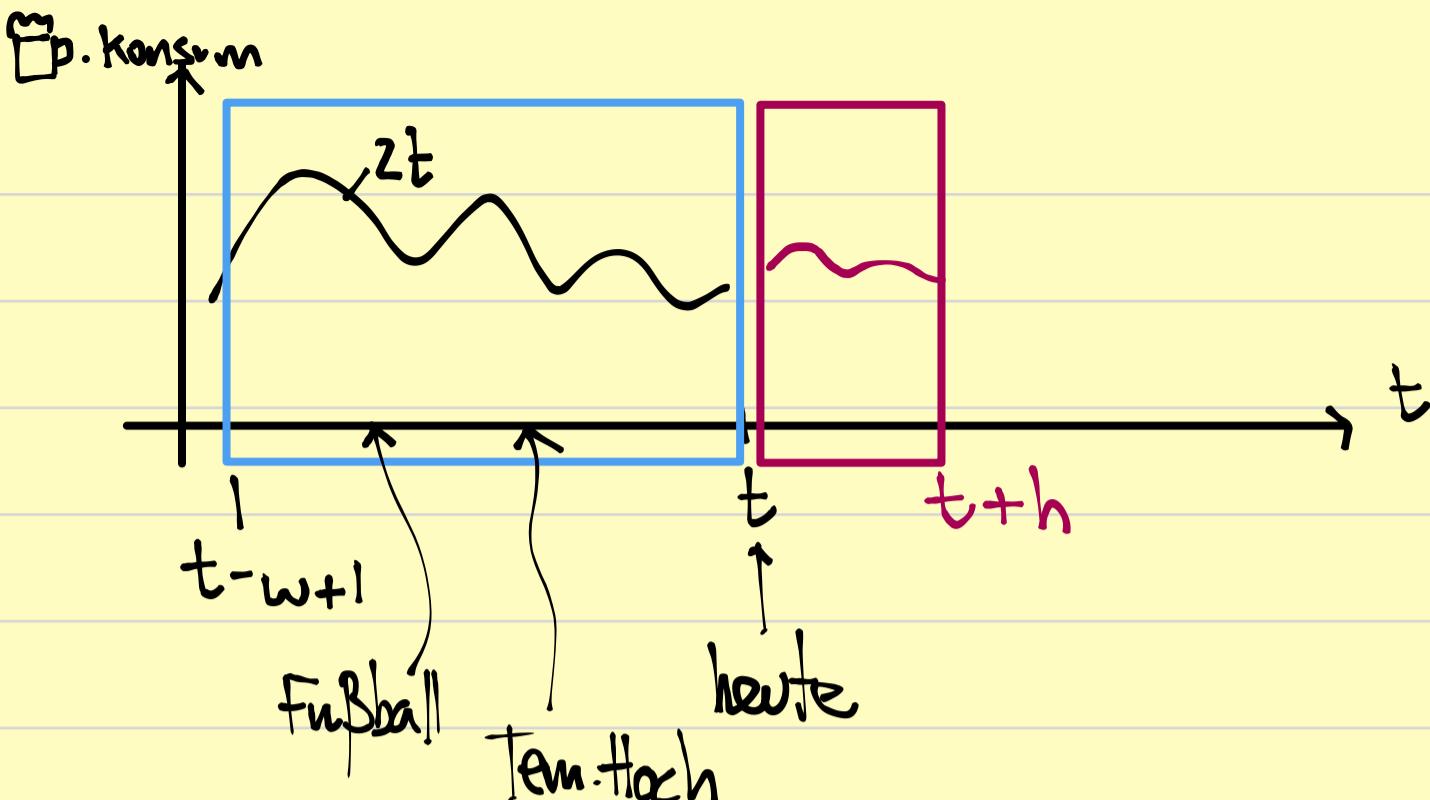
Wir lernen, Zeitreihen als überwachte Lernaufgabe aufzubereiten, korrekt & kausal zu skalieren, solide Baselines einzusetzen und dann LSTM / GRU (Gated Recurrent Units) sowie kausale 1D · CNN einzusetzen.  
 Beispiel: Umsatzprognose (Tage) mit „Vergangenheit → Zukunft“ Muster.

