Artificial Neural Networks
Examination, March 2003

Instructions

There are SIXTY questions (worth up to 60 marks). The exam mark (maximum 60) will be added to the mark obtained in the laborations (maximum 5). The total pass mark for the course is 35 out of 65.

For each question, please select a maximum of ONE of the given answers (either A, B, C, D or E). You should select the one answer that represents the BEST possible reply to the question (in some cases, there may be no obvious "wrong" answers, but one answer should always be better than the others). Every time you select the correct answer, you will be awarded +1 mark. However, every time you select an incorrect answer, a penalty score will be subtracted from your total mark. This penalty depends on the number of possible answers to the question, as follows:

Number of possible answers	Score for correct answer	Score for incorrect answer
2	+1	-1
3	+1	$-\frac{1}{2}$
4	+1	$-\frac{1}{3}$
5	+1	$-\frac{1}{4}$

If you do not give any answer to a question, no marks will be added to your total score and there will be no penalty. If you give more than one answer to a question, this will be counted as an *incorrect* answer. So please be *very* careful, and make sure that ONLY one letter (A or B or C or D or E) is visible in each of your written answers. Please write your answers very clearly so that they can be read by an average examiner!

Advice: read all of the questions before you start to answer.
Tools required: calculator.

Questions

1   Pattern Recognition

What is *classification*?

    A. Deciding which features to use in a pattern recognition problem.

    B. Deciding which class an input pattern belongs to.

    C. Deciding which type of neural network to use.

2   Pattern Recognition

Many pattern recognition problems require the original input variables to be combined together to make a smaller number of new variables. These new input variables are called

    A. *patterns.*

    B. *classes.*

    C. *features.*

3   Pattern Recognition

The process described in question 2 is

    A. a type of pre-processing which is often called *feature extraction.*

    B. a type of pattern recognition which is often called *classification.*

    C. a type of post-processing which is often called *winner-takes-all.*

4   Classical Pattern Recognition

During training, which parameters must be calculated in a *minimum distance classifier*?

    A. The mean vector of each class.

    B. The mean vector and covariance matrix of each class.

    C. The weights, connections and bias values of each class.

5   Classical Pattern Recognition

During training, which parameters must be calculated in a *Bayes optimal classifier*?

    A. The mean vector of each class.

    B. The mean vector and covariance matrix of each class.

    C. The weights, connections and bias values of each class.

6 Classical Pattern Recognition

Design a minimum distance classifier with two classes using the following training data:

Class 1: $\begin{bmatrix} -2.0 \\ 1.0 \end{bmatrix}, \begin{bmatrix} -2.0 \\ 3.0 \end{bmatrix}, \begin{bmatrix} -1.5 \\ 2.0 \end{bmatrix}, \begin{bmatrix} -2.5 \\ 2.0 \end{bmatrix}$ Class 2: $\begin{bmatrix} 2.0 \\ -2.5 \end{bmatrix}, \begin{bmatrix} 3.0 \\ -2.0 \end{bmatrix}, \begin{bmatrix} 1.0 \\ -2.0 \end{bmatrix}, \begin{bmatrix} 2.0 \\ -1.5 \end{bmatrix}$

What are the prototype vectors of the trained classifier?

A. $\mathbf{m}_1 = \begin{bmatrix} -2.0 \\ 2.0 \end{bmatrix}, \mathbf{m}_2 = \begin{bmatrix} 2.0 \\ -2.0 \end{bmatrix}$

B. $\mathbf{m}_1 = \begin{bmatrix} 2.0 \\ -2.0 \end{bmatrix}, \mathbf{m}_2 = \begin{bmatrix} -2.0 \\ 2.0 \end{bmatrix}$

C. $\mathbf{m}_1 = \begin{bmatrix} 2.0 \\ 2.0 \end{bmatrix}, \mathbf{m}_2 = \begin{bmatrix} -2.0 \\ -2.0 \end{bmatrix}$

D. $\mathbf{m}_1 = \begin{bmatrix} -2.0 \\ -2.0 \end{bmatrix}, \mathbf{m}_2 = \begin{bmatrix} 2.0 \\ 2.0 \end{bmatrix}$

7 Classical Pattern Recognition

Classify the test vector $[1.0, -2.0]^T$ with the trained classifier in question 6. Which class does this vector belong to?

A. Class 1.

B. Class 2.

8 Classical Pattern Recognition

The decision function for a minimum distance classifier is $d_j(\mathbf{x}) = \mathbf{x}^T \mathbf{m}_j - \frac{1}{2} \mathbf{m}_j^T \mathbf{m}_j$ where $\mathbf{m}_j$ is the prototype vector for class j. What is the value of the decision function for each of the classes in question 6 for the test vector $[0, -1.0]^T$?

