Artificial Neural Networks
Examination, March 2005

Instructions

There are SIXTY questions. (The pass mark is 30 out of 60.)

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

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. *features.*

C. *weights.*

D. *classes.*

2 Pattern Recognition

The process described in question 1 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.*

3 Pattern Recognition

Is the following statement true or false? "In supervised learning, there is a target output vector which tells the pattern recognition system the correct class for a given input vector."

A. TRUE.

B. FALSE.

4 Pattern Recognition

Is the following statement true or false? "The decision theory approach to pattern recognition is based on decision functions. To classify an unknown pattern vector, the value of the decision function is calculated for each class. The vector is assigned to the class with the highest decision value."

A. TRUE.

B. FALSE.

5 Classical Pattern Recognition

For a minimum distance classifier with three input variables, what is the *decision boundary* between two classes?

A. A line.

B. A curve.

C. A plane.

D. A hyperplane.

E. A discriminant value.

6 Classical Pattern Recognition

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

$$\text{Class 1:} \quad \begin{bmatrix} -1.0 \\ -0.5 \end{bmatrix}, \begin{bmatrix} -1.2 \\ -1.5 \end{bmatrix}, \begin{bmatrix} -0.8 \\ -1.0 \end{bmatrix}$$

$$\text{Class 2:} \quad \begin{bmatrix} 1.3 \\ 0.5 \end{bmatrix}, \begin{bmatrix} 0.9 \\ -0.2 \end{bmatrix}, \begin{bmatrix} 0.8 \\ -0.3 \end{bmatrix}$$

$$\text{Class 3:} \quad \begin{bmatrix} -1.0 \\ 0.8 \end{bmatrix}, \begin{bmatrix} -1.0 \\ 0.9 \end{bmatrix}, \begin{bmatrix} -1.0 \\ 1.3 \end{bmatrix}$$

Then classify the test vector $[-0.5, 1.0]^T$ with the trained classifier. Which class does this vector belong to?

A. Class 1.

B. Class 2.

C. Class 3.

7 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 6 for the test vector $[0, -0.5]^T$?

A. $d_1(\mathbf{x}) = -0.875$, $d_2(\mathbf{x}) = -0.375$, $d_3(\mathbf{x}) = -2.375$.

B. $d_1(\mathbf{x}) = -0.375$, $d_2(\mathbf{x}) = -0.875$, $d_3(\mathbf{x}) = -0.875$.

C. $d_1(\mathbf{x}) = -1.5$, $d_2(\mathbf{x}) = -0.5$, $d_3(\mathbf{x}) = -0.5$.

D. $d_1(\mathbf{x}) = -0.5$, $d_2(\mathbf{x}) = -0.5$, $d_3(\mathbf{x}) = -1.5$.

8 Classical Pattern Recognition

Give the equation of the decision line between classes 2 and 3 for the minimum distance classifier in question 6.

A. $x_2 = 0$

B. $x_2 = -2x_1 - \frac{1}{2}$

C. $x_2 = 2x_1 + \frac{1}{2}$

9 Classical Pattern Recognition

For a Bayes classifier with one input variable, what is the *decision boundary* between two classes?

A. A line.

B. A curve.

C. A plane.

D. A hypercurve.

E. A discriminant value.

10 Classical Pattern Recognition

The covariance between two variables x and y can be calculated using the formula

$$COV_{xy} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N-1},$$

where μ_x refers to the mean value of variable x, and μ_y refers to the mean value of variable y. What is the covariance between the two input variables for class 2 in question 6?

A. 0.050

B. 0.095

C. 0.115

11 Classical Pattern Recognition

Is the following statement true or false? "A minimum distance classifier should be used instead of a Bayes classifier if the variation in the training vectors is large compared to the distance between classes."

A. TRUE.

B. FALSE.

12 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.

13 Training and Testing

What is an *outlier*?

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

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

C. An input pattern which produces a classification error.

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

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

14 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 strengths 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 answers is NOT a general characteristic of artificial neural networks?

A. Learning.

B. Generalization.

C. Parallel processing.

D. Robust performance.

E. Fast processor speed.

16 Artificial Neural Networks

Is the following statement true or false? "Artificial neural networks are usually synchronous, but we can simulate an asynchronous network by updating the nodes in a random order."

A. TRUE.

B. FALSE.

17 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. Rescaling to lie between 0 and 1.

D. Principal component analysis (PCA).

E. Deleting outliers from the training set.

18 Artificial Neural Networks

Which of the following artificial neural networks uses soft competitive learning?

