ELE888/EE8209 Intelligent Systems (2009)
Final Exam (A) - Winter 2009

- *This is a closed-book, closed-notes test. The duration of the test is 3 hours (8:10am - 11:10am).*
- *One formula sheet is provided.*
- *Only calculators with no text storage facilities are allowed. Internet, computers or phones should not be accessed during the duration of the exam.*
- *No questions related to the content of this question paper will be answered by the instructor administering the test. Any assumptions made should be clearly stated in your answer sheet.*
- *The total marks for the exam is 60 marks (worth 30% of your total final course mark). There are 4 questions and all questions carry equal marks.*
- *Absolutely no talking during the exam period and also no peeping into others answer sheets.*

Question 1 (5 + 5 + 5 Marks)

(a) List and describe the 5 key components of a pattern classification system. Sketch a block diagram to show their relationship.

(b) Enumerate the steps in designing a minimum conditional-risk classifier for the two category (ω_1 = sea-bass, ω_2 = salmon) in the fish packing plant problem. Use Bayes decision rule with a high penalty for classifying sea-bass as salmon. Answer in point form (no lengthy descriptions), provide the loss function matrix and clearly explain all assumptions.

(c) Given a training set with discrete features, arrive at a decision for whether the given condition X, is suitable for playing a round of Golf. State any assumptions made.

$X = <$ Bank Balance $=$ *Low*, Humidity $=$ *Normal*, Wind $=$ *Strong*, Location $=$ *Canada* $>$

Bank Balance	Season	Humidity	Wind	Location	Play Golf
Low	Summer	High	Strong	Canada	No
Med	Spring	Normal	Strong	Canada	No
Low	Summer	High	Weak	Bahamas	No
High	Winter	High	Weak	Canada	No
High	Summer	Normal	Weak	Canada	Yes
Low	Winter	High	Strong	Bahamas	No
Med	Spring	Normal	Weak	Bahamas	Yes
Low	Summer	Low	Strong	Canada	No
Med	Winter	High	Strong	Australia	No
Med	Summer	Low	Weak	Australia	Yes
High	Summer	Normal	Strong	Canada	No
Low	Fall	High	Strong	Australia	Yes
Low	Fall	High	Strong	Canada	No
Med	Summer	High	Strong	Canada	Yes
Low	Winter	Normal	Weak	Australia	Yes

Question 2 (2 + 4 + 2 + 4 + 3 Marks)

A 2-2-1 Multilayer Neural Network (MNN) with bias is used to solve the **XOR** problem according to the following conditions:

- Activation function (output and hidden nodes) $= \frac{1}{1+e^{net}}$
- Learning rate $\eta(\cdot) = 0.3$
- Input patterns are of the form $\mathbf{x}_i^T = [bias\ x_{i1}\ x_{i2}]$, where $bias = 1$
- Input-to-hidden weights $[w_{ji}]$: ${w_{1i}}^T = [+0.1\ -0.1\ +0.2]$; ${w_{2i}}^T = [-0.4\ +0.2\ -0.2]$
- Hidden-to-output weights $[w_{kj}]$: ${w_{1j}}^T = [+0.5\ +0.1\ -0.5]$

(a) Sketch the network and label the nodes, weights, inputs and outputs at each layer.

(b) Given the input pattern ${\mathbf{x}_i}^T = [+1\ -1\ +1]$, calculate z_k for the output node.

(c) If the Manhattan metric: $J = \sum_{k=1}^{c} |t_k - z_k|$ is used in back propagation training, show that the sensitivity δ_k of the output node is given by:

$$\delta_k = f'(net_k).sgn(t_k - z_k)$$

$$[\text{hint: } \tfrac{d}{dx}|x| = \tfrac{x}{|x|} = sgn(x)].$$

(d) Find an expression for $\triangle w_{kj}$ and use it to perform one iteration of backpropagation to update **hidden-to-output weights only** (i.e. neglect input-to-hidden weight updates).

(e) Explain the significance of adding or reducing the number of nodes (n_H) in the hidden layer of a MNN with respect to its effect on the resulting classifier.