A. $d_1(\mathbf{x}) = -2.0, d_2(\mathbf{x}) = -6.0$

B. $d_1(\mathbf{x}) = -3.0, d_2(\mathbf{x}) = -4.0$

C. $d_1(\mathbf{x}) = -4.0, d_2(\mathbf{x}) = -3.0$

D. $d_1(\mathbf{x}) = -6.0, d_2(\mathbf{x}) = -2.0$

9 Classical Pattern Recognition

Give the equation of the decision boundary between classes 1 and 2 for the minimum distance classifier in question 6.

A. $x_2 = 0$

B. $x_2 = x_1$

C. $x_2 = -2x_1 - 1$

D. $x_2 = 2x_1 + 1$

10   Training and Testing

What is an *outlier*?

    A. An input pattern which is not included in the test set.

    B. An input pattern which produces a classification error.

    C. An input pattern which is not included in the training set.

    D. An input pattern which is very similar to the prototype vector of the patterns in the same class.

    E. An input pattern which is very different from the prototype vector of the patterns in the same class.

11   Biological Neurons

Which of the following statements is the best description of Pavlov's learning rule?

    A. "If a particular input stimulus is always active when a neuron fires then its weight should be increased."

    B. "If a stimulus acts repeatedly at the same time as a response then a connection will form between the neurons involved. Later, the stimulus alone is sufficient to activate the response."

    C. "The connection strenghts of the neurons involved are modified to reduce the error between the desired and actual outputs of the system."

12   Perceptrons

A perceptron with a unipolar step function has two inputs with weights $w_1 = 0.2$ and $w_2 = -0.5$, and a threshold $\theta = -0.2$ (θ can therefore be considered as a weight for an extra input which is always set to -1). For a given training example $\mathbf{x} = [1, 1]^T$, the desired output is 1 (one). Does the perceptron give the correct answer (that is, is the actual output the same as the desired output)?

    A. Yes.

    B. No.

13   Perceptrons

The perceptron in question 12 is trained using the learning rule $\triangle \mathbf{w} = \eta \, (d - y) \, \mathbf{x}$, where $\mathbf{x}$ is the input vector, η is the learning rate, $\mathbf{w}$ is the weight vector, d is the desired output, and y is the actual output. What are the new values of the weights and threshold after one step of training with the input vector $\mathbf{x} = [0, 1]^T$ and desired output 1, using a learning rate $\eta = 0.2$?

    A. $w_1 = 0.0, w_2 = -0.5, \theta = 0.0$.

    B. $w_1 = 0.2, w_2 = -0.3, \theta = 0.0$.

    C. $w_1 = 0.2, w_2 = -0.3, \theta = -0.4$.

    D. $w_1 = 0.2, w_2 = -0.5, \theta = -0.2$.

    E. $w_1 = 0.4, w_2 = -0.5, \theta = -0.4$.
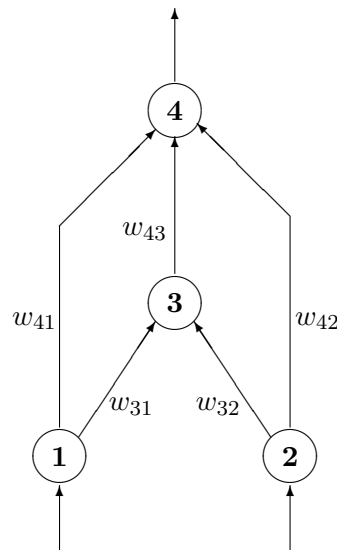
14 Perceptrons

A single-layer perceptron has 5 input units and 3 output units. How many weights does this network have?

 A. 5

 B. 9

 C. 15

 D. 25

 E. 30

15 Perceptrons

The following network is a multi-layer perceptron, where all of the units have binary inputs (0 or 1), unipolar step functions and binary outputs (0 or 1).

The weights for this network are $w_{31} = 1$, $w_{32} = 1$, $w_{41} = 1$, $w_{42} = 1$ and $w_{43} = 1$. The threshold of the hidden unit (**3**) is 1.5 and the threshold of the output unit (**4**) is 0.5. The threshold of both input units (**1** and **2**) is 0.5, so the output of these units is exactly the same as the input. Which of the following Boolean functions can be computed by this network?

 A. AND.

 B. OR.

 C. XOR.

 D. All of the above answers.

 E. None of the above answers.

16 Perceptrons

Would it be possible to train the multi-layer perceptron in question 15 to solve the XOR problem using the back-propagation algorithm?

 A. Yes.

 B. No.

17   Multi-Layer Feedforward Networks

Is the following statement true or false? "For any feedforward network, we can always create an equivalent feedforward network with separate layers."

    A. TRUE.

    B. FALSE.

18   Multi-Layer Feedforward Networks

Is the following statement true or false? "A multi-layer feedforward network with linear activation functions is no more powerful than a single-layer feedforward network with linear activation functions."

    A. TRUE.

    B. FALSE.

19   Multi-Layer Feedforward Networks

A multi-layer feedforward network has 5 input units, one hidden layer with 4 units, and 3 output units. How many weights does this network have?

