Prof. Dr. U. Stadtmüller
M. Zampiceni

Sommer 2013
27.05.2013
Blatt 05

Funktionale Datenanalyse

(Abgabe: Mo., 03.06.2013, vor 13:15)

1. Lade den Datensatz `sample` von der Vorlesungshomepage herunter. Nehme an, dass es sich um Messungen handelt, die man für 200 Individuen bezüglich dreier Merkmale aufgenommen hat.

   (a) Berechne die empirische Kovarianzmatrix von sample und bestimme ihre Eigenwerte und Eigenfunktionen.

   (b) *Fur die, die `R 3.0.0` oder spätere Versionen haben:* Installiere und lade das Paket `rgl`.
   - Plotte den Datensatz mit `plot3d`.
   - Füge dem Plot die 3 Eigenvektoren und die 3 Geraden, die von den Eigenvektoren erzeugt werden, hinzu.
     Die Befehle `lines3d(...)` und `abclines3d(...)` können nützlich sein.
   - Benutze `ellipse3d(...)`, um ein Ellipsoid dem Plot hinzuzufügen, das eine 95% Konfidenzregion für den gegebenen Datensatz hergibt.

   (b') *Für die, die ältere Versionen von `R` haben:*
   - Plotte die erste gegen die zweite Komponente des Datensatzes.
   - Füge dem Plot die entsprechende Projektionen der 3 Eigenvektoren und die (Projektionen der) 3 Geraden, die von den Eigenvektoren erzeugt werden, hinzu.
   - Wiederhole das Ganze für die zweite gegen die dritte Komponente und für die erste gegen die dritte Komponente.

   (c) Wie viele Dimensionen kann man benutzen, um den Datensatz vernünftig zu beschreiben?

   (d) Berechne, entweder direkt oder mit dem Befehl `princomp`, die Hauptkomponenten des Datensatzes und plotte sie.

   (6 Punkte)

2. In der Vorlesung wurde gezeigt, dass

$$CV = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \hat{m}_{-j}(t_j))^2 \approx \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i - \hat{m}(t_i)}{1 - S_{ii}}\right)^2$$

   Nun wollen wir es anhand eines konkreten Beispiels rechnen.

   (a) Simuliere mit `R` folgende Situation: $m(t) = (1-t^2)^2\mathbb{I}_{[-1,1]}(t)$ und wir nehmen an, dass m an den Punkten $t_0 = -1, t_1 = -0.99, \ldots, t_n = 1$ gemessen wurde, und dass die Messfehler ϵ_i iid sind, mit $\epsilon \sim \mathcal{N}(0, 0.04)$.

   (b) Wir betrachten als Basis $\{1, \cos(\pi t), \ldots \cos(k\pi t)\}$ für $k = 1, \ldots, 10$.

(c) Berechne wie in der Vorlesung (für $k = 1, \ldots, 10$) die Matrix $S = \Phi(\Phi^t\Phi)^{-1}\Phi^T$ und den Schätzer $(\hat{m}(t_1), \ldots, \hat{m}(t_n))^T$, und die Werte von

$$a(k) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i - \hat{m}(t_i)}{1 - S_{,ii}}\right)^2.$$

(d) Berechne nun (für $k = 1, \ldots, 10$) die Werte von

$$CV(k) = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \hat{m}_{-j}(t_j))^2$$

und vergleiche diese mit den vorherigen.
Kommt man zum gleichen Ergebnis (wie viele Basisfunktionen funktionieren am besten)?

Hinweis: Die Matrix S_{-j} kann berechnet werden durch: $S_{-j} = (S_{\bullet,1}, \ldots, S_{\bullet,j-1}S_{\bullet,j+1}\ldots S_{\bullet,n})$, wobei $S_{\bullet,i}$ die $i-$te Spalte von S ist.

(6 Punkte)