### A1. Begriffe

- Zeitreihe: Geordnete Werte  $y_1, y_2, \dots, y_n$  in festen Abständen
- Horizont  $h$ : Wie weit in die Zukunft wir voraussagen.
- Mehrhorizont ..  $H$ : Vektor  $(\hat{y}_{t+1}, \dots, \hat{y}_{t+H})$ .
- Exogene Größen  $x_t$ : weitere, zum Zeitpunkt  $t$  bekannte Prädiktoren (i.e. Ereignis, Wetter, ...)
- Intuition: Prognose ist eine Abbildung  
 Vergangenheitsfenster + Kontext  $\rightarrow$  Zukunft

### A2. Von der Reihe zum Lernbeispiel (Sliding Window, Strategien)

- Fensterung: Aus einem Vergangenheitsfenster der Länge  $w$ :  $z_t = [y_{t-w+1}, \dots, y_t]$  + exogene Information  $x_{t-w+1}$ : t lernen wir eine Abbildung  $f: z_t \rightarrow y_{t+h}$



Strategien:

(1) Rekursiv. Trainiere auf  $h=1$ , benutze  $\hat{y}_{t+1}$  als neues lag für  $t+2$  usw.  
EINFACHES MODELL, ABER FEHLER KUMMULIEREN SICH.

(2) Direkt. Für jedes  $h$  wird ein separates Modell  $f_h$  gebaut.  
STABIL, ABER AUFWENDIG.

(3) Mehr.ziel.Direkt. Ein Modell gibt  $(\hat{y}_{t+1}, \dots, \hat{y}_{t+h})$  auf einmal aus.  
HORIZONTELERNEN GEMEINSAM.

Intuition: Rekursiv ist leicht, aber ... Schneeball-Effekt.  
Direkt/Mehrziel ist robuster für lange Horizonte.

Praxis: wähle  $w$  mindestens so groß wie die wichtigste Periode (i.e.  $w > 7$  bei Wochentypen)

## A3. Skalierung, Kausalität & Feature-Design.

- kausale Fenster: NIEMALS ZU VENFTIGE WERTE IN FEATURES MISCHEN.
- Skalierung: Min/Max nur für Training fitten und dieselben min/max Werte auf VAL/TEST.  
$$x_i^* = \frac{x_i - x_{\min}^{\text{training}}}{x_{\max}^{\text{training}} - x_{\min}^{\text{training}}} \rightarrow x_i^* = \frac{x_i - x_{\min}^{\text{training}}}{x_{\max}^{\text{training}} - x_{\min}^{\text{training}}} \quad \begin{matrix} \text{TRAINING} \\ \text{Test/Val} \end{matrix}$$
- Features: Zeit-Cosinus/Sinus, binäre Flags, externe Größen, ...
- Intuition: ERST SPLITTEN, DANN FITTEN

## A4. Baselines & klassische ML für Zeitreihen.

- Naiv.  $\hat{y}_{t+1} = y_t$ .
- Saisonal-Naiv. (Periode  $s$ ):  $\hat{y}_{t+1} = y_{t+1-s}$
- Ridge/Lasso
- Trees / Random Forest.

## A5. LSTM & GRU . Gedächtnis für Sequenzen (mit Intuition).

Intuition: Drei „Schalter“ steuern, was behalten, was neu speichern, was ausgeben. Wie ein Notizbuch mit Lese-/Schreib-/Lösche-Funktionen.

GRU: Vereinfachte Version (Update- / Reset-Tor)

PRAxis: Dropout, Gradient-clipping, Fensterlängen, ... sind wichtige Hebel.