    A. 12

    B. 20

    C. 27

    D. 32

    E. 40

20   Multi-Layer Feedforward Networks

What is the *credit assignment problem* in a multi-layer feedforward network?

    A. The problem of adjusting the weights for the hidden units.

    B. The problem of adjusting the weights for the output units.

    C. The problem of avoiding local minima in the error function.

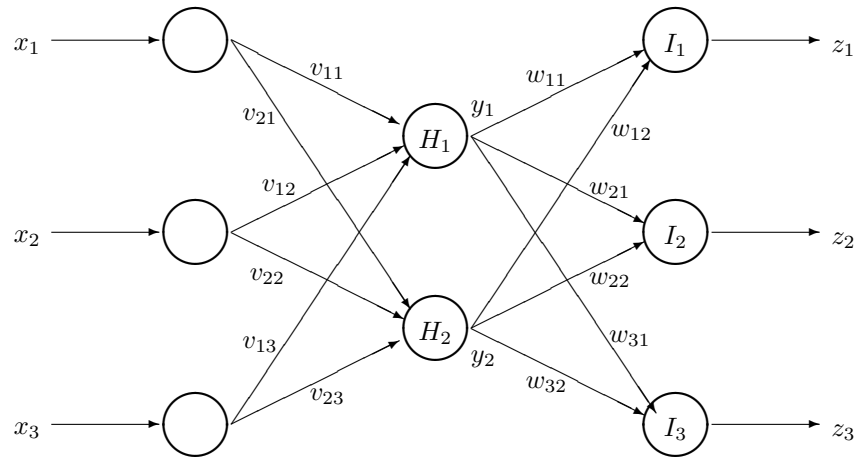   D. The problem of defining an error function for linearly separable problems.

21   Multi-Layer Feedforward Networks

What is the most general type of decision region that can be formed by a feedforward network with NO hidden layers?

    A. Arbitrary decision regions – the network can approximate any function (the accuracy of the approximation depends on the number of hidden units).

    B. Convex decision regions – for example, the network can approximate any Boolean function.

    C. Decision regions separated by a line, plane or hyperplane.

    D. None of the above answers.

22 Multi-Layer Feedforward Networks

A training pattern, consisting of an input vector $\mathbf{x} = [x_1, x_2, x_3]^T$ and desired outputs $\mathbf{t} = [t_1, t_2, t_3]^T$, is presented to the following neural network. What is the usual sequence of events for training the network using the backpropagation algorithm?

- A. (1) calculate $y_j = f(H_j)$, (2) calculate $z_k = f(I_k)$, (3) update w_{kj}, (4) update v_{ji}.

- B. (1) calculate $y_j = f(H_j)$, (2) calculate $z_k = f(I_k)$, (3) update v_{ji}, (4) update w_{kj}.

- C. (1) calculate $y_j = f(H_j)$, (2) update v_{ji}, (3) calculate $z_k = f(I_k)$, (4) update w_{kj}.

- D. (1) calculate $z_k = f(I_k)$, (2) update w_{kj}, (3) calculate $y_j = f(H_j)$, (4) update v_{ji}.

23 Multi-Layer Feedforward Networks

After some training, the units in the neural network of question 22 have the following weight vectors:

$$\mathbf{v}_1 = \begin{bmatrix} -0.7 \\ 1.8 \\ 2.3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1.2 \\ -0,6 \\ 2.1 \end{bmatrix}, \quad \mathbf{w}_1 = \begin{bmatrix} 1.0 \\ -3.5 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 0.5 \\ -1.2 \end{bmatrix} \quad \text{and} \quad \mathbf{w}_3 = \begin{bmatrix} 0.3 \\ 0.6 \end{bmatrix}.$$

Assume that all units have sigmoid activation functions given by

$$f(x) = \frac{1}{1 + \exp(-x)}$$

and that each unit has a bias $\theta = 0$ (zero). If the network is tested with an input vector $\mathbf{x} = [2.0, 3.0, 1.0]^T$ then the output of the first hidden neuron y_1 will be

- A. -2.1000

- B. 0.1091

- C. 0.5000

- D. 0.9982

- E. 6.3000

(*Hint:* on some calculators, $\exp(x) = e^x$ where $e = 2.7182818$)

24 Multi-Layer Feedforward Networks

For the same neural network described in questions 22 and 23, the output of the second hidden neuron y_2 will be

    A. -2.1000

    B. 0.1091

    C. 0.5000

    D. 0.9982

    E. 6.3000

(Assume exactly the same weights, activation functions, bias values and input vector as described in the previous question.)

25 Multi-Layer Feedforward Networks

For the same neural network described in questions 22 and 23, the output of the first output neuron z_1 will be

    A. 0.0570

    B. 0.2093

    C. 0.5902

    D. 0.5910

    E. 0.6494

(Assume exactly the same weights, activation functions, bias values and input vector as in question 23.)

