Artificial Neural Networks
Examination, August 2004

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 a *feature* in a pattern classification problem?

A. An output variable.

B. An input variable.

C. A hidden variable or weight.

2 Pattern Recognition

What is a *decision boundary* in a pattern classification problem with two input variables?

A. A histogram defined for each image in the training set.

B. A line or curve which separates two different classes.

C. A plane or hypercurve defined in the space of possible inputs.

3 Pattern Recognition

What is a *statistically optimal classifier*?

A. A classifier which calculates the nearest neighbour for a given test example.

B. A classifier which gives the lowest probability of making classification errors.

C. A classifier which minimises the sum squared error on the training set.

4 Classical Pattern Recognition

How are the *prototype vectors* calculated in a minimum distance classifier?

A. For each class, calculate the mean vector from the input vectors in the training set.

B. For each class, calculate the covariance matrix from the input vectors in the training set.

C. For each class, calculate the prior probability from the input vectors in the training set.

5 Classical Pattern Recognition

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

$$\text{Class 1:} \quad \begin{bmatrix} -2.0 \\ -1.0 \end{bmatrix}, \begin{bmatrix} -2.0 \\ -3.0 \end{bmatrix} \qquad \text{Class 2:} \quad \begin{bmatrix} 2.0 \\ 1.0 \end{bmatrix}, \begin{bmatrix} 2.0 \\ -1.0 \end{bmatrix} \qquad \text{Class 3:} \quad \begin{bmatrix} -2.0 \\ 1.0 \end{bmatrix}, \begin{bmatrix} -2.0 \\ 3.0 \end{bmatrix}$$

Then classify the test vector $[1.0, -2.0]^T$ with the trained classifier. Which class does this vector belong to?

A. Class 1.

B. Class 2.

C. Class 3.

6 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 ω_j. What is the value of the decision function for each of the three classes in question 5 for the test vector $[0, -1.0]^T$?

A. $d_1(\mathbf{x}) = -2.0$, $d_2(\mathbf{x}) = -2.0$, $d_3(\mathbf{x}) = -6.0$.

B. $d_1(\mathbf{x}) = -6.0$, $d_2(\mathbf{x}) = -2.0$, $d_3(\mathbf{x}) = -2.0$.

C. $d_1(\mathbf{x}) = -1.5$, $d_2(\mathbf{x}) = -3.5$, $d_3(\mathbf{x}) = -3.5$.

D. $d_1(\mathbf{x}) = -3.5$, $d_2(\mathbf{x}) = -1.5$, $d_3(\mathbf{x}) = -9.5$.

7 Classical Pattern Recognition

Give the equation of the decision line between classes 2 and 3 for the minimum distance classifier in question 5.

A. $x_2 = 0$

B. $x_2 = -2x_1 - 1$

C. $x_2 = 2x_1 + 1$

8 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.

9 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.

10 Training and Testing

What is *generalization*?

A. The ability of a pattern recognition system to approximate the desired output values for pattern vectors which are not included in the test set.

B. The ability of a pattern recognition system to approximate the desired output values for pattern vectors which are not included in the training set.

C. The ability of a pattern recognition system to extrapolate on pattern vectors which are not included in the training set.

D. The ability of a pattern recognition system to interpolate on pattern vectors which are not included in the test set.

11 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.

12 Biological Neurons

What are the advantages of biological neural networks (BNNs) compared to conventional Von Neumann computers?

(i) BNNs have the ability to learn from examples.
(ii) BNNs have a high degree of parallelism.
(iii) BNNs require a mathematical model of the problem.
(iv) BNNs can acquire knowledge by "trial and error".
(v) BNNs use a sequential algorithm to solve problems.

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

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

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

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

E. (i), (iv) and (v).

13 Biological Neurons

Which of the following statements is the best description of Hebb'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."

14 Artificial Neural Networks

Which of the following statements is the best description of the Delta 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."

15 Artificial Neural Networks

Which of the following techniques can NOT be used for pre-processing the inputs to an artificial neural network?

A. Normalization.

B. Winner-takes-all.

C. Fast Fourier Transform (FFT).

D. Principal component analysis (PCA).

E. Deleting outliers from the training set.

16 Artificial Neural Networks

A neuron with 4 inputs has the weight vector $\mathbf{w} = [1, 2, 3, 4]^T$ and a bias $\theta = 0$ (zero). The activation function is linear, where the constant of proportionality equals 2 — that is, the activation function is given by $f(net) = 2 \times net$. If the input vector is $\mathbf{x} = [4, 8, 5, 6]^T$ then the output of the neuron will be

A. 1.

B. 56.

C. 59.

D. 112.

E. 118.

17 Artificial Neural Networks

Which of the following neural networks uses supervised learning?

