

# Artificial Neural Networks

## Examination, March 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 *class* in a pattern recognition problem?

- A. An input variable.
- B. A hidden variable or weight.
- C. An output variable.

### 2 Pattern Recognition

Is the following statement true or false? “In unsupervised 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.

### 3 Classical Pattern Recognition

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

- A. A line.
- B. A plane.
- C. A curve.
- D. A hyperplane.
- E. A hypercurve.

### 4 Classical Pattern Recognition

For a Bayes classifier with three input variables, what is the shape of the *decision boundary* between two classes?

- A. A line.
- B. A plane.
- C. A curve.
- D. A hyperplane.
- E. A hypercurve.

## 5 Classical Pattern Recognition

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

$$\text{Class 1: } \begin{bmatrix} 12.0 \\ 97.0 \end{bmatrix}, \begin{bmatrix} 34.0 \\ 76.0 \end{bmatrix}, \begin{bmatrix} 74.0 \\ 52.0 \end{bmatrix} \quad \text{Class 2: } \begin{bmatrix} 52.0 \\ 22.0 \end{bmatrix}, \begin{bmatrix} 79.0 \\ 64.0 \end{bmatrix}, \begin{bmatrix} 100.0 \\ 16.0 \end{bmatrix}$$

What are the prototype vectors of the trained classifier?

- A.  $\mathbf{m}_1 = \begin{bmatrix} 40.0 \\ 75.0 \end{bmatrix}, \mathbf{m}_2 = \begin{bmatrix} 77.0 \\ 27.0 \end{bmatrix}.$
- B.  $\mathbf{m}_1 = \begin{bmatrix} 40.0 \\ 77.0 \end{bmatrix}, \mathbf{m}_2 = \begin{bmatrix} 75.0 \\ 34.0 \end{bmatrix}.$
- C.  $\mathbf{m}_1 = \begin{bmatrix} 40.0 \\ 75.0 \end{bmatrix}, \mathbf{m}_2 = \begin{bmatrix} 77.0 \\ 34.0 \end{bmatrix}.$

## 6 Classical Pattern Recognition

Classify the test vector  $[66.0, 75.0]^T$  with the trained classifier in question 5. Which class does this vector belong to?

- A. Class 1.
- B. Class 2.

## 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  $j$ . What is the value of the decision function for each of the classes in question 5 for the test vector  $[-10.0, 40.0]^T$ ?

- A.  $d_1(\mathbf{x}) = -1012.5, d_2(\mathbf{x}) = -2952.5.$
- B.  $d_1(\mathbf{x}) = -1084.5, d_2(\mathbf{x}) = -2780.5.$
- C.  $d_1(\mathbf{x}) = -1012.5, d_2(\mathbf{x}) = -3019.0.$

## 8 Classical Pattern Recognition

Give the equation of the decision boundary between classes 1 and 2 for the minimum distance classifier in question 5.

- A.  $x_2 = 0.7708x_1 + 5.9063$
- B.  $x_2 = 0.8140x_1 + 8.6977$
- C.  $x_2 = 0.9024x_1 + 1.7073$

## 9 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  $\mu_x$  refers to the mean value of variable  $x$ , and  $\mu_y$  refers to the mean value of variable  $y$ . What is the covariance between the two input variables for class 2 in question 5?

- A.  $-27.0$
- B.  $-522.0$
- C.  $-702.0$

## 10 Classical Pattern Recognition

A Bayes classifier with normally distributed classes is trained using the same data as the minimum distance classifier in question 5. Give the covariance matrix of the trained Bayes classifier for class 2.

- A.  $\begin{bmatrix} 579.0 & -27.0 \\ -27.0 & 684.0 \end{bmatrix}$
- B.  $\begin{bmatrix} 624.0 & -522.0 \\ -522.0 & 1629.0 \end{bmatrix}$
- C.  $\begin{bmatrix} 988.0 & -702.0 \\ -702.0 & 507.0 \end{bmatrix}$

## 11 Training and Testing

A classification system may be trained on one data set and tested on a second data set. The system designer can then experiment with different parameters in the classifier, for example, the number of hidden units in an artificial neural network. For real world applications, it is therefore important to use a *third* data set to evaluate the final performance of the system. Why?

- A. The error on the third data set is used to choose between lots of different possible systems.
- B. The error on the third data set provides a better (unbiased) estimate of the true generalization error.
- C. It's not important – testing on the second data set indicates the generalization performance of the system.