26 Multi-Layer Feedforward Networks

For the same neural network described in questions 22 and 23, the output of the third output neuron z_3 will be

    A. 0.0570

    B. 0.2093

    C. 0.5902

    D. 0.5910

    E. 0.6494

(Assume exactly the same weights, activation functions, bias values and input vector as in question 23.)
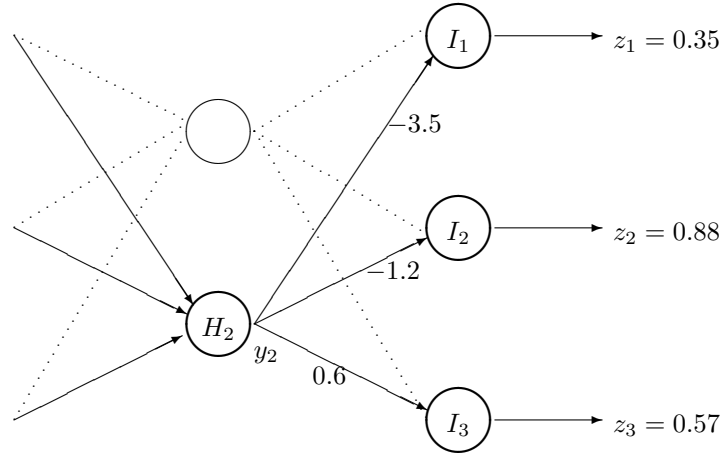
27 Multi-Layer Feedforward Networks

The following figure shows part of the neural network described in questions 22 and 23. In this question, a new input pattern is presented to the network and training continues as follows. The actual outputs of the network are given by $\mathbf{z} = [0.35, 0.88, 0.57]^T$ and the corresponding target outputs are given by $\mathbf{t} = [1.00, 0.00, 1.00]^T$. The weights w_{12}, w_{22} and w_{32} are also shown below.

For the output units, the Generalized Delta Rule can be written as

$$\triangle w_{kj} = \eta \delta_k y_j$$

where

$$\delta_k = f'(I_k)(t_k - z_k)$$

where $\triangle w_{kj}$ is the change to the weight from unit j to unit k, η is the learning rate, δ_k is the error for unit k, and $f'(net)$ is the derivative of the activation function $f(net)$.

For the sigmoid activation function given in question 23, the derivative can be rewritten as

$$f'(I_k) = f(I_k)[1 - f(I_k)].$$

What is the error for each of the output units?

A. $\delta_{\mathsf{output\_1}} = 0.4225$, $\delta_{\mathsf{output\_2}} = -0.1056$, and $\delta_{\mathsf{output\_3}} = 0.1849$.

B. $\delta_{\mathsf{output\_1}} = 0.1479$, $\delta_{\mathsf{output\_2}} = -0.0929$, and $\delta_{\mathsf{output\_3}} = 0.1054$.

C. $\delta_{\mathsf{output\_1}} = -0.4225$, $\delta_{\mathsf{output\_2}} = 0.1056$, and $\delta_{\mathsf{output\_3}} = -0.1849$.

D. $\delta_{\mathsf{output\_1}} = -0.1479$, $\delta_{\mathsf{output\_2}} = 0.0929$, and $\delta_{\mathsf{output\_3}} = -0.1054$.

(Assume exactly the same weights, activation functions and bias values as described in question 23.)

28   Multi-Layer Feedforward Networks

For the hidden units of the same network, the Generalized Delta Rule can be written as

$$\triangle v_{ji} = \eta \delta_j x_i$$

where

$$\delta_j = f'(H_j) \sum_k \delta_k w_{kj}$$

where $\triangle v_{ji}$ is the change to the weight from unit i to unit j, η is the learning rate, δ_j is the error for unit j, and $f'(net)$ is the derivative of the activation function $f(net)$.

For the sigmoid activation function given in question 23, the derivative can be rewritten as

$$f'(H_j) = f(H_j)[1 - f(H_j)].$$

What is the error for hidden unit 2 given that its activation for the pattern being processed is currently $y_2 = 0.74$?

   A. $\delta_{\texttt{hidden\_2}} = -0.2388$

   B. $\delta_{\texttt{hidden\_2}} = -0.0660$

   C. $\delta_{\texttt{hidden\_2}} = 0.0000$

   D. $\delta_{\texttt{hidden\_2}} = 0.0660$

   E. $\delta_{\texttt{hidden\_2}} = 0.2388$

(Assume exactly the same weights, activation functions and bias values as described in question 23, and exactly the same output vectors **t** and **z** as described in the previous question.)

29   Multi-Layer Feedforward Networks

Which of the following techniques is NOT a strategy for dealing with local minima in the backpropagation algorithm?

