



Blatt 10

Aufgabe 10.1

(2 Punkte)

Leiten Sie den modifizierten EM-Algorithmus für Sequenzmotive unter Nutzung der ϕ Parametrisierung her.

Aufgabe 10.2

(2 Punkte)

Leiten Sie den *Gibbs Sampling Algorithmus* für ein PWM Mischmodell, ein WAM Mischmodell und ein Mischmodell aus PWMs und WAMs her.

Aufgabe 10.3

(2 Punkte)

Leiten Sie den *Collapsed Gibbs Sampling Algorithmus* für ein PWM Mischmodell, ein WAM Mischmodell und ein Mischmodell aus PWMs und WAMs her.

Aufgabe 10.4

(4 Punkte)

Auf der Seite zur Vorlesung finden Sie einen Datensatz (`mix.txt`) mit künstlich erzeugten Sequenzen der Länge $L = 7$. Ein Teil dieser Sequenzen wurde von einer PWM und ein anderer Teil der Sequenzen von einer WAM gezogen. Für beide Modelle wurde eine *equivalent sample size* von $\epsilon = 32$ genutzt.

- (a) Lernen Sie ein Mischmodell aus einer PWM und einer WAM in der θ -Parametrisierung mit dem modifiziertem EM-Algorithmus. Berechnen Sie dann für jede Sequenz i die Werte $\gamma_{i,k}$ mit $k = 1, 2$. Ordnen Sie nun jede Sequenz i der Komponente k (PWM oder WAM) mit dem maximalen $\gamma_{i,k}$ zu.
 - (b) Starten Sie den Gibbs Sampling Algorithmus für ein Mischmodell aus einer PWM und einer WAM. Bestimmen Sie für jede Sequenz i , wie oft sie (im Zuge des Gibbs Samplings) der Klasse $k = 1$ (PWM) oder der Klasse $k = 2$ (WAM) zugeordnet wurde und wählen Sie die Klasse k mit der größeren Anzahl von Zuordnungen.
 - (c) Auf der Seite zur Vorlesung finden Sie außerdem einen Datensatz (`mix_classes.txt`), der für jede der Sequenzen i die realen, bei der Generierung genutzten, Klassenzuordnungen enthält. Verwenden Sie diesen Datensatz, um Ihre Ergebnisse aus Aufgabe a) und b) zu überprüfen. Welches der beiden Verfahren hat die Zuordnung besser rekonstruiert?
-