## 12 Biological Neurons

What are the advantages of biological neural networks (BNNs) compared to conventional 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) and (iii).
- B. (i), (ii) and (iv).
- C. (i), (iii) and (iv).
- D. (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 strengths of the neurons involved are modified to reduce the error between the desired and actual outputs of the system.”

## 14 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.

## 15 Artificial Neural Networks

Which of the following equations is the best description of Hebbian learning?

- A.  $\Delta \mathbf{w}_k = \eta y_k \mathbf{x}$ ,
- B.  $\Delta \mathbf{w}_k = \eta (d_k - y_k) \mathbf{x}$ ,
- C.  $\Delta \mathbf{w}_k = \eta (\mathbf{x} - \mathbf{w}_k)$ ,
- D.  $\Delta \mathbf{w}_j = \eta_j (\mathbf{x} - \mathbf{w}_j)$ , where  $\eta_j < \eta$  and  $j \neq k$ ,

where  $\mathbf{x}$  is the input vector,  $\eta$  is the learning rate,  $\mathbf{w}_k$  is the weight vector,  $d_k$  is the target output, and  $y_k$  is the actual output for unit  $k$ .

## 16 Artificial Neural Networks

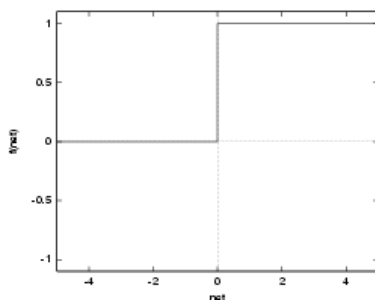
Which of the following algorithms can be used to train a feedforward neural network with a single layer of weights?

- A. Hard competitive learning.
- B. Soft competitive learning.
- C. The Delta rule.
- D. A genetic algorithm.
- E. All of the above answers.

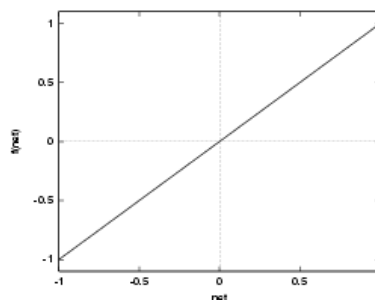
## 17 Artificial Neural Networks

Identify each of the following activation functions.

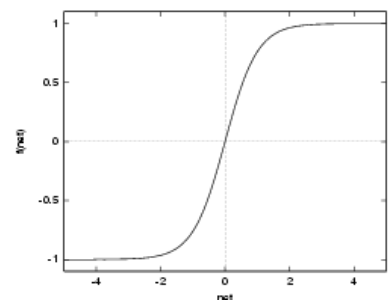
(i)



(ii)



(iii)



- A. (i) Unipolar step, (ii) Unipolar sigmoid, (iii) Linear.
- B. (i) Unipolar step, (ii) Linear, (iii) Bipolar step.
- C. (i) Unipolar sigmoid, (ii) Linear, (iii) Bipolar sigmoid.
- D. (i) Unipolar step, (ii) Linear, (iii) Bipolar sigmoid.
- E. (i) Bipolar step, (ii) Linear, (iii) Bipolar sigmoid.

## 18 Perceptrons

A simple perceptron has two input units, a unipolar step function, weights  $w_1 = 0.2$  and  $w_2 = -0.5$ , and a threshold  $\theta = -0.2$  ( $\theta$  can therefore be considered as a weight for an extra input which is always set to -1). What is the actual output of this perceptron for the input pattern  $\mathbf{x} = [1, 1]^T$ ?

- A. -1.
- B. 0.
- C. 1.

## 19 Perceptrons

The perceptron in question 18 is trained using the learning rule  $\Delta \mathbf{w} = \eta (d - y) \mathbf{x}$ , where  $\mathbf{x}$  is the input vector,  $\eta$  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.  $w_1 = 0.4, w_2 = -0.5, \theta = -0.4$ .

## 20 Perceptrons

What is the equation of the decision line for the perceptron in question 18 (before training)?

- A.  $x_2 = 0$
- B.  $x_2 = 0.667x_1$
- C.  $x_2 = 0.400x_1 + 0.400$
- D.  $x_2 = 0.667x_1 + 1.333$
- E.  $x_2 = 0.800x_1 + 0.800$

## 21 Perceptrons

What is the equation of the decision line for the perceptron in question 19 (after training)?