A. Self-organizing feature map (SOFM).

B. The Hopfield network.

C. Simple recurrent network (SRN).

D. All of the above answers.

E. None of the above answers.

18 Artificial Neural Networks

Which of the following algorithms can be used to train a single-layer feedforward network?

A. Hard competitive learning.

B. Soft competitive learning.

C. A genetic algorithm.

D. All of the above answers.

E. None of the above answers.

19 Artificial Neural Networks

What is the biggest difference between Widrow & Hoff's *Delta Rule* and the *Perceptron Learning Rule* for learning in a single-layer feedforward network?

A. There is no difference.

B. The Delta Rule is defined for step activation functions, but the Perceptron Learning Rule is defined for linear activation functions.

C. The Delta Rule is defined for sigmoid activation functions, but the Perceptron Learning Rule is defined for linear activation functions.

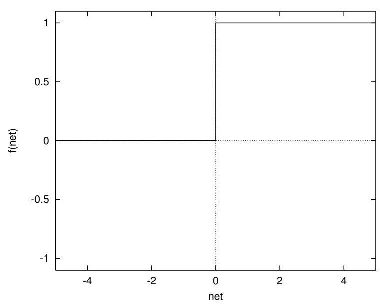
D. The Delta Rule is defined for linear activation functions, but the Perceptron Learning Rule is defined for step activation functions.

E. The Delta Rule is defined for sigmoid activation functions, but the Perceptron Learning Rule is defined for step activation functions.

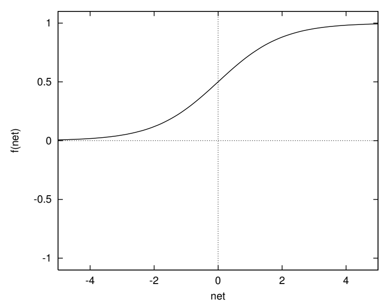
20 Artificial Neural Networks

Identify each of the following activation functions.

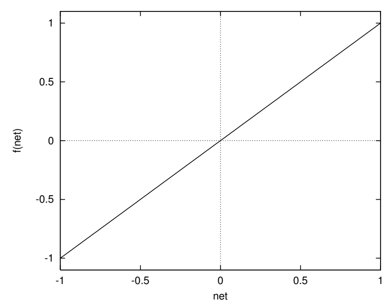
(i)



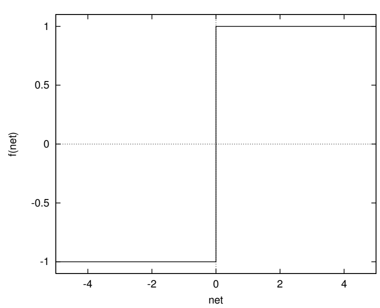
(ii)



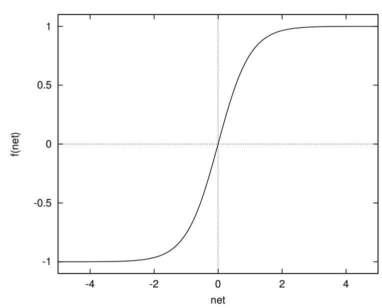
(iii)



(iv)



(v)



A. (i) Unipolar step [`hardlim`], (ii) Unipolar sigmoid [`logsig`], (iii) Linear [`purelin`], (iv) Bipolar step [`hardlims`], (v) Bipolar sigmoid [`tansig`].

B. (i) Unipolar step [`hardlim`], (ii) Bipolar sigmoid [`tansig`], (iii) Linear [`purelin`], (iv) Bipolar step [`hardlims`], (v) Unipolar sigmoid [`logsig`].

C. (i) Bipolar step [`hardlims`], (ii) Unipolar sigmoid [`logsig`], (iii) Linear [`purelin`], (iv) Unipolar step [`hardlim`], (v) Bipolar sigmoid [`tansig`].

D. (i) Bipolar step [`hardlims`], (ii) Bipolar sigmoid [`tansig`], (iii) Linear [`purelin`], (iv) Unipolar step [`hardlim`], (v) Unipolar sigmoid [`logsig`].