   A. Add random noise to the weights or input vectors during training.

   B. Train using the Generalized Delta Rule with momentum.

   C. Train and test using the hold-one-out strategy.

   D. Test with a committee of networks.

30   Multi-Layer Feedforward Networks

Training with the "1-of-M" coding is best explained as follows:

   A. Set the actual output to 1 for the correct class, and set all of the other actual outputs to 0.

   B. Set the actual outputs to the posterior probabilities for the different classes.

   C. Set the target output to 1 for the correct class, and set all of the other target outputs to 0.

   D. Set the target outputs to the posterior probabilities for the different classes.

31 Multi-Layer Feedforward Networks

Which of the following statements is the best description of *underfitting*?

    A. The network becomes "specialized" and learns the training set too well.

    B. The network can predict the correct outputs for test examples which lie outside the range of the training examples.

    C. The network does not contain enough adjustable parameters to find a good approximation to the unknown function which generated the training data.

    D. None of the above answers.

32 Multi-Layer Feedforward Networks

Which of the following statements is the best description of *interpolation*?

    A. The network becomes "specialized" and learns the training set too well.

    B. The network can predict the correct outputs for test examples which lie outside the range of the training examples.

    C. The network does not contain enough adjustable parameters to find a good approximation to the unknown function which generated the training data.

    D. None of the above answers.

33 Recurrent Artificial Neural Networks

Why is the *context layer* important in a simple recurrent network (SRN)?

    A. It allows the network to discover useful representations in the hidden layer.

    B. It allows the network to remove redundant information from the input vector.

    C. It allows the network to store information about previous input vectors.

    D. It allows the network to extrapolate on the patterns contained in the training set.

34 Recurrent Artificial Neural Networks

A simple recurrent network (SRN) has 5 input units, 4 hidden units and 3 output units. How many weights does this network have?

    A. 20

    B. 28

    C. 32

    D. 48

    E. 64

35   Genetic Algorithms

Which of the following statements is the best description of *reproduction*?

    A. Randomly modify the strings using ranking.

    B. Randomly change a small part of some strings.

    C. Randomly pick strings to make the next generation.

    D. Randomly generate the initial values for the strings.

    E. Randomly combine the genetic information from two strings.

36   Genetic Algorithms

Which of the following statements is the best description of *cross-over*?

    A. Randomly modify the strings using ranking.

    B. Randomly change a small part of some strings.

    C. Randomly pick strings to make the next generation.

    D. Randomly generate the initial values for the strings.

    E. Randomly combine the genetic information from two strings.

37   Genetic Algorithms

Which of the following statements is the best description of *mutation*?

    A. Randomly modify the strings using ranking.

    B. Randomly change a small part of some strings.

    C. Randomly pick strings to make the next generation.

    D. Randomly generate the initial values for the strings.

    E. Randomly combine the genetic information from two strings.

38   Genetic Algorithms

The *weighted roulette wheel* is a technique used for

    A. selecting the best chromosome.

    B. randomly selecting the chromosomes.

    C. crossing-over the selected chromosomes.

    D. mutating the fittest chromosomes.

    E. measuring the fitness of the chromosomes.

39 Genetic Algorithms

Elitism is a technique used for

    A. copying the fittest member of the population into the mating pool.

    B. obtaining the selection probabilities for reproduction.

    C. allowing many similar individuals to survive into the next generation.

    D. deleting undesirable members of the population.

40 Unsupervised Learning

Is the following statement true or false? "A cluster is a group of patterns that have similar feature values."

    A. TRUE.

    B. FALSE.

41 Unsupervised Learning

A self-organizing feature map (SOFM) has 8 input units, and 100 output units arranged in a two-dimensional grid. How many weights does this network have?

    A. 80

    B. 100

    C. 800

    D. 1000

    E. 1500

42 Unsupervised Learning

Which of the following statements is NOT true for *hard competitive learning* (HCL)?

    A. There is no target output in HCL.

    B. There are no hidden units in a HCL network.

    C. During testing, HCL is equivalent to a Bayes optimal classifier.

    D. The input vectors are often normalized to have unit length — that is, $\parallel \mathbf{x} \parallel = 1$.

    E. During training, the weights of the winning unit are adapted to be more similar to the input vector.
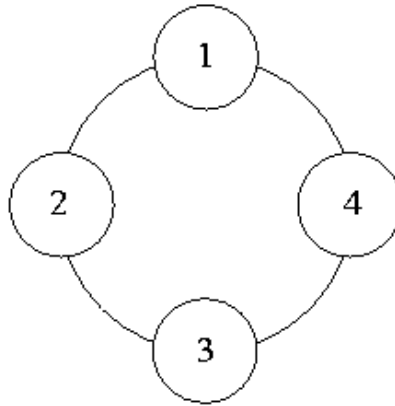
43 Unsupervised Learning

Which of the following statements is NOT true for a self-organizing feature map (SOFM)?

    A. The size of the neighbourhood is decreased during training.

    B. The units are arranged in a regular geometric pattern such as a square or ring.

    C. The learning rate is a function of the distance of the adapted units from the winning unit.

    D. The weights of the winning unit k are adapted by $\triangle \mathbf{w}_k = \eta \, (\mathbf{x} - \mathbf{w}_k)$, where $\mathbf{x}$ is the input vector.