- A.  $x_2 = 0$
- B.  $x_2 = 0.667x_1$
- C.  $x_2 = 0.400x_1 + 0.400$
- D.  $x_2 = 0.667x_1 + 1.333$
- E.  $x_2 = 0.800x_1 + 0.800$

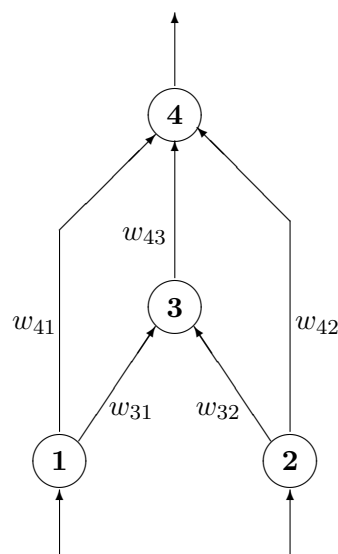
## 22 Perceptrons

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

- A. 6
- B. 9
- C. 18
- D. 25

## 23 Perceptrons

In the following network, 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.

## 24 Perceptrons

Is the following statement true or false? “A multi-layer perceptron with separate layers of weights could perform exactly the same function as the neural network in question 23.”

- A. TRUE.
- B. FALSE.

## 25 ADALINE

Which of the following statements is NOT true for an ADALINE network?

- A. There are separate phases for training and testing.
- B. The weight vector is constantly adapted.
- C. The outputs are continuous valued.
- D. The inputs are continuous valued.
- E. The activation function is linear.

## 26 Multi-Layer Feedforward Networks

A multi-layer feedforward network has 5 input units, a first hidden layer with 4 units, a second hidden layer with 3 units, and 2 output units. How many weights does this network have?

- A. 14
- B. 18
- C. 20
- D. 26
- E. 38

## 27 Multi-Layer Feedforward Networks

Is the following statement true or false? “The Delta rule solves the credit assignment problem in the training of multi-layer feedforward networks.”

- A. TRUE.
- B. FALSE.

## 28 Multi-Layer Feedforward Networks

Why is it not a good idea to have linear activation functions in the hidden units of a multi-layer feedforward network?

- A. This type of function is smooth but non-decreasing.
- B. It is not possible to calculate the derivative for this type of function.
- C. A multi-layer network with this type of activation functions would be no more powerful than a single-layer network.
- D. When the errors are backpropagated, this function gives no information about the amount of adjustment required to the weights of the hidden units.

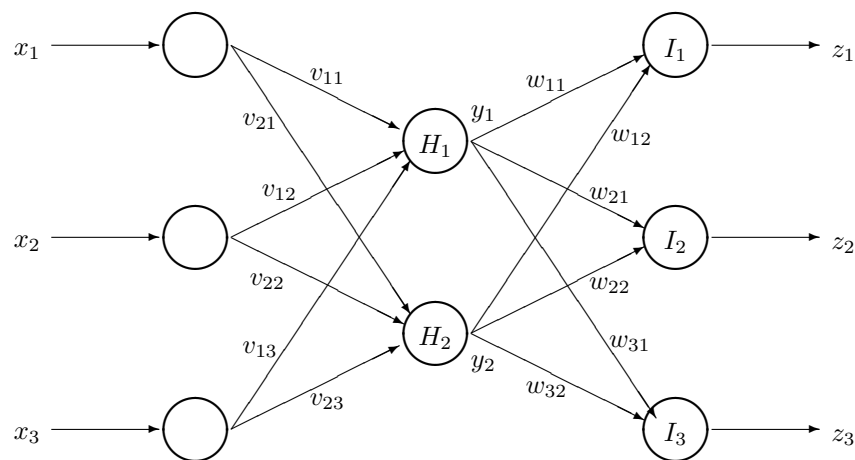
## 29 Multi-Layer Feedforward Networks

Why is it not a good idea to have step activation functions in the hidden units of a multi-layer feedforward network?

- A. This type of function is smooth but non-decreasing.
- B. It is not possible to calculate the derivative for this type of function.
- C. A multi-layer network with this type of activation functions would be no more powerful than a single-layer network.
- D. When the errors are backpropagated, this function gives no information about the amount of adjustment required to the weights of the hidden units.