A. The Hopfield network.

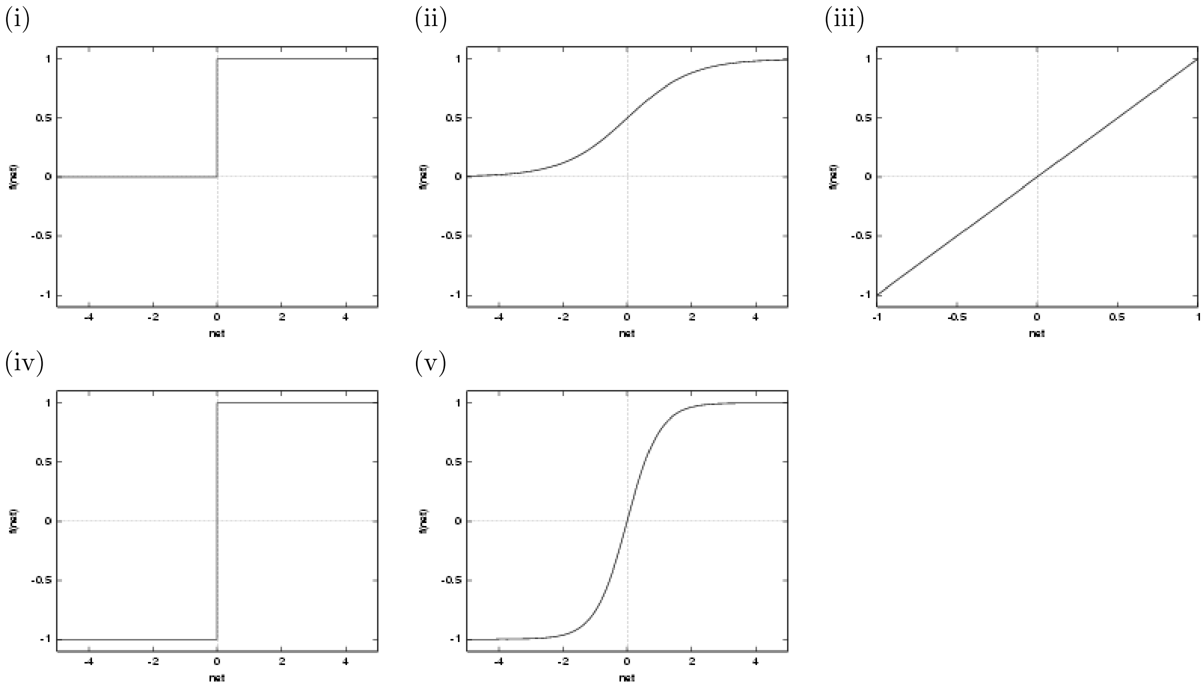
B. Simple recurrent network.

C. Self-organizing feature map.

D. Multi-layer feedforward network.

19 Artificial Neural Networks

Identify each of the following activation functions.



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`].

20 Perceptrons

Which of the following 2 input Boolean logic functions are linearly separable?

(i) AND

x_1	x_2	d
0	0	0
0	1	0
1	0	0
1	1	1

(ii) OR

x_1	x_2	d
0	0	0
0	1	1
1	0	1
1	1	1

(iii) XOR

x_1	x_2	d
0	0	0
0	1	1
1	0	1
1	1	0

(iv) NAND

x_1	x_2	d
0	0	1
0	1	1
1	0	1
1	1	0

(v) NOT XOR

x_1	x_2	d
0	0	1
0	1	0
1	0	0
1	1	1

A. (i) and (iv).

B. (iii) and (v).

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

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

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

21 Perceptrons

A perceptron with a unipolar step function has two inputs with weights $w_1 = 0.5$ and $w_2 = -0.2$, and a threshold $\theta = 0.3$ (θ can therefore be considered as a weight for an extra input which is always set to -1). For a given training example $\mathbf{x} = [0, 1]^T$, 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 $\eta = 0.5$?

A. $w_1 = 1.0, w_2 = -0.2, \theta = -0.2$.

B. $w_1 = 0.5, w_2 = 0.3, \theta = 0.7$.

C. $w_1 = 0.5, w_2 = 0.3, \theta = -0.2$.

D. $w_1 = 0.5, w_2 = -0.3, \theta = 0.2$.

E. $w_1 = 0.5, w_2 = -0.2, \theta = 0.3$.

23 Perceptrons

Is the following statement true or false? "The XOR problem can be solved by a multi-layer perceptron, but a multi-layer perceptron with step activation functions cannot learn to do this."

