HS Osnabrück
IuI – Biermann

Praktikum zu Wissensbasierten Systemen

Blatt 3

Vorgesehen für die Praktikumssitzungen vom 19., 22., 26., 29. April

1. Aufgabe: In der Datei „wbsa0239s.dat“ finden Sie stundenweise berechnete Temperaturmittelwerte, die auf Messungen beruhen, die an einer Meßstelle des physikalischen Instituts der Universität Osnabrück in den Jahren 2010 bis 2014 durchgeführt wurden[1]. In dieser Aufgabe soll auf Grundlage dieser Daten ein automatischer Lernvorgang durchgeführt werden, der es ermöglichen soll, eine an einem gegebenen Tag zu einer einer gebenen Stunde herrschende Außentemperatur als “sehr kalt“, “kalt“, “mittel“, “warm“ oder “sehr warm“ einzuschätzen. Diese Einschätzung soll in Abhängigkeit von der Jahres- und der Tageszeit erfolgen. So würde man zum Beispiel eine Temperatur von 10° im Januar als warm und im Juli dagegen als kalt empfinden.

   Um eine solche tages- und stundenabhängige Einschätzung durchführen zu können, soll mit Hilfe von Regressionsverfahren aus den vorliegen Daten für jeden Tag und jede Stunde eine Normaltemperatur ermittelt werden und für eine gegebene Tagestemperatur die Abweichung von diesem Normalwert bestimmt werden.

   (a) Numerieren Sie zunächst die in der Datei enthaltenen Stunden und Temperaturwerte mit Null beginnend durch. Der t-te Temperaturwerte sei damit x_t mit $t = 0, \dots, n-1$, wobei n die Anzahl der Datensätze ist. Der Wert t selber ist die Anzahl der Stunden, die seit dem 01.01.2010, 00:00 Uhr vergangen sind.

   Vor den weiteren Berechnungen sollen die Temperaturdaten um einen möglichen Trendanteil bereinigt werden. Berechnen Sie hierfür zu den Werten x_t die Regressionsgerade $a_1 \cdot t + a_2$ und ziehen Sie die Geradenwerte von den Temperaturwerten ab:

   $$x_t \quad \to \quad x_t - (a_1 \cdot t + a_2) \qquad \text{für} \quad t = 0, \dots, n-1 \tag{0.1}$$

   Geben Sie die beiden Regressionskoeffizienten a_1 und a_2 aus, und deuten Sie diese.

   (b) Es ist zu erwarten, daß die (nunmehr trendbereinigten) Temperaturwerte x_t der Jahres- und der Tagesperiode unterliegen. Nehmen Sie daher eine periodische Regression zu den Frequenzen $f_1 = 1/(365 \cdot 24)$ sowie $f_2 = 1/24$ vor.

   Zeichnen Sie die Werte x_t gemeinsam mit der von der periodischen Regression gelieferten Funktion[2] $\mathrm{g}(t)$.

   (c) Berechnen Sie die Abweichungen der Temperaturwerte x_t von der periodischen Funktion $\mathrm{g}(t)$:

   $$e_t \quad = \quad x_t - \mathrm{g}(t) \qquad \text{für} \quad t = 0, \dots, n-1 \tag{0.2}$$

   Wie lautet die Standardabweichung der e_t? Vergewissern Sie sich, daß deren Mittelwert ungefähr Null ist.

[1] In der Datei „wbsa0239s.dat“ befinden sich vier Datenspalten: Tag, Monat, Jahr, Stunde und Temperaturmittelwert des betreffenden Tages

[2] Diese besitzt hier die Gestalt $\mathrm{g}(t) = a_1 \cos(2\pi f_1 t) + b_1 \sin(2\pi f_1 t) + a_2 \cos(2\pi f_2 t) + b_2 \sin(2\pi f_2 t)$.

(d) Prüfen Sie nach, ob die (trendbereinigten) Temperaturdaten x_t tatsächlich die Perioden $T_1 = 365$ Tage und $T_2 = 24$ Stunden besitzen. Berechnen Sie dazu das Periodogramm der x_t, zeichnen Sie es, und finden Sie die beiden größten Periodogrammwerte.

(e) Verwenden Sie die "K-Mittelwertmethode", um die Abweichungen e_t von der Normaltemperatur in die fünf Klassen "sehr kalt", "kalt", "mittel", "warm" und "sehr warm" einzuteilen.[3]

Erstellen Sie eine Funktion, der ein Datum (zwischen 2010 und 2020), eine Uhrzeit (als volle Stunde) und ein Temperaturwert übergeben wird und die darauf angibt, wie dieser Temperaturwert bezüglich der genannten Einteilung einzuschätzen ist.

Berechnen Sie dazu als erstes den zu der angegebenen Zeit gehörigen normalen Temperaturwert und stellen dann fest, zu welcher der zuvor gefundenen Klassen die Abweichung zwischen diesem und dem übergebenen Temperaturwert gehört.

Hinweis: Die zu der betreffenden Zeit gehörige Normaltemperatur erhält man mit Hilfe der zuvor bestimmten Regressionsfunktionen (periodisch und linear). Da die Regressionsfunktionen als Eingabe die Anzahl der Stunden, die seit dem 01.01.2010, 00:00 Uhr vergangen sind, benötigen, muß eine entsprechende Umrechnung durchgeführt werden. Diese Umrechnung kann in *Matlab* mit folgender Funktion durchgeführt werden:

```
function std=StundenAb010110(tag,monat,jahr,stunde)
std=etime([jahr,monat,tag,stunde,0,0],[10,01,01,0,0,0])/3600;
```

2. Aufgabe: In der Datei „wbsa0230c.dat“ finden Sie aus drei Komponenten bestehende Datenwerte der Form $[x, y, z]$. Lesen Sie diese Daten in *Matlab* ein und erzeugen Sie mit `plot3` ein dreidimensionales Diagramm dieser Punkte. Man erkennt, daß die Punkte ungefähr einen Paraboloiden (eine „räumliche Parabel“) nachbilden. Will man eine Funktion finden, die die funktionale Abhängigkeit $\mathrm{f}(x, y) = z$ näherungsweise darstellen, so bietet sich hier eine zweidimensionale polynominale Regression zweiter Ordnung an. Gesucht ist ein Regressionspolynom zweier Veränderlicher

$$\mathrm{p}(x, y) \;=\; a_1x^2 + a_2xy + a_3y^2 + a_4x + a_5y + a_6 \qquad (0.3)$$

Finden Sie ein solches Polynom zu den vorliegenden Datenwerten mit Hilfe eines MKQ-Ansatzes. Orientieren Sie sich dabei sowohl bei der Vorgehensweise als auch bei den Bezeichnungen an den in der Vorlesung erläuterten Verfahren[4] Zeichnen Sie Ihre gefundenes Regressionspolynom gemeinsam mit den Datenpunkten (*Matlab*-Funktionen `plot3`, `meshgrid` und `mesh`).

[3] Informieren Sie sich über die Matlabfunktion `kmeans`.

[4] Verwenden Sie bitte keine im Internet oder in der sonstigen Literatur gefundene vorgefertigte Lösung.