

Universität Dortmund SoSe 2004  
Übungen zu *Wissensentdeckung in Datenbanken*

Blatt 9. Abgabe bis Montag, den 28.6.2004

Als Lösungen der Aufgaben werden die vollständigen Rechenwege erwartet. (Die Lösung muss nicht per Email eingereicht werden, sondern kann auch auf Papier in unseren Büros in GB IV, R. 119/120 abgegeben werden.)

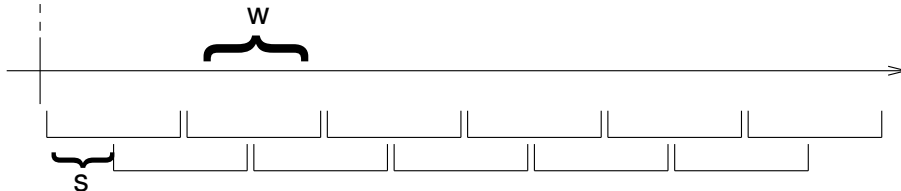
**Aufgabe 1** Bei der Analyse zeitlicher Daten kann die Abtastrate, also die Häufigkeit der Messungen, mit der die Zeitreihenwerte gewonnen wurden, eine Rolle spielen. Ein Signal werde durch die Funktion  $\sin(\frac{1}{8}\pi t)$  erzeugt und zu den Zeitpunkten  $t = 0, 1, 2, \dots$  gemessen. Finden Sie eine Funktion mit längerer Periode, die zu diesen Messzeitpunkten die gleichen Messwerte erzeugt (aber dazwischen andere Werte annimmt). Hinweis: Es gilt  $\sin(x) = \sin(2\pi + x)$ .

**Aufgabe 2** In der Vorlesung wurden parametrisierte Modelle zur Modellierung von Zeitreihen vorgestellt. Wie findet man jedoch geeignete Parameterwerte für ein Modell bei einer gegebenen Zeitreihe? Eine einfache Methode, die sogenannte *Kleinste-Quadrate-Methode*, minimiert die Summe der quadrierten punktwisen Abstände des Modells zur Zeitreihe. Sei  $m_t$  der Wert, den das Modell zum Zeitpunkt  $t$  liefert, so wird also der Wert  $Q = \sum_{t=1}^T (y_t - m_t)^2$  minimiert. Dazu wird  $Q$  partiell nach jedem Parameterwert abgeleitet und die Ableitung gleich Null gesetzt, so dass man ein Gleichungssystem mit ebenso vielen Gleichungen wie Parameterwerten erhält. Die Lösung des Gleichungssystems liefert geeignete Parameterwerte für das Modell.

Führen Sie dieses Verfahren für das einfachste Modell des linearen Trends durch, also für  $m_t = \beta_1 + \beta_2 t$ . Geben Sie mit Hilfe des Ergebnisses eine Trendgerade für die Zeitreihe  $(y)_t = 9, 16, 16, 23, 25, 31$  an.

**Aufgabe 3** Für alle Sliding Window-Ansätze (siehe Vorlesungsskript, Teil 10, S. 21f) spielt die Effizienz der eingesetzten Verfahren eine entscheidende Rolle. In dieser Aufgabe soll der Zusammenhang zwischen dem Aufwand zur Bearbeitung eines Zeitfensters und der Bearbeitung einer kompletten Zeitreihe untersucht werden. Sei  $s$  die Schrittweite,  $w$  die

Fensterbreite und  $n$  die Anzahl der Werte in der Zeitreihe (siehe Abbildung).



Es ergibt sich dann ein Überlappungsgrad von  $g = \frac{w}{s}$ . Liegt der Überlappungsgrad unter 1, so spricht man von überlappungsfreier Fensterung.

1. Bei gegebenem  $s$ ,  $w$  und  $n$ , wie viele Fenster müssen berechnet werden?
2. Wir gehen nun davon aus, dass ein einzelnes Fenster in  $w$  Schritten bearbeitet werden kann (d.h. jeder Punkt innerhalb des Fensters wird einmal betrachtet). Zeigen Sie, dass die komplette Zeitreihe in  $gn - gw + w$  Schritten bearbeitet werden kann.
3. Gehen wir davon aus, dass ein Fenster in  $w^2$  Schritten bearbeitet werden kann. Zeigen Sie, dass die gesamte Zeitreihe in  $gnw - gw^2 + w^2$  Schritten bearbeitet werden kann.
4. Gehen wir davon aus, dass ein Fenster in  $w \log(w)$  Schritten bearbeitet werden kann. Zeigen Sie, dass die gesamte Zeitreihe in  $gn \log(w) - gw \log(w) + w \log(w)$  Schritten bearbeitet werden kann.