Question 3 (5 + 2 + 4 + 4 Marks)

(a) Consider the K-means algorithm for unsupervised clustering of the following two-dimensional data samples. Given initial mean estimates for the case of c=2 below, perform two iterations of the K-means algorithm and plot the new mean estimates for each iteration. Do you expect the algorithm to continue or converge in the next iteration? Justify.

$$\mathbf{x}_1 = \begin{bmatrix} -3 \\ 2 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} -2 \\ 4 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix} -1 \\ -2 \end{bmatrix}, \mathbf{x}_5 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \mathbf{x}_6 = \begin{bmatrix} 3 \\ -2 \end{bmatrix},$$

$$\hat{\boldsymbol{\mu}}_\mathbf{1}(\mathbf{0}) = \begin{pmatrix} -4 \\ 0 \end{pmatrix}, \hat{\boldsymbol{\mu}}_\mathbf{2}(\mathbf{0}) = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

(b) What is the main difference between the K-means and Fuzzy K-means algorithm in terms of their approach to estimating means at each new iteration? List one advantage and disadvantage of each algorithm?

(c) Construct a cluster dendrogram for the one dimensional data {3, 7, 9, 14, 17, 22} using d_{max} as the distance metric. Show explicit calculations of the distances in arriving at the clusters, and label the dendrogram appropriately.

(d) Briefly explain (with the use of a diagram) the nature and purpose of the window function $\Lambda(\mid y - y^* \mid)$ in a Self-Organizing Map (SOM). What does y^* represent in this context? In your answer, describe how learning would proceed in a SOM network with an output layer of at least three nodes, when presented with a single input data sample. You may use data from the K-means question, along with your own choice of initial SOM weights to aid in your description (*no calculations necessary*).

Question 4 (3 + 3 + 3 + 4 + 2 Marks)

(a) Explain, using simple figures, the *'crossover'* and *'mutation'* operations in the context of a genetic algorithm.

(b) Outline two alternative ways a genetic algorithm can be used to represent a Multilayer Neural Network (MNN) classifier, and identify one benefit of training a MNN with the genetic algorithm versus standard backpropagation.

(c) Explain the significance of *'score'* and *'rank'* in the context of a genetic algorithm. If a genetic algorithm was used to identify the parameters of an unsupervised classifier such as K-means, explain why the score of the classifier would need to consider the concept of *'scatter'*.

(d) Compute the Jackknife and the Variance of the Jackknife estimate of the **mean** of the following data: {10, 10, 10, 20, 30, 30}.

(e) Compute the Bootstrap estimate of the mean of the data in part (d). Use a B value of 2 and the random indexes r_1 = 4, 3, 5, 4, 5, 6 and r_2 = 2, 3, 6, 1, 4, 3.

ELE888/EE8209 Intelligent Systems (Formula Sheet)

$$P(\omega_j \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega_j) \cdot P(\omega_j)}{\sum_{k=1}^{c} p(\mathbf{x} \mid \omega_k) \cdot P(\omega_k)} \tag{1}$$

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} \, |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}\,(\mathbf{x}-\boldsymbol{\mu})^T \, \boldsymbol{\Sigma}^{-1} \, (\mathbf{x}-\boldsymbol{\mu})\right] \tag{2}$$

$$\boldsymbol{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^T \equiv \int (\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^T p(\mathbf{x})d\mathbf{x} \tag{3}$$

$$R = \int R(\alpha(\mathbf{x}) \mid \mathbf{x})p(\mathbf{x})d\mathbf{x} \tag{4}$$

$$R(\alpha_i \mid \mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i \mid \omega_j)P(\omega_j \mid \mathbf{x}) \tag{5}$$

$$J(\mathbf{w}) \equiv \frac{1}{2}\sum_{k=1}^{c}(t_k - z_k)^2 = \frac{1}{2}\|\mathbf{t}-\mathbf{z}\|^2 \tag{6}$$

$$\triangle\mathbf{w} = -\eta\frac{\partial J}{\partial \mathbf{w}} \tag{7}$$

$$\frac{\partial J}{\partial w_{kj}} = \frac{\partial J}{\partial net_k}\frac{\partial net_k}{\partial w_{kj}} \tag{8}$$