21 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 0 (zero). Does the perceptron give the correct answer (that is, is the actual output the same as the desired output)?

A. Yes.

B. No.

22 Perceptrons

The perceptron in question 21 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. $w_1 = 0.2, w_2 = -0.3, \theta = -0.4$.

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

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

D. $w_1 = 0.4, w_2 = -0.5, \theta = -0.4$.

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

23 Perceptrons

A single-layer perceptron has 6 input units and 3 output units. How many weights does this network have?

A. 6

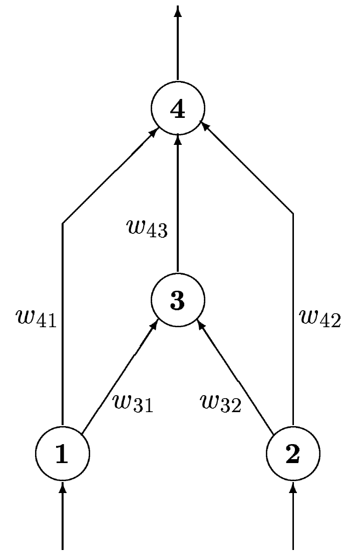
B. 9

C. 18

D. 25

24 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 $w_{43} = -2$. 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.

25 Perceptrons

Would it be possible to train the multi-layer perceptron in question 24 to solve the XOR problem using the back-propagation algorithm?

A. Yes.

B. No.

26 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.

27 Multi-Layer Feedforward Networks

Is the following statement true or false? "A multi-layer feedforward network with linear activation functions is more powerful than a single-layer feedforward network with linear activation functions."

A. TRUE.

B. FALSE.

28 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 output units.

B. The problem of adapting the neighbours of the winning unit.

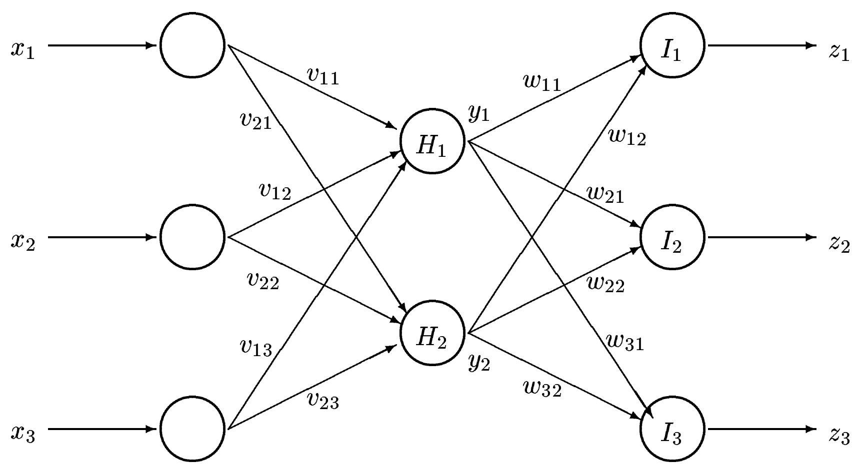
C. The problem of defining an error function for linearly inseparable problems.

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

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

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

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

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

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

30 Multi-Layer Feedforward Networks

After training, the units in the neural network of question 29 have the following weight vectors:

$$\mathbf{v}_1 = \begin{bmatrix} -2.0 \\ 2.0 \\ -2.0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1.0 \\ 1.0 \\ -1.0 \end{bmatrix}, \quad \mathbf{w}_1 = \begin{bmatrix} 1.0 \\ -3.5 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 0.5 \\ -1.2 \end{bmatrix} \quad \text{and} \quad \mathbf{w}_3 = \begin{bmatrix} 0.3 \\ 0.6 \end{bmatrix}.$$

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 second hidden neuron y_2 will be

A. 0.018

B. 0.500

C. 0.982

D. 1.000

E. 4.000

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

31 Multi-Layer Feedforward Networks

For the same neural network described in questions 29 and 30, the output of the second output neuron z_2 will be