    E. The weights of the neighbours j of the winning unit are adapted by $\triangle \mathbf{w}_j = \eta_j \, (\mathbf{x} - \mathbf{w}_j)$, where $\eta_j > \eta$ and $j \neq k$.

44 Unsupervised Learning

A self-organizing feature map has four cluster units arranged in a one-dimensional ring, as shown in the following diagram:

The weight vectors of the four units are given as follows:

$$\begin{aligned}
\mathbf{w}_1 &= [-1.00, -1.50, 0.50]^T \\
\mathbf{w}_2 &= [2.00, -2.00, 5.20]^T \\
\mathbf{w}_3 &= [1.50, 6.00, 4.30]^T \\
\mathbf{w}_4 &= [-4.00, 7.00, 0.60]^T
\end{aligned}$$

An input vector $\mathbf{x} = [-1.40, 2.30, 0.20]^T$ is presented to the network. Which unit is nearest to $\mathbf{x}$ in terms of Euclidean distance?

A. Unit 1.

B. Unit 2.

C. Unit 3.

D. Unit 4.

45 Unsupervised Learning

Adapt the weight vector of the winning unit in question 44 according to the SOFM learning algorithm with a learning rate of 0.5, using the same input vector as before. What is the new weight vector?

A. $\mathbf{w}_{\texttt{winner}} = [-2.70, 4.65, 0.40]^T$

B. $\mathbf{w}_{\texttt{winner}} = [-1.20, 0.40, 0.35]^T$

C. $\mathbf{w}_{\texttt{winner}} = [0.05, 4.15, 2.25]^T$

D. $\mathbf{w}_{\texttt{winner}} = [0.30, 0.15, 2.70]^T$

46 Unsupervised Learning

Adapt the weight vectors of the neighbours of the winning unit in question 44 according to the SOFM learning algorithm with a learning rate of 0.2, using the same input vector as before. What are the new weight vectors for the two units?

    A. $\mathbf{w_{neighbour1}} = [-3.48, 6.06, 0.52]^T$ and $\mathbf{w_{neighbour2}} = [1.32, -1.14, 4.20]^T$

    B. $\mathbf{w_{neighbour1}} = [-3.48, 6.06, 0.52]^T$ and $\mathbf{w_{neighbour2}} = [0.92, 5.26, 3.48]^T$

    C. $\mathbf{w_{neighbour1}} = [-1.08, -0.74, 0.44]^T$ and $\mathbf{w_{neighbour2}} = [0.92, 5.26, 3.48]^T$

    D. $\mathbf{w_{neighbour1}} = [-1.08, -0.74, 0.44]^T$ and $\mathbf{w_{neighbour2}} = [1.32, -1.14, 4.20]^T$

47 Unsupervised Learning

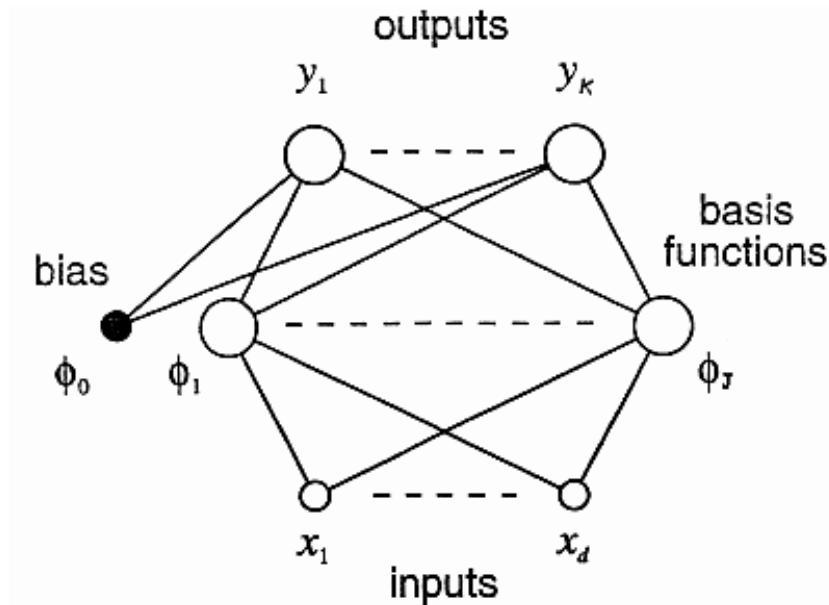What is the *topological mapping* in a self-organizing feature map (SOFM)?

    A. A map which organizes the robots and tells them where to go.

    B. A mapping where similar inputs produce similar outputs, which preserves the probability distribution of the training data.

    C. An approximation of a continuous function, which maps the input vectors onto their posterior probabilities.

    D. A mapping where similar inputs produce different outputs, which preserves the possibility distribution of the training data.

48 Radial Basis Function Networks

Which one of the following statements is NOT true in the comparison of RBF networks and multi-layer perceptrons (MLPs)?