$$\frac{\partial J}{\partial w_{ji}} = \frac{\partial J}{\partial y_j}\frac{\partial y_j}{\partial w_{ji}} \tag{9}$$

$$J_e = \sum_{j=1}^{c}\sum_{\mathbf{x}\in D_j}\|\mathbf{x}-\boldsymbol{\mu}_j\|^2 \tag{10}$$

$$J_{fuzzy} = \sum_{j=1}^{c}\sum_{i=1}^{n}[P(\omega_j|\mathbf{x}_i,\theta)]^b\|\mathbf{x}_i-\boldsymbol{\mu}_j\|^2 \tag{11}$$

$$\hat{P}(\omega_j|\mathbf{x}_i,\theta) = \frac{(1/d_{ij})^{\frac{1}{b-1}}}{\sum_{j=1}^{c}(1/d_{ij})^{\frac{1}{b-1}}} \tag{12}$$

$$\hat{\boldsymbol{\mu}}_j = \frac{\sum_{i=1}^{n}[P(\omega_j|\mathbf{x}_i,\theta)]^b\mathbf{x}_i}{\sum_{i=1}^{n}[P(\omega_j|\mathbf{x}_i,\theta)]^b} \tag{13}$$

$$\mathbf{w}_{ki}(t+1) = \mathbf{w}_{ki}(t) + \eta(t)\Lambda(|\mathbf{y}-\mathbf{y}^*|,t)\,[\mathbf{x}_i - \mathbf{w}_{ki}(t)] \tag{14}$$

$$\begin{gathered}
d_{min}(D_i,D_j) = min_{\mathbf{x}\in D_i,\mathbf{x'}\in D_j}\|\mathbf{x}-\mathbf{x'}\| \\
d_{max}(D_i,D_j) = max_{\mathbf{x}\in D_i,\mathbf{x'}\in D_j}\|\mathbf{x}-\mathbf{x'}\| \\
d_{avg}(D_i,D_j) = \tfrac{1}{n_i n_j}\textstyle\sum_{\mathbf{x}\in D_i}\sum_{\mathbf{x'}\in D_j}\|\mathbf{x}-\mathbf{x'}\| \\
d_{mean}(D_i,D_j) = \|\mathbf{m}_i-\mathbf{m}_j\|
\end{gathered} \tag{15}$$

$$\begin{gathered}
\mathbf{S}_j = \textstyle\sum_{\mathbf{x}\in D_j}(\mathbf{x}-\mathbf{m}_j)(\mathbf{x}-\mathbf{m}_j)^T \\
\mathbf{S}_W = \textstyle\sum_{j=1}^{c}\mathbf{S}_j \\
\mathbf{S}_B = \textstyle\sum_{j=1}^{c} n_j(\mathbf{m}_j-\mathbf{m})(\mathbf{m}_j-\mathbf{m})^T
\end{gathered} \tag{16}$$

$$XB(c) = \frac{1}{n}\cdot\frac{\sum_{k=1}^{c}\sum_{i=1}^{n}P(\omega_k|\mathbf{x}_i,\theta)\|\mathbf{x}_i-\boldsymbol{\mu}_k\|}{min\|\boldsymbol{\mu}_p-\boldsymbol{\mu}_q\|} \tag{17}$$

$$CH(c) = \frac{tr(\mathbf{S}_B)}{tr(\mathbf{S}_W)}\cdot\frac{(n-c)}{(c-1)} \tag{18}$$

$$Bias_{jack} = (n-1)\cdot(\hat{\theta}_{(.)}-\hat{\theta}) \tag{19}$$

$$Var_{jack} = \frac{(n-1)}{n}\cdot\sum_{i=1}^{n}\left[\hat{\theta}_{(i)}-\hat{\theta}_{(.)}\right]^2 \tag{20}$$

$$Bias_{boot} = (\hat{\theta}^{*(.)}-\hat{\theta}) \tag{21}$$

$$Var_{boot} = (1/B)\cdot\sum_{b=1}^{B}\left[\hat{\theta}^{*(b)}-\hat{\theta}^{*(.)}\right]^2 \tag{22}$$