Wann LSTM? Wenn Abhängigkeiten länger sind als das, was ein flacher Layer-Ansatz erfassen kann.

$$\text{LSTM: } i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (\text{Input-Tor})$$

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (\text{Vergessenes Tor})$$

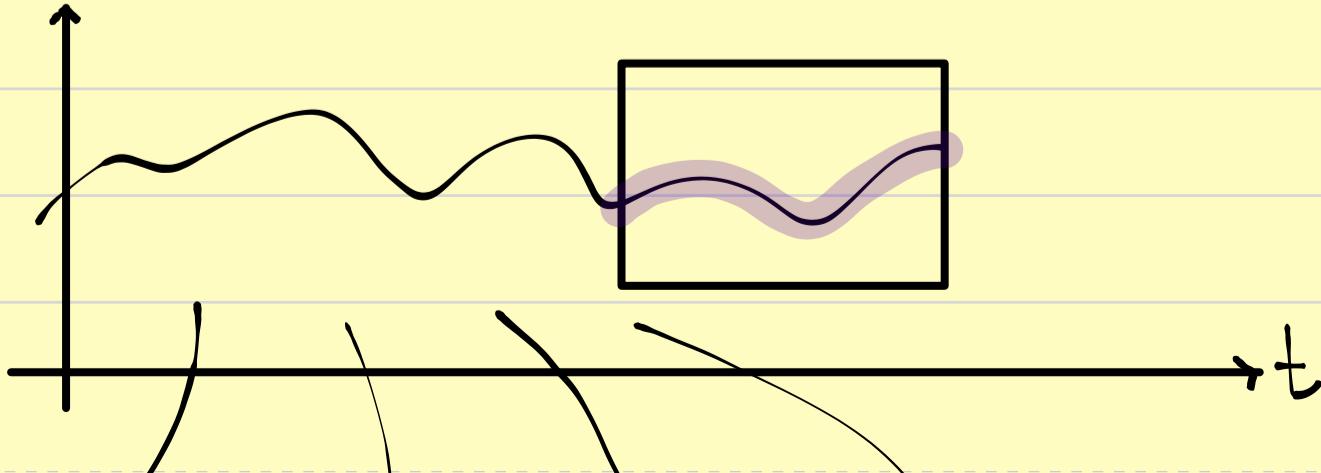
$$\bar{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (\text{Kandidat})$$

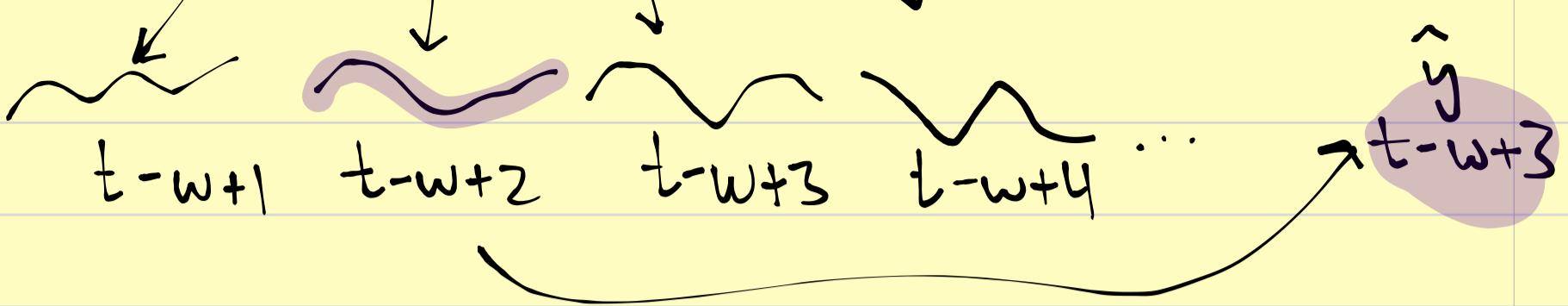
$$c_t = f_t \odot c_{t-1} + i_t \odot \bar{c}_t \quad (\text{Update})$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (\text{Output-Tor})$$

$$h_t = o_t \odot \tanh(c_t)$$

• Konsum





A6. Kausale 1D-CNN (TCN) - großer Faktor, stabiler Gradient

$$\text{Kausale Faltung: } y_t = \sum_{k=0}^{k-1} w_k \cdot x_{t-d-k}$$

$d$ : Abstand im Kernel

### A7. Praxis Check-liste.

- 1) Ziel: Horizonte & Prognose sauber definieren.
- 2) Exogene Merkmale kausal bauen.
- 3) Erstmal baselines (naiiv, s.naiiv, ...), erst dann LSTM/TCN!
- 4) Erfolg messen.

