Übungen Algorithmen der Bioinformatik II
Wintersemester 2004/05
Stefan Posch, Birgit Möller, Denis Williams

Institut für Informatik
Universität Halle

Blatt 13

Aufgabe 13.1 Betrachten Sie zwei (verschiedene) Bayesbäume $S1, S2$ auf diskreten Zufallsvariablen $(X_1, \ldots, X_L)$, deren zugehörigen ungerichteten Bäume identisch sind.

(a) Gegeben seien nun konkrete Parameter $\vec{\theta}_{S_1}$ für den Bayesbaum S_1. Geben Sie ein Verfahren an, um daraus Parameter $\vec{\theta}_{S_2}$ für S_2 zu bestimmen, sodaß

$$P(\vec{X}|S_1, \vec{\theta}_{S_1}) = P(\vec{X}|S_2, \vec{\theta}_{S_2})$$

(b) Zeigen Sie mit den Resultaten der Vorlesung und die ML-Schätzwerte $\hat{\vec{\theta}}^{\mathrm{ML}}_{S_1}$, $\hat{\vec{\theta}}^{\mathrm{ML}}_{S_2}$:

$$P(\vec{X}|S_1, \vec{\theta}^{\mathrm{ML}}_{S_1}) = P(\vec{X}|S_2, \hat{\vec{\theta}}^{\mathrm{ML}}_{S_2})$$

Aufgabe 13.2 Betrachten Sie nun zwei binäre Zufallsvariable X_1, X_2 und die zwei möglichen S_1 und S_2 für die Menge.

(a) Geben Sie zunächst die MP-Schätzwerte für die Parameter von S_1 und S_2 an, und zwar in Abhängigkeit von den absoluten Häufigkeiten einer beliebigen Beobachtung $N_{12}(a, b), a, b, \in \{0, 1\}$ und den Hyperparametern der Beta-Priors.

(b) Geben Sie eine konkrete Wahl der Hyperparameter für S_1 und S_2 an, sodaß

$$P(\vec{X}|S_1, \vec{\theta}^{\mathrm{MP}}_{S_1}) = P(\vec{X}|S_2, \hat{\vec{\theta}}^{\mathrm{MP}}_{S_2})$$

für beliebige Beobachtungen gilt.

Wie ändert sich die Situation, wenn wir statt des MP-Schätzers nun den MAP-Schätzer verwenden?

Aufgabe 13.3 Betrachten Sie das Strukturlernen für Bayesnetze mittels *greedy hill climbing* und der Likelihood unter den ML-Schätzwerten für die Parameter

$$P(\vec{d}_N|S, \hat{\vec{\theta}}^{\mathrm{ML}}_S)$$

als scoring Funktion.

(a) Geben Sie eine möglichst einfache (d.h. mit wenig Aufwand berechenbare) Formel für die Änderung der scores für lokale Kantenänderungen an. Als lokale Kantenänderung betrachten wir

Abgabe: 27.1.05

- entfernen einer Kante
- umdrehen der Richtung einer Kante (falls der Graph azyklisch bleibt)
- einfügen einer Kante (falls der Graph azyklisch bleibt)

(b) Überlegen Sie eine effiziente Datenstruktur für die Durchführung des *greedy hill climbing*.

Aufgabe 13.4 Betrachten Sie die Datensätze aus Aufgabe 12.3.

(a) Verwenden Sie eine 10-fache Crossvalidation und einen Bayes-Klassifikator mit iMM(0) für die TFBS und hMM(0) für den Hintergrund um den Testfehler (erwarteter Klassifikationsfehler) zu approximieren.

(b) Wiederholen Sie (a) z-fach, mit $z = 10, 50, 100, 500, 1000$ und bestimmen Sie das arithmetische Mittel sowie die empirische Standardabweichung des Klassifikationsfehler über die z Wiederholungen.

Versuchen Sie den Verlauf der empirischen Standardabweichung mit einem früheren Ergebnis der Vorlesung in Zusammenhang zu bringen.

Abgabe: 27.1.05