A. TRUE.

B. FALSE.

24 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 of weights."

A. TRUE.

B. FALSE.

25 Multi-Layer Feedforward Networks

In the backpropagation algorithm, how is the error function usually defined?

A. $\frac{1}{2} \sum_j (weight_j \times input_j)$ for all inputs j.

B. $\frac{1}{2} \sum_j (target_j - output_j)^2$ for all outputs j.

C. $\frac{1}{2} \sum_j (target_j - output_j)$ for all outputs j.

26 Multi-Layer Feedforward Networks

Is the following statement true or false? "The *credit assignment problem* in a multi-layer feedforward network is the problem of determining the number of hidden units for a given set of training data."

A. TRUE.

B. FALSE.

27 Multi-Layer Feedforward Networks

A common technique for training MLFF networks is to calculate the generalization error on a separate data set after each epoch of training. Training is stopped when the generalization error starts to increase. This technique is called

A. momentum.

B. underfitting.

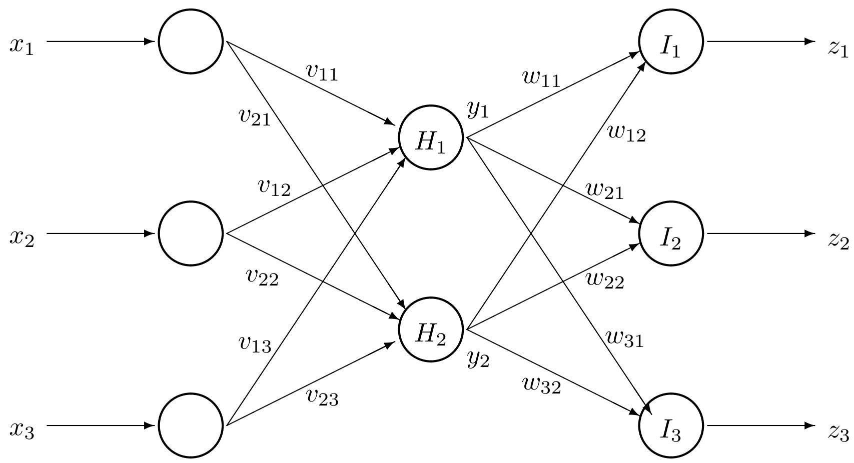
C. hold-one-out.

D. generalization.

E. early stopping.

28 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}.

29 Multi-Layer Feedforward Networks

After some training, the units in the neural network of question 28 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. 0.9982

B. 0.8909

C. 0.5000

D. 0.1091

E. 0.0018

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

30 Multi-Layer Feedforward Networks

For the same neural network described in questions 28 and 29, the output of the third output neuron z_3 will be

A. 0.6494

B. 0.5910

C. 0.5902

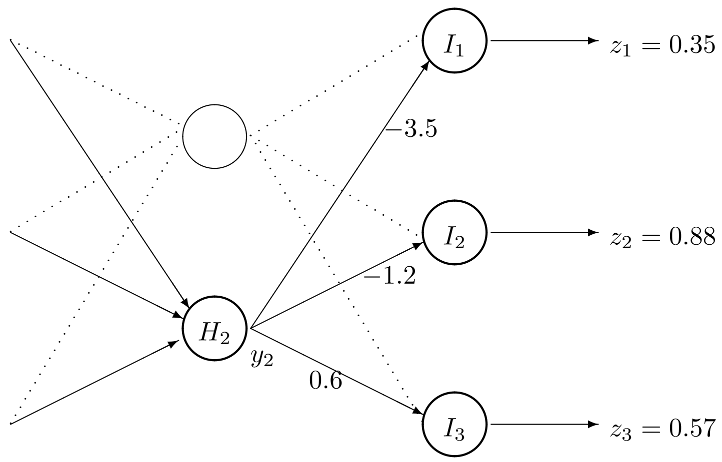
D. 0.2093

E. 0.0570

(Assume exactly the same weights, activation functions, bias values and input vector as in question 29.)

31 Multi-Layer Feedforward Networks

The following figure shows part of the neural network described in questions 28 and 29. 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 29, 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. $\delta_{\texttt{output\_1}} = 0.4225$, $\delta_{\texttt{output\_2}} = -0.1056$, and $\delta_{\texttt{output\_3}} = 0.1849$.

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

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

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

(Assume exactly the same weights, activation functions and bias values as described in question 29.)

