



## Blatt 9

### Aufgabe 9.1

(6 Punkte)

Wir betrachten ein inhomogenes Markovmodell erster Ordnung und nehmen als A-Priori-Dichte der Parameter  $\theta$  des Links-Rechts-Modells ein Produkt aus Dirichlet-Dichten (wie in der Vorlesung gegeben) mit einer *equivalent sample size*  $\epsilon$  und ansonsten gleichverteilten Hyperparametern an.

- Leiten Sie für  $L = 3$  und  $A = 2$  die A-Priori-Dichte der Parameter  $\phi$  und den MAP-Schätzer von  $\phi$  her.
- Wie lauten die A-Priori-Dichte und der MAP-Schätzer der Parameter  $\phi$  des Rechts-Links-Modells? Vergleichen Sie die Dichte und den MAP-Schätzer des Links-Rechts-Modells mit denen des Rechts-Links-Modells. Transformieren Sie hierzu den MAP-Schätzer des Links-Rechts-Modells in den Raum der Parameter des Rechts-Links-Modells und zeigen sie entweder, dass sich die A-Posteriori-Dichte nicht ändert, oder, dass zumindest der MAP-Schätzer des Links-Rechts-Modells in den MAP-Schätzer des Rechts-Links-Modells transformiert wird.
- Berechnen Sie für den Datensatz `seq_1` den maximalen Wert der A-Posteriori-Dichte für  $\theta_{\text{links-rechts}}$ ,  $\theta_{\text{rechts-links}}$ ,  $\phi_{\text{links-rechts}}$  sowie  $\phi_{\text{rechts-links}}$  und vergleichen Sie die 4 Werte.

### Aufgabe 9.2

(4 Punkte)

Leiten Sie den M-Schritt des modifizierten EM-Algorithmus für ein PWM-Mischmodell sowie ein WAM-Mischmodell in  $\phi$  Parametrisierung her.

### Aufgabe 9.3

(5 Punkte)

Im Folgenden wollen wir die Klassifikationsgenauigkeit der  $\phi$  Parametrisierung mit der Klassifikationsgenauigkeit der  $\theta$  Parametrisierung vergleichen. Als Maß für die Klassifikationsgenauigkeit bietet sich die Fläche unter der ROC-Kurve (AUC) an.

Auf der Seite zur Übung finden Sie vier Datensätze: `art1train`, `art2train`, `art1test`, `art2test`. Diese Daten wurden jeweils generiert von einem WAM-Mischmodell mit zwei Komponenten. Die Wahrscheinlichkeiten  $(\pi_1, \pi_2)$  der Komponenten wurden von einer Beta-Dichte mit Hyperparametern  $(32, 32)$  gezogen. Die Parameter beider WAM-Modelle wurden unabhängig von  $(\pi_1, \pi_2)$  und unabhängig voneinander jeweils von einem Produkt aus Dirichlet-Dichten mit Hyperparametern  $(8, 8, 8, 8)$  für

---

die Startwahrscheinlichkeiten und  $(2, 2, 2, 2)$  für die Übergangswahrscheinlichkeiten gezogen. Die Datensätze `art1train` und `art1test` sind Trainings- und Testdatensatz für Klasse 1. Die Datensätze `art2train` und `art2test` sind Trainings- und Testdatensatz für Klasse 2.

- (a) Lernen Sie auf den Trainingsdaten die WAM-Mischmodelle in der  $\theta$  Parametrisierung, analog zu Aufgabe 5.2, und die WAM-Mischmodelle in der  $\phi$  Parametrisierung, jeweils mit  $K = 2$  Komponenten und  $\epsilon = 64$ . Vergleichen Sie die Klassifikationsgenauigkeit der beiden Parametrisierungen auf den Testdatensätzen anhand der AUC.
- (b) Lernen Sie auf den Trainingsdaten `seq_1`, `seq_3` aus Aufgabe 5.1 die WAM-Mischmodelle in der  $\phi$  Parametrisierung mit  $K = 2$  Komponenten und  $\epsilon = 64$ . Vergleichen Sie die Klassifikationsgenauigkeit auf den Testdatensätzen `seq_2`, `seq_4` anhand der AUC mit der aus Aufgabe 5.1.
- (c) Variieren Sie nun die *equivalent sample size*  $\epsilon$  zwischen dem kleinstmöglichen Wert und 256, und plotten Sie die AUC-Werte als Funktionen von  $\epsilon$ . Vergleichen Sie die 4 Kurven mit denen aus Aufgabe 5.2.

#### **Aufgabe 9.4**

(5 Punkte)

Zusatzaufgabe zu Aufgabe 9.1: Leiten Sie die entsprechende A-Priori-Dichte der Parameter des Rechts-Links-Modells für beliebiges  $L$  und/oder beliebiges  $A$  her. Wie lautet die A-Posteriori-Dichte der Parameter des Rechts-Links-Modells, und wie lautet der MAP-Schätzer?

**Abgabetermin: 16. Januar**

---