Question Code P/HPC

Diagnostic Exam Fall 2003

High Performance Computing Paper

This examination is based on the paper, *Hidden Markov Models in Computational Biology: Applications to Protein Modeling*, by Anders Krogh, et al.

Consider the following multiple alignment of six amino acid sequences from the protein-binding region of astroviruses and related RNA viruses, obtained recently by group of researchers at ODU and Eastern Virginia Medical School . Construct a Hidden Markov Model that could output these sequences. Indicate the columns in the multiple alignment that you choose to correspond to *match* states, and identify columns corresponding to *insertion* states and *delete states* in each sequence. Tabulate the transition probabilities from each state to the next state, and compute the emission probabilities in each state. Avoid overfitting the data by using Laplace's rule to obtain pseudocounts as needed.

seq 1	Q	K	K	-	K	G	K	T	K
seq 2	Q	A	K	G	K	T	K	-	K
seq 3	E	K	K	-	K	G	K	T	K
seq 4	Q	K	K	-	K	G	K	N	K
seq 5	E	G	K	N	K	G	K	T	K
seq 6	E	G	K	-	K	G	K	T	K