UNIVERSITY OF
CAMBRIDGE

M. PHIL. IN COMPUTATIONAL BIOLOGY

Wednesday 30 May 2007 2.00 to 4.00

COMPUTATIONAL BIOLOGY

*Attempt **ALL** questions.*

*There are **THREE** questions in total.*

The questions carry equal weight.

STATIONERY REQUIREMENTS
Cover sheet
Treasury Tag
Script paper

SPECIAL REQUIREMENTS
None

You may not start to read the questions printed on the subsequent pages until instructed to do so by the Invigilator.

1 Functional Genomics

(a) Describe how a microarray measures mRNA expression levels. List three sources of noise that affect the expression levels and suggest methods for how we might control for them during quantitative analysis of the data. 40%

(b) Describe and compare hierarchical and partitioning methods for clustering of samples. Describe two examples of using clustering techniques to analyse microarray data. 30%

(c) Describe, with the aid of diagrams, how the K means clustering technique finds K representative centroids. In particular, define the initialisation of cluster centroids, how centroids move during updates, and when to terminate updates. 30%

2 Computational Neuroscience

(a) Describe the key steps in the generation of an action potential in an axon. In your answer, describe what happens to sodium and potassium ions, and their corresponding channels, during the different phases of an action potential. 40%

(b) Describe how the Hodgkin-Huxley model (below) captures the dynamics of action potential propagation. Your answer should include a definition for each term in the equation, and a brief description of how m, n, h vary over time.

$$C_m \frac{dV}{dt} = -g_l(V - E_l) - g_{Na} m^3 h (V - E_{Na}) - g_K n^4 (V - E_K) + I_e$$

How does the Hodgkin-Huxley model compare to a passive integrate and fire neuron model? 30%

(c) Define the following terms, in the context of supervised learning:

1. multi-layer perceptrons
2. training set
3. error function
4. linear separability
5. gradient descent learning rule 30%

3 Statistical Methods in Bioinformatics

(a) Suppose that $(Y_n, X_n), n = 0, 1, \ldots, N$ is a hidden Markov model in which the hidden chain X has transition matrix $P = p(i,j)$ and initial distribution ρ, and the output distribution is specified by $B = b(j,k)$. Define

$$\alpha_n(j) = P(Y_0^n, X_n = j), \quad \beta_n(j) = P(Y_{n+1}^N \mid X_n = j).$$

Show that for each n,

$$P(Y_0^N) = \sum_j \alpha_n(j)\beta_n(j)$$

15%

(b) Establish that

$$\alpha_{n+1}(j) = b(j, Y_{n+1}) \sum_i \alpha_n(i) p(i,j), \quad n = 0, 1, \ldots, N-1$$

35%

(c) Answer the following, in the context of hidden Markov models (HMMs):

1. What is the Viterbi algorithm used for?
2. What is a dynamic Bayesian network?
3. How can the HMM in part (a) be generalised?
4. Describe three examples of HMMs in bioinformatics.

50%

END OF PAPER