A. -1.000

B. -0.928

C. -0.087

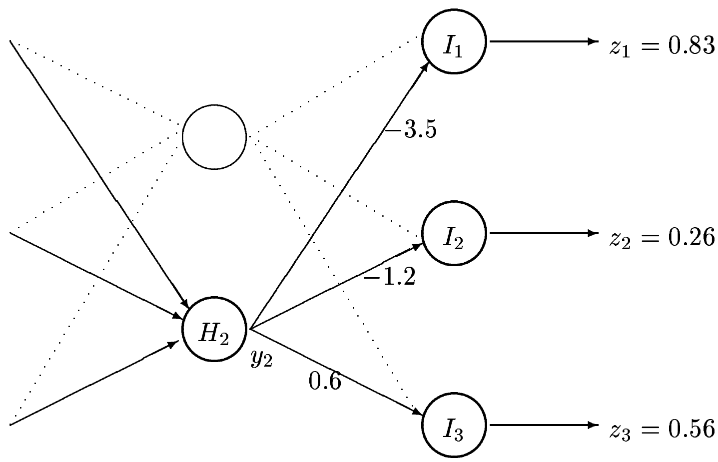
D. 0.283

E. 0.527

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

32 Multi-Layer Feedforward Networks

The following figure shows part of the neural network described in questions 29 and 30. 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 $\mathbf{z} = [0.83, 0.26, 0.56]^T$ and the corresponding target outputs are given by $\mathbf{t} = [0.58, 0.70, 0.20]^T$. 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 30, 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. $\delta_{\texttt{output\_1}} = -0.0425$, $\delta_{\texttt{output\_2}} = 0.3256$, and $\delta_{\texttt{output\_3}} = -0.1584$.

B. $\delta_{\texttt{output\_1}} = -0.0353$, $\delta_{\texttt{output\_2}} = 0.0847$, and $\delta_{\texttt{output\_3}} = -0.0887$.

C. $\delta_{\texttt{output\_1}} = 0.0425$, $\delta_{\texttt{output\_2}} = -0.3256$, and $\delta_{\texttt{output\_3}} = 0.1584$.

D. $\delta_{\texttt{output\_1}} = 0.0353$, $\delta_{\texttt{output\_2}} = -0.0847$, and $\delta_{\texttt{output\_3}} = 0.0887$.

(Assume exactly the same weights, activation functions and bias values as described in question 30.)

33 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 30, 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 $y_2 = 0.69$?

A. $\delta_{\texttt{hidden\_2}} = -0.0721$

B. $\delta_{\texttt{hidden\_2}} = -0.0067$

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

D. $\delta_{\texttt{hidden\_2}} = 0.0067$

E. $\delta_{\texttt{hidden\_2}} = 0.0721$

(Assume exactly the same weights, activation functions and bias values as described in question 30, and exactly the same output vectors **t** and **z** as described in the previous question.)

34 Multi-Layer Feedforward Networks

Which of the following equations best describes the *Generalized Delta Rule with momentum*?

A. $\triangle w_{ji}(t+1) = \eta \delta_j x_i$

B. $\triangle w_{ji}(t+1) = \alpha \delta_j x_i$

C. $\triangle w_{ji}(t+1) = \eta \delta_j x_i + \alpha \triangle w_{ji}(t)$

D. $\triangle w_{ji}(t+1) = \eta \delta_j x_i + \alpha \delta_j x_i(t)$

E. $\triangle w_{ji}(t+1) = \eta\,(x_i - w_{ji}) + \alpha \triangle w_{ji}(t)$

where $\triangle w_{ji}(t)$ is the change to the weight from unit i to unit j at time t, η is the learning rate, α is the momentum coefficient, δ_j is the error term for unit j, and x_i is the i^{th} input to unit j.

35 Multi-Layer Feedforward Networks

One method for dealing with local minima is to use a *committee* of networks. What does this mean?

A. A large number of different networks are trained and tested. The network with the lowest sum squared error on a separate validation set is chosen as the best network.

B. A large number of different networks are trained and tested. All of the networks are used to solve the real-world problem by taking the average output of all the networks.

C. A large number of different networks are trained and tested. The networks are then combined together to make a network of networks, which is biologically more realistic and computationally more powerful than a single network.

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

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

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

D. None of the above answers.

37 Multi-Layer Feedforward Networks

What is the most general type of decision region that can be formed by a feedforward network with TWO hidden layers?

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

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

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

D. None of the above answers.

38 Multi-Layer Feedforward Networks

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

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 (e.g., hidden units) to find a good approximation to the unknown function which generated the training data.

39 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.

40 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.

41 Genetic Algorithms

Which of the following applications can be implemented by a genetic algorithm?