    A. The decision boundaries formed by the hidden units are hyper-planes for the MLP and hyper-spheres for an RBF network.

    B. The supervised training is highly non-linear with problems of local minima for the MLP, but linear for an RBF network.

    C. All of the parameters in an RBF network are determined at the same time using a global training strategy, but a MLP network is often trained in two stages (supervised and unsupervised learning).

    D. The training of an RBF network is usually faster than that of a MLP.

    E. RBF and MLP networks are both able to approximate arbitrary non-linear functions.

49 Radial Basis Function Networks

The above figure shows a two-layer radial basis function (RBF) network with Gaussian radial basis functions in the hidden units defined by

$$\phi_j(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{m}_j\|^2}{2\sigma_j^2}\right)$$

where $\mathbf{m}_j$ is the mean vector and σ_j^2 is the variance for hidden unit j, $\mathbf{x}$ is the input vector, and $\|\mathbf{a} - \mathbf{b}\|$ refers to the Euclidean distance between two vectors $\mathbf{a}$ and $\mathbf{b}$. What is the correct sequence for training this network?

(i) Decide the number of basis functions.

(ii) Calculate the variance σ_j^2 from the training data for each cluster j.

(iii) Calculate the mean vectors $\mathbf{m}_j$ using an unsupervised learning algorithm such as the SOFM.

(iv) Train the weights for the output units using a supervised learning algorithm such as the Delta rule.

A. (i), (ii), (iii), (iv).

B. (i), (iii), (ii), (iv).

C. (ii), (i), (iii), (iv).

D. (ii), (iii), (iv), (i).

50 Radial Basis Function Networks

A radial basis function (RBF) network has 1 input unit, 10 radial basis function units and 5 output units. How many adjustable parameters does this network have?

A. 50

B. 60

C. 70

D. 80

E. 90

51 Associative Memory

Which of the following statements is NOT true for a Hopfield network?

    A. The output of the units is often specified by a bipolar step function.

    B. The weight matrix is symmetric — that is, $w_{ij} = w_{ji}$ for all units i and j.

    C. A unit cannot be connected to itself — that is, $w_{ii} = 0$ for all units i.

    D. The Hopfield network minimizes an energy function during recall.

    E. The error function can be defined as $E = -\frac{1}{2} \sum_i \sum_j w_{ij} S_i S_j$.

52 Associative Memory

A Hopfield network has 10 units. How many adjustable parameters does this network have?

    A. 45

    B. 50

    C. 90

    D. 100

    E. 362880

53 Associative Memory

Calculate the weight matrix for a Hopfield network to store the pattern $\begin{bmatrix} 1 & -1 & 1 & -1 & 1 \end{bmatrix}$.

A. $\mathbf{W} = \begin{bmatrix} -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 & -1 \end{bmatrix}$

B. $\mathbf{W} = \begin{bmatrix} 0 & 1 & -1 & 1 & -1 \\ 1 & 0 & 1 & -1 & 1 \\ -1 & 1 & 0 & 1 & -1 \\ 1 & -1 & 1 & 0 & 1 \\ -1 & 1 & -1 & 1 & 0 \end{bmatrix}$

C. $\mathbf{W} = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 \end{bmatrix}$

D. $\mathbf{W} = \begin{bmatrix} 0 & -1 & 1 & -1 & 1 \\ -1 & 0 & -1 & 1 & -1 \\ 1 & -1 & 0 & -1 & 1 \\ -1 & 1 & -1 & 0 & -1 \\ 1 & -1 & 1 & -1 & 0 \end{bmatrix}$

54 Associative Memory

Calculate the weight matrix for a Hopfield network to store two patterns $\begin{bmatrix} 1 & -1 & 1 & -1 & 1 \end{bmatrix}$ and $\begin{bmatrix} -1 & -1 & 1 & 1 & 1 \end{bmatrix}$.

A. $\mathbf{W} = \begin{bmatrix} 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & -2 & 0 & -2 \\ 0 & -2 & 0 & 0 & 2 \\ -2 & 0 & 0 & 0 & 0 \\ 0 & -2 & 2 & 0 & 0 \end{bmatrix}$

B. $\mathbf{W} = \begin{bmatrix} 0 & 0 & 0 & -2 & 2 \\ 0 & 0 & -2 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 & -2 \\ 2 & 0 & 0 & -2 & 0 \end{bmatrix}$

C. $\mathbf{W} = \begin{bmatrix} 0 & 2 & -2 & 0 & -2 \\ 2 & 0 & -2 & 0 & -2 \\ -2 & -2 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & -2 & 2 & 0 & 0 \end{bmatrix}$

D. $\mathbf{W} = \begin{bmatrix} 0 & 0 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & -2 \\ 2 & 0 & 0 & -2 & 0 \\ -2 & 0 & -2 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 \end{bmatrix}$

55 Applications

Which type of artificial neural network was used as an adaptive filter for echo cancellation in telephone circuits?