32 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 29, 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 29, and exactly the same output vectors $\mathbf{t}$ and $\mathbf{z}$ as described in the previous question.)

33 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. Repeat the training using a new set of random weights.

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

E. Use a committee of networks.

34 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.

35 Multi-Layer Feedforward Networks

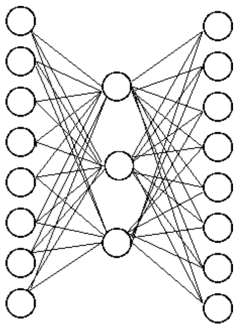
Is the following statement true or false? "Using too few hidden units in a multi-layer feedforward network can lead to overfitting of the training data."

A. TRUE.

B. FALSE.

36 Multi-Layer Feedforward Networks

This $8 \times 3 \times 8$ network was trained to learn the identity function, using the eight training examples shown below. After 5000 training epochs, the three hidden units encode the eight distinct inputs as shown.



Input	Hidden Values	Output
10000000	.89 .04 .08	10000000
01000000	.15 .99 .99	01000000
00100000	.01 .97 .27	00100000
00010000	.99 .97 .71	00010000
00001000	.03 .05 .02	00001000
00000100	.01 .11 .88	00000100
00000010	.80 .01 .98	00000010
00000001	.60 .94 .01	00000001

Which of the following statements is NOT true for this type of artificial neural network?

A. The target output pattern is the same as the input pattern.

B. A pattern is compressed by presenting it to the input layer and calculating the activations for the hidden layer.

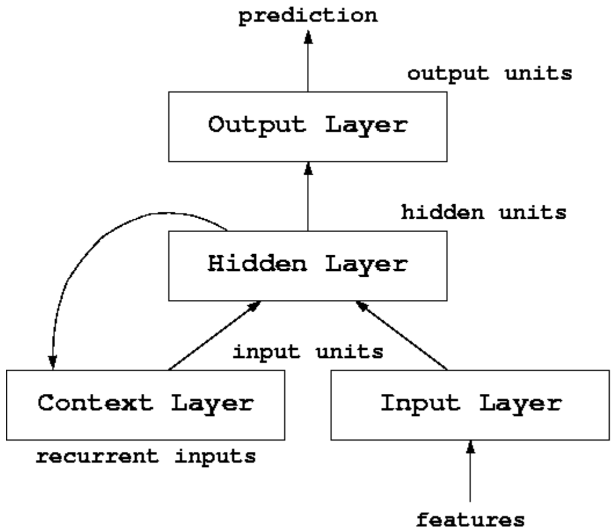
C. After successful training, the hidden units compute what is essentially the principal components of the input data.

D. No loss in information is expected, because the network will re-create perfectly all the training examples on the output layer.

E. To reconstruct the input vector, the compressed vector is presented to the hidden layer and the activations are propagated to the output layer.

37 Multi-Layer Feedforward Networks

The following schematic diagram shows an artificial neural network that can be used for processing time series data.



Which of the following statements is NOT true for this type of artificial neural network?

A. The test examples must be presented to the network in the correct order.

B. The training examples must be presented to the network in the correct order.

C. The number of context units should be the same as the number of input units.

D. This type of network can predict the next chunk of data in the series from the past history of data.

E. The hidden units encode an internal representation of the data in the series that precedes the current input.

38 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.

39 Unsupervised Learning

A self-organizing feature map (SOFM) has 15 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

40 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, $\| \mathbf{x} \| = 1$.

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

41 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.

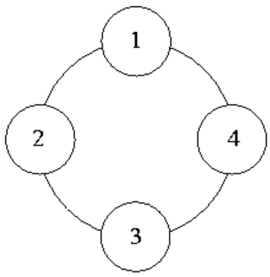
C. The learning rate is a function of the distance of the adapted units from the winning unit.

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 $j \neq k$.

42 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.

43 Unsupervised Learning

Adapt the weight vector of the winning unit in question 42 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. $\mathbf{w}_{\texttt{winner}} = [0.30, 0.15, 2.70]^T$

44 Unsupervised Learning

Adapt the weight vectors of the neighbours of the winning unit in question 42 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 $\mathbf{w}_{\texttt{neighbour2}} = [1.32, -1.14, 4.20]^T$

45 Unsupervised Learning

What is the biggest difference between a self-organizing map (SOM) and an ensemble of SOMs (ESOM) for classification?