A. Feature selection for pattern recognition problems.

B. Learning the weights in a neural network.

C. Robot path planning.

D. All of the above answers.

E. None of the above answers.

42 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.

43 Genetic Algorithms

Ranking is a technique used for

A. copying the fittest member of each 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.

44 Unsupervised Learning

Is the following statement true or false? “Clusters that are similar in some way should be far apart.”

A. TRUE.

B. FALSE.

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

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 $j \neq k$.

46 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 SOFM training algorithm is based on soft competitive learning.

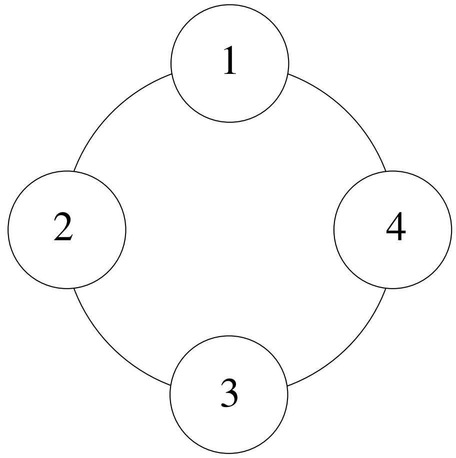
C. The cluster units are arranged in a regular geometric pattern such as a square or ring.

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

E. The network can grow during training by adding new cluster units when required.

47 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.

48 Unsupervised Learning

Adapt the weight vector of the winning unit in question 47 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. $\mathbf{w}_{\texttt{winner}} = [0.30, 0.15, 2.70]^T$

49 Unsupervised Learning

Adapt the weight vectors of the neighbours of the winning unit in question 47 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. $\mathbf{w}_{\texttt{neighbour1}} = [-3.48, 6.06, 0.52]^T$ and $\mathbf{w}_{\texttt{neighbour2}} = [1.32, -1.14, 4.20]^T$

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

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

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

50 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.

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

How many hidden layers are there in a Hopfield network?

A. None (0).

B. One (1).

C. Two (2).

53 Associative Memory

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

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

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

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

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

54 Associative Memory

The same Hopfield network in question 53 is trained to store a second pattern $\begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}$. What is the new weight matrix?

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

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

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

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

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. Multi-layer feedforward network.

C. Simple recurrent network.

D. The Hopfield network.

E. Self-organizing feature map.

56 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. Multi-layer feedforward network.

C. Simple recurrent network.

D. The Hopfield network.

E. Self-organizing feature map.

57 Applications

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

A. Linear feedforward network.

B. Multi-layer feedforward network.

C. Simple recurrent network.

D. The Hopfield network.

E. Self-organizing feature map.

58 Applications

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

A. Linear feedforward network.

B. Multi-layer feedforward network.

C. Simple recurrent network.

D. The Hopfield network.

E. Self-organizing feature map.

59 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. Multi-layer feedforward network.

C. Simple recurrent network.

D. The Hopfield network.

E. Self-organizing feature map.

60 Applications

Which type of learning system was used for gesture recognition on a mobile robot?

A. A linear feedforward network.

B. A multi-layer feedforward network.

C. A simple recurrent network.

D. A self-organising feature map.

E. A genetic algorithm.

Answers

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

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

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

31.		D	$+1$	$-\frac{1}{4}$	
32.		B	$+1$	$-\frac{1}{3}$	
33.		B	$+1$	$-\frac{1}{4}$	
34.		C	$+1$	$-\frac{1}{4}$	
35.		B	$+1$	$-\frac{1}{2}$	
36.		C	$+1$	$-\frac{1}{3}$	
37.		B	$+1$	$-\frac{1}{3}$	
38.		B	$+1$	$-\frac{1}{2}$	
39.		B	$+1$	$-\frac{1}{4}$	
40.		E	$+1$	$-\frac{1}{4}$	

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

51.		E	+1	$-\frac{1}{4}$	
52.		A	+1	$-\frac{1}{2}$	
53.		A	+1	$-\frac{1}{3}$	
54.		C	+1	$-\frac{1}{3}$	
55.		A	+1	$-\frac{1}{4}$	
56.		D	+1	$-\frac{1}{4}$	
57.		E	+1	$-\frac{1}{4}$	
58.		B	+1	$-\frac{1}{4}$	
59.		A	+1	$-\frac{1}{4}$	
60.		B	+1	$-\frac{1}{4}$	