

Algorithms in bioinformatics – Quiz

PSSM

- 1) In equation $p = (a*f + b*g)/(a+b)$ what is a and b, and how do you estimate their value. Likewise what is f and g?
- 2) How does the formula reduce if you only have one sequence from which to make the calculation
- 3) What is the step to calculate the weigh-matrix (PSSM) elements from the values of p?
- 4) How do you score a peptide to the constructed PSSM
- 5) What is sequence weighting?
- 6) What are r and s in the equation $1/(r*s)$, and how do you calculate the weight of a peptide using this equation?

Alignment

- 1) Why is the O3 algorithm slower than the O2?
- 2) To fill out the D matrix in the O3 algorithm, you only need to calculate and compare 3 values? Correct or false
- 3) What information is stored in the P and Q matrices of the O2 algorithm
- 4) What part of the O3 and O2 algorithms should be updated to implement an sequence profile scoring scheme rather than the current Blosum scheme?

GibbsSampling

- 1) In the equation $p = \min(1, \exp(dE/T))$, what is dE and T?
- 2) What is the effect of having a value of $T > 0$?
- 3) Is GibbsSampling guaranteed to find the optimal solution to a problem?

HMM

- 1) What answers do you get when you apply the Viterbi algorithm to score a sequence to an HMM?
- 2) What answers do you get when you apply the Forward algorithm to score a sequence to an HMM?
- 3) What answers do you get when you apply the posterior decoding algorithm to score a sequence to an HMM?

Cross validation and training of data driven prediction methods

- 1) What is Cross-validation?
- 2) What is the single most important issue to deal with when making cross-validation partitions?
- 3) What is early stopping?
- 4) If you use early stopping, why can you not use the test data to report the model performance? And how do you the in the case make a setup so that you can estimate the model performance

SMM

- 1) What is the effect of the second term in the Error function $E = 1/2*(O-t)^2 + \sum \lambda w^2$
- 2) What is gradient descent, and how do you use it to find the optimal model parameters?
- 3) Explain the equation $w' = w - \epsilon * dE/dw$
- 4) How do you calculate dE/dw_1 if O has the form $O = I_1*w_1 + I_2*w_2 + I_3*w_3$?

Hobohm and data redundancy

- 1) When is it important to deal with data redundancy?
- 2) How do the two Hobohm algorithms work?
- 3) Which of the two algorithms is faster, and why?
- 4) In what situation will the time for running Hobohm1 be comparable to that of Hobohm2?