A. In a SOM, the squared Euclidean distance is used to determine the winning unit.

B. In a SOM, there are two phases during training: soft competitive learning and fine tuning.

C. In an ESOM, a separate SOM is trained for each class in the training data.

D. In an ESOM, the number of cluster units can be increased by adding new units if the distribution of the input data is non-stationary.

46 Associative Memory

A Hopfield network has 63 units. How many adjustable parameters does this network contain?

A. 0.

B. 63.

C. 1953.

D. 3906.

E. 3969.

47 Associative Memory

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

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

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

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

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

48 Associative Memory

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

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

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

C. $\mathbf{W} = \begin{bmatrix} 0 & -1 & 1 \\ -1 & 0 & -3 \\ 1 & -3 & 0 \end{bmatrix}$

D. $\mathbf{W} = \begin{bmatrix} 0 & 3 & -3 \\ 3 & 0 & 1 \\ -3 & 1 & 0 \end{bmatrix}$

49 Associative Memory

The theoretical maximum number of patterns that could be stored by a Hopfield network is given as

$$p_{max} = \frac{N}{2 \ln N}$$

where N is the number of network units. So what is the theoretical maximum number of patterns that could be stored by the network in question 48?

A. 1.098

B. 1.365

C. 7.603

D. 8.286

50 Genetic Algorithms

The English moth *Biston betularia* provides a good example of natural evolution. In this moth there are two colour types, light and dark. In 1848, dark moths formed less than 2% of the population. But in 1898, which was the time of England's industrial revolution, 95% of the moths in Manchester and other highly industrialized areas were of the dark type. Why?

A. The mutant moths had light-sensitive skin cells which were able to change colour and adapt to the bright electric street-lights in the industrial cities.

B. Soot from factories collected on the trees the moths lived in, so birds could see the light moths better and ate more of them. As a result, more dark moths survived and produced offspring.

C. The weight for the dark-light neuron in the moth chromosome was increased by Hebbian learning, so the Manchester moths gradually changed colour from light to dark.

D. Pavlov's dog became colour-blind as a result of the stimulus-response experiments being performed in Manchester. The dog couldn't see the dark moths, so he ate all of the light moths.

51 Genetic Algorithms

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

A. Randomly change a small part of some strings.

B. Randomly change the fitness function.

C. Randomly combine the genetic information from 2 strings.

D. Randomly generate small initial values for the weights.

E. Randomly pick strings to make the next generation.

52 Genetic Algorithms

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

A. Randomly change a small part of some strings.

B. Randomly change the fitness function.

C. Randomly combine the genetic information from 2 strings.

D. Randomly generate small initial values for the weights.

E. Randomly pick strings to make the next generation.

53 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.

54 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.

55 Genetic Algorithms

Is the following statement true or false? "A genetic algorithm could be used to search the space of possible weights for an artificial neural network, without requiring any gradient information."

A. TRUE.

B. FALSE.

56 Genetic Algorithms

Is the following statement true or false? "A genetic algorithm (global search) could be combined with a gradient-based algorithm (local search) to produce a hybrid GA. The hybrid algorithm could perform better than both the global and local search methods on their own."

A. TRUE.

B. FALSE.

57 Applications

Which application in intelligent mobile robotics made use of a single-layer feedforward network?

A. Path planning.

B. Route following.

C. Gesture recognition.

D. Obstacle avoidance and wall following.

E. Estimating the direction to an odour source.

58 Applications

Which application in intelligent mobile robotics made use of a multi-layer feedforward network?

A. Path planning.

B. Route following.

C. Gesture recognition.

D. Obstacle avoidance and wall following.

E. Estimating the direction to an odour source.

59 Applications

Which application in intelligent mobile robotics made use of a simple recurrent network?

A. Path planning.

B. Route following.

C. Gesture recognition.

D. Obstacle avoidance and wall following.

E. Estimating the direction to an odour source.

60 Applications

Which application in intelligent mobile robotics made use of a self-organizing feature map?

A. Path planning.

B. Route following.

C. Gesture recognition.

D. Obstacle avoidance and wall following.

E. Estimating the direction to an odour source.

Answers

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

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

21.		A	+1	-1	
22.		C	+1	$-\frac{1}{4}$	
23.		A	+1	-1	
24.		A	+1	-1	
25.		B	+1	$-\frac{1}{2}$	
26.		B	+1	-1	
27.		E	+1	$-\frac{1}{4}$	
28.		A	+1	$-\frac{1}{3}$	
29.		A	+1	$-\frac{1}{4}$	
30.		C	+1	$-\frac{1}{4}$	

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

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

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