

## Example exam questions for Lectures 6-8

### Q1 (Lecture 6)

Explain why “genetic” is used in the term “genetic algorithms”? In your answer make reference to DNA and agents.

### Q2 (Lecture 7)

What is the Baldwin effect? Explain and contrast with Lamarckian evolution, i.e. the transfer of traits (from generation to generation) *learned* during the individual’s lifetime.

### Q3 (Lecture 8)

Micro-array technology allows the monitoring of expression of genes in selected tissues. Sometimes a disease can be identified by looking at the expression pattern of a patient in relation to an expected pattern (established from a set of other patients with known disease states). Since there are thousands of genes and only a small number of patterns with known disease states, it is easy to “overfit” a model (i.e. find a model that nicely classifies the training samples but fails brutally on novel samples). One resort is to select a much smaller set of genes that seem to be genuinely relevant for the disease, and to build a model and classify novel samples based on these selected genes only.

- a. Explain how *either* K-means or a Support Vector Machine could be used to classify novel samples on basis of selected genes, and into two classes (say A and B).

To reduce the number of input features to a classifier, Golub *et al* established how well each gene *correlated* with a disease state (types of Leukemia in their case). Only the genes with the strongest correlation were used to classify novel samples.

- b. Describe an alternative way (to correlation) of selecting a smaller set of input features to perform accurate classification. If you can’t think of one, Principal Components Analysis has been used by several and Eisen *et al* used Hierarchical Cluster Analysis.