 A. Linear feedforward network.

 B. Self-organizing feature map (SOFM).

 C. Multi-layer feedforward network.

 D. Simple recurrent network (SRN).

 E. The Hopfield network.

56 Applications

Which type of artificial neural network was used to learn the angle to an odour source using gas sensors on a mobile robot?

 A. Linear feedforward network.

 B. Self-organizing feature map (SOFM).

 C. Multi-layer feedforward network.

 D. Simple recurrent network (SRN).

 E. The Hopfield network.

57 Applications

Which type of artificial neural network was used to detect abnormalities in microscopic images of breast tissue?

    A. Linear feedforward network.

    B. Self-organizing feature map (SOFM).

    C. Multi-layer feedforward network.

    D. Simple recurrent network (SRN).

    E. The Hopfield network.

58 Applications

Which type of artificial neural network can be used to recover a clean version of a stored image given a noisy version of that image?

    A. Linear feedforward network.

    B. Self-organizing feature map (SOFM).

    C. Multi-layer feedforward network.

    D. Simple recurrent network (SRN).

    E. The Hopfield network.

59 Applications

Which type of artificial neural network was used to control the ALVINN autonomous land vehicle?

    A. Linear feedforward network.

    B. Self-organizing feature map (SOFM).

    C. Multi-layer feedforward network.

    D. Simple recurrent network (SRN).

    E. The Hopfield network.

60 Applications

Which type of artificial neural network was used for learning simple behaviours such as obstacle avoidance and wall following on a mobile robot?

    A. Linear feedforward network.

    B. Self-organizing feature map (SOFM).

    C. Multi-layer feedforward network.

    D. Simple recurrent network (SRN).

    E. The Hopfield network.

Answers

1.		B	+1	$-\frac{1}{2}$	
2.		C	+1	$-\frac{1}{2}$	
3.		A	+1	$-\frac{1}{2}$	
4.		A	+1	$-\frac{1}{2}$	
5.		B	+1	$-\frac{1}{2}$	
6.		A	+1	$-\frac{1}{3}$	
7.		B	+1	-1	
8.		D	+1	$-\frac{1}{3}$	
9.		B	+1	$-\frac{1}{3}$	
10.		E	+1	$-\frac{1}{4}$	

11.		B	+1	$-\frac{1}{2}$	
12.		B	+1	-1	
13.		C	+1	$-\frac{1}{4}$	
14.		C	+1	$-\frac{1}{4}$	
15.		B	+1	$-\frac{1}{4}$	
16.		B	+1	-1	
17.		A	+1	-1	
18.		A	+1	-1	
19.		D	+1	$-\frac{1}{4}$	
20.		A	+1	$-\frac{1}{3}$	

21.		C	+1	$-\frac{1}{3}$	
22.		A	+1	$-\frac{1}{3}$	
23.		D	+1	$-\frac{1}{4}$	
24.		B	+1	$-\frac{1}{4}$	
25.		E	+1	$-\frac{1}{4}$	
26.		C	+1	$-\frac{1}{4}$	
27.		B	+1	$-\frac{1}{3}$	
28.		B	+1	$-\frac{1}{4}$	
29.		C	+1	$-\frac{1}{3}$	
30.		C	+1	$-\frac{1}{3}$	

31.		C	+1	$-\frac{1}{3}$	
32.		D	+1	$-\frac{1}{3}$	
33.		C	+1	$-\frac{1}{3}$	
34.		D	+1	$-\frac{1}{4}$	
35.		C	+1	$-\frac{1}{4}$	
36.		E	+1	$-\frac{1}{4}$	
37.		B	+1	$-\frac{1}{4}$	
38.		B	+1	$-\frac{1}{4}$	
39.		A	+1	$-\frac{1}{3}$	
40.		A	+1	-1	

41.		C	+1	$-\frac{1}{4}$	
42.		C	+1	$-\frac{1}{4}$	
43.		E	+1	$-\frac{1}{4}$	
44.		A	+1	$-\frac{1}{3}$	
45.		B	+1	$-\frac{1}{3}$	
46.		A	+1	$-\frac{1}{3}$	
47.		B	+1	$-\frac{1}{3}$	
48.		C	+1	$-\frac{1}{4}$	
49.		B	+1	$-\frac{1}{3}$	
50.		D	+1	$-\frac{1}{4}$	

51.		E	+1	$-\frac{1}{4}$	
52.		A	+1	$-\frac{1}{4}$	
53.		D	+1	$-\frac{1}{3}$	
54.		A	+1	$-\frac{1}{3}$	
55.		A	+1	$-\frac{1}{4}$	
56.		D	+1	$-\frac{1}{4}$	
57.		B	+1	$-\frac{1}{4}$	
58.		E	+1	$-\frac{1}{4}$	
59.		C	+1	$-\frac{1}{4}$	
60.		A	+1	$-\frac{1}{4}$	