

BIOINFORMATIKA – ZADÁNÍ

2. SAMOSTATNÉ PRÁCE

Skryté markovské modely / Genová exprese

Podmínky

Pro získání plného počtu bodů (tj. 20) je potřeba si vybrat jeden z dále uvedených úkolů a vypracovat jej

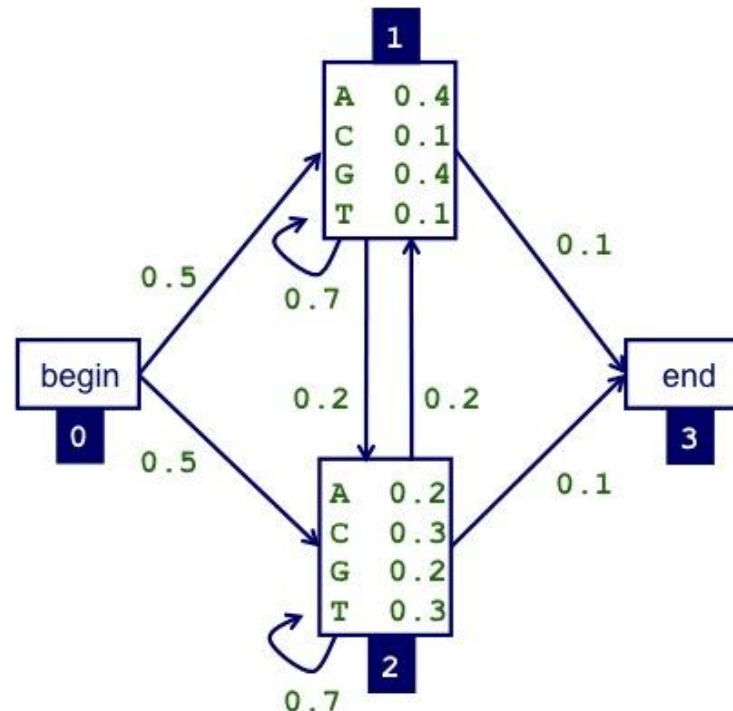
Úkoly můžete odevzdávat do 13. cvičení, přičemž pro pozdní odevzdání budou platit podmínky na webu Marka Cravena ([tady](#)) - za každý den navíc se vám bude odečítat 1 bod.

Varianta 1: Skryté markovské modely (implementace)

1. Naprogramujte Forward algoritmus a Viterbiho algoritmus a použijte je pro zodpovězení otázek na následujícím slajdu.
2. Není potřeba, abyste řešili uživatelský vstup. Stačí, když napíšete funkce, které převezmou na vstupu vámi zvolenou reprezentaci skrytého markovského modelu a sekvenci a na základě toho vypočítají pravděpodobnost této sekvence (Forward algoritmus), resp. nejpravděpodobnější posloupnost stavů (Viterbiho algoritmus).
3. Pozor! Je potřeba, aby vaše implementace opravdu používala dynamické programování.

Varianta 1: Skryté markovské modely

2. Spočítejte pravděpodobnost sekvence CGTCAG pomocí vámi implementovaného Forward algoritmu pro následující model.
3. Nalezněte nejpravděpodobnější sekvenci stavů, která vygenerovala sekvenci CGTCAG pomocí vámi implementovaného Viterbiho algoritmu.



Varianta 2: Analýza dat genové exprese

- **Vypracujte alespoň 75% úkolů z:**
- <http://www-users.cselabs.umn.edu/classes/Spring-2012/csci3003/files/lab7/Lab7.pdf>
- **Za jeden úkol se považuje úkol nebo otázka označená malým písmenem (a), (b),..., což odpovídá v odkazovaném dokumentu podotázkám. Pokud nějaká otázka nemá podotázky, považuje se celá za jeden úkol.**

Varianta 3: Gene finding pomocí HMM

- Náročnější (a zajímavější) úkol za 40 bodů (tj. nebudete muset dělat úkol číslo 3)
- <http://www.biostat.wisc.edu/~craven/776/hw3.html>