# NANYANG TECHNOLOGICAL UNIVERSITY

## SEMESTER 1 EXAMINATION 2006-2007

### H6429: Computational Intelligence – Methods & Applications

November/December 2006 Time allowed: 3 hours

**INSTRUCTIONS**

1. This paper contains **FOUR (4)** questions and comprises **FOUR (4)** pages.

2. Answer **all FOUR (4)** questions.

3. All questions carry equal marks.

4. This is an open book exam; you are allowed to have with you **one book** of your choice, but **no other** printed material, notes, or copy-books.

---

**1. General questions related to Computational Intelligence (CI) concepts and definitions.**

(a) CI is usually defined by listing its three core branches and a few additional branches. Give an example of such definition, explain why it is not sufficient, and give a definition that covers all CI areas.

(4 marks)

(b) CI methods draw inspirations from many sources, including neurobiology and psychology. List a few neurobiological inspirations and link them to CI methods. Do the same for psychological inspirations. Explain how they differ from each other.

(5 marks)

Question No. 1 continues on Page 2

(c) Write the summation rules for joint probabilities, including the case of continuous variable *X*. Explain the meaning of conditional probabilities and introduce their summation rules. Explain how Bayes theorem, which calculates posterior probabilities from class conditional, prior and evidence probabilities, results from the definition of conditional probabilities.

(5 marks)

(d) Two rare subtypes of a disease have been identified, $D_1$ and $D_2$. Two experimental medical procedures were used, M1 and M2. For disease $D_2$ procedure $M_1$ was used only 1 time and in all other cases the number of uses was $N(M_i,D_j)=3$. However, it was found that using procedure $M_1$ instead of $M_2$ is 4 times more risky because of later complications. Use Bayesian risk estimation to determine which procedure should be used in which situation.

(5 marks)

(e) You want to create a visual map displaying similarities between different music groups. Explain how to generate numerical data from description of music groups, and how the popularity of these groups will affect such representation. Explain how to generate numerical data from comparisons between different groups and note that such data may be asymmetric. Which method would you use to visualize such data and how should such data be preprocessed to create useful maps?

(6 marks)

**2. Questions related to exploratory data analysis.**

(a) Present vertices (0000), (0001), (0011), (0111), and (1111) of a 4-dimensional cube using star plots in a legible way. (Hint: avoid collapsing star plot representations into a single point.)

(5 marks)

(b) Describe the Independent Component Analysis (ICA) as a means for 2-D data visualization. What condition does it try to optimize? How does it differ from the Principal Component Analysis (PCA)?

(6 marks)

Question No. 2 continues on Page 3

(c) Derive the formula for first direction of the Fisher projection in two-class case. How can the second Fisher direction be obtained?

(7 marks)

(d) Compare multidimensional scaling (MDS), Principal Component Analysis and Self-Organized Maps, including computational complexity of these 3 methods. Which of these methods will you use on typical gene microarray data (>10.000 features, <100 samples)? Which methods is preferred for medical screening data (~10 features, >100.000 samples)?

(7 marks)

**3. Questions about the theory of adaptive systems.**

(a) Two $d$-dimensional Gaussian probability density functions are given:

$$P(\mathbf{X}|\omega_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{X}-\boldsymbol{\mu}_i)\right)$$

They have identical diagonal covariance matrices and differ only by their means $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$. Assuming equal *a priori* probabilities, derive the simplest MAP rule for this case, and show that it is equivalent to the nearest prototype classifier. What kind of decision borders are produced by such a classifier if more than two prototypes are used?

(6 marks)

(b) Why it is necessary to select a model with appropriate bias for a given data? What are the main tradeoffs involved? Describe a cross-validation procedure for model selection, its variants, strong and weak points, and its applications.

(6 marks)

(c) Define joint information $H(X,Y)$, conditional information $H(X|Y)$, mutual information $MI(X;Y)$ and prove that $H(X,Y) = H(X|Y) + H(Y|X) + MI(X;Y)$, and that $H(X,Y) = H(X) + H(Y) - MI(X;Y)$. Consider only the discrete case.

(6 marks)

Question No. 3 continues on Page 4

(d) Draw the network structure for a single-layer Separable Function Network (SFN), and write the general form of mappings implemented by such networks. Describe the similarities and the differences of SFN networks and the Radial Basis Function networks. Why do such networks provide better models than polynomial expansions?

(7 marks)

**4. Questions about computational intelligence algorithms.**

(a) What criterion does the CHAID decision tree algorithm use and why is it a good criterion? How is the tree constructed using this decision?

(6 marks)

(b) Describe the logistic discrimination model. When is this model exact? Present classification rules for logistic discrimination. How do the decision borders look like? How can one optimize parameters in this model?

(6 marks)

(c) Formulate the SVM large margin optimization problem, write Lagrangian for this problem, and derive the formula for discriminant function g(X) in terms of dot products between the training vectors and the vector X. Show how this discriminant function may be replaced by its kernel form.

(6 marks)

(d) The changes in variables describing an electric circuit are recorded as +1, 0 or −1. Ohm's law binds 3 variables: $V=I*R$. Draw a cube that represents these 3 values for each variable and mark the points for which Ohms law may be true with small circles (for example $(V,I,R)=(+1,+1,+1)$ or (0,0,0)) and those for which it is certainly false with small crosses (for example $(V,I,R)=(+1,-1,-1)$). What other relations between variables have the same geometrical representation? How can this representation be used to estimate possible changes of variables in larger systems, where many variables have values constrained by this type of relations?

(7 marks)

**End of Paper**