Prof. Dr. U. Stadtmüller
M. Zampiceni

Sommer 2013
01.07.2013
Blatt 09

Funktionale Datenanalyse

(Abgabe: Mo., 08.07.2013, vor 13:15)

1. Sei $X = \{X(t), t \in T\} \in L_2(T)$ ein stochastischer Prozess mit $\mathbb{E}(\|X\|^2) < +\infty, \mathbb{E}(X^2(t)) < +\infty$ für alle $t \in T$ und $\mathbb{E}(X(t)) = 0$ für alle $t \in T$. Sei C die Kovarianzfunktion von X, d.h.: $C(t, u) = \text{cov}(X(t), X(u))$ für $t, u \in T$.

   (a) Sei $\mathcal{L}_X : L_2(S \times T) \to L_2(S)$ durch

   $$\mathcal{L}_X(f)(s) = \int_T X(u) f(s, u)\, du$$

   definiert. Überprüfe, dass der adjungierte Operator $\mathcal{L}_X^* : L_2(S) \to L_2(S \times T)$ durch

   $$\mathcal{L}_X^*(g)(s, t) = X(t) g(s)$$

   gegeben ist. Wie kann man das Regressionsproblem (bekannt aus der Vorlesung):

   $$Y(s) = \alpha(s) + \int_T X(t)\beta(s, t) dt + \epsilon(s)$$

   mithilfe von $\mathcal{L}_X$ umschreiben?

   (b) Sei $\Gamma : L_2(S \times T) \to L_2(S \times T)$, mit

   $$\Gamma(h)(s, t) := \int_T C(t, u) h(s, u)\, du.$$

   Zeige, dass $\Gamma = \Gamma^*$ und dass

   $$\Gamma = \mathbb{E}(\mathcal{L}_X^* \mathcal{L}_X).$$

   (2.5+4.5 Punkte)

2. Wir betrachten jetzt den Datensatz `daily` aus dem Packet `fda`.

   (a) Berechne die Mittelwertfunktion für die täglichen Temperaturen an den 35 Wetterstationen, und ziehe diese bei jeder Kurve ab. Speichere den neuen Datensatz. gehe davon aus, dass die Temperaturen immer um 12:00 aufgenommen wurden (also beim 0.5ten Tag, beim 1.5ten Tag usw)

   (b) Diesmal wollen wir eine Fourierbasis erzeugen: Sie muss Periode 365 haben, und 50 Sinus- (bzw Cosinusfunktionen) enthalten. Speichere die zentrierte Messungen für die Temperatur als *functional Data Objects* und plotte das Ergebnis. Transformiere auch die Mittelwertfunktion; Vergleiche graphisch die "geglättete" Temperaturen mit den "rauhen".

   (c) Führe nun die funktionale Hauptkomponentenanalyse durch, lasse hierbei 20 Hauptfunktionen berechnen, und plotte die Schätzungen, die man für die Temperaturkurven bekommt, wenn man

   - Nur die erste Hauptfunktion betrachtet (und natürlich die Mittelwertfunktion)
   - Die ersten zwei Hauptfunktionen (und Mittelwertfunktion)

Erzeuge ein Scatterplot bestehend aus der ersten zwei Koeffizienten jeder Kurve $((\xi_i^{(1)}, \xi_i^{(2)})$, $i = 1, \ldots, n)$. Beschriebe qualitativ, was jede der zwei ersten Hauptkomponenten "darstellt".

(d) Im (vollständigeren) Datensatz `CanadianWeather` sind auch die Klima-zonen der 35 Stationen angegeben (Pacific, Artic, Antartic, Continental). Plotte wieder die 35 Kurven aus Teilaufgabe (d) in unterschiedliche Farben, je nach Klimazone, und wiederhole es für die ersten 2 Scores. Interpretiere die Bilder.
Plotte nun die ersten 3 Scores, immer noch mit der Klimazone-farbe. Kann man jetzt die 4 Gruppen besser erkennen?

(0.5+0.5+2+2 Punkte)