## 30 Multi-Layer Feedforward Networks

Consider the following feedforward network with one hidden layer of units:



The input vector to the network is  $\mathbf{x} = [x_1, x_2, x_3]^T$ , the vector of hidden layer outputs is  $\mathbf{y} = [y_1, y_2]^T$ , the vector of actual outputs is  $\mathbf{z} = [z_1, z_2, z_3]^T$ , and the vector of desired outputs is  $\mathbf{t} = [t_1, t_2, t_3]^T$ . The network has 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  $\mathbf{x} = [1.0, 2.0, 3.0]^T$  then the activation  $H_1$  of the first hidden neuron will be

- A. -4.000
- B. 0.000
- C. 0.018
- D. 0.500
- E. 2.000

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

### 31 Multi-Layer Feedforward Networks

Assuming exactly the same neural network and the same input vector as in the previous question, what is the activation  $I_3$  of the third output neuron?

- A. -1.7320
- B. -0.5910
- C. 0.1503
- D. 0.3054
- E. 0.5758

### 32 Multi-Layer Feedforward Networks

For the output units of the network in question 30, the generalized Delta rule can be written as

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

where  $\Delta w_{kj}$  is the change to the weights from unit  $j$  to unit  $k$ ,  $\eta$  is the learning rate,  $\delta_k$  is the error term for unit  $k$ , and  $y_j$  is the  $j^{th}$  input to unit  $k$ . In the backpropagation algorithm, what is the error term  $\delta_k$ ?

- A.  $\delta_k = f'(H_j)(t_k - z_k)$ .
- B.  $\delta_k = f'(I_k) \sum_k \delta_k w_{kj}$ .
- C.  $\delta_k = f'(I_k)(t_k - z_k)$ .
- D.  $\delta_k = f'(H_j) \sum_k \delta_k w_{kj}$ .

where  $f'(net)$  is the derivative of the activation function  $f(net)$ .

### 33 Multi-Layer Feedforward Networks

For the hidden units of the network in question 30, the generalized Delta rule can be written as

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

where  $\Delta v_{ji}$  is the change to the weights from unit  $i$  to unit  $j$ ,  $\eta$  is the learning rate,  $\delta_j$  is the error term for unit  $j$ , and  $x_i$  is the  $i^{th}$  input to unit  $j$ . In the backpropagation algorithm, what is the error term  $\delta_j$ ?

- A.  $\delta_j = f'(H_j)(t_k - z_k)$ .
- B.  $\delta_j = f'(I_k) \sum_k \delta_k w_{kj}$ .
- C.  $\delta_j = f'(I_k)(t_k - z_k)$ .
- D.  $\delta_j = f'(H_j) \sum_k \delta_k w_{kj}$ .

where  $f'(net)$  is the derivative of the activation function  $f(net)$ .

### 34 Multi-Layer Feedforward Networks

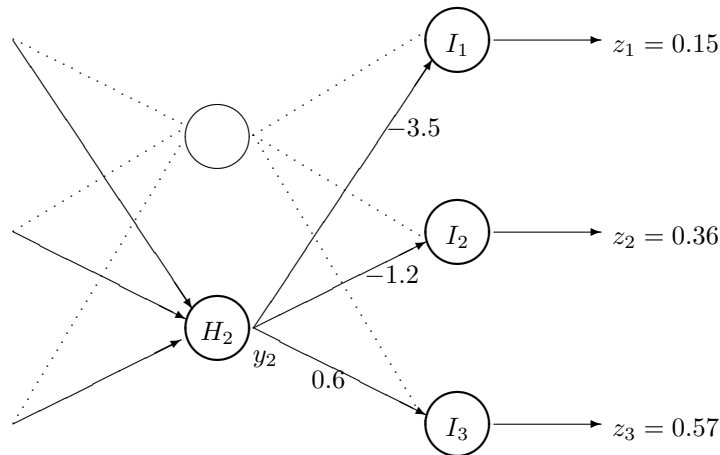
Which of the following equations best describes the generalized Delta rule with momentum?

- A.  $\Delta^p w_{kj}(t+1) = \eta \delta_k y_j(t)$
- B.  $\Delta^p w_{kj}(t+1) = \alpha \delta_k y_j(t)$
- C.  $\Delta^p w_{kj}(t+1) = \eta \delta_k y_j + \alpha \Delta^p w_{kj}(t)$
- D.  $\Delta^p w_{kj}(t+1) = \eta \delta_k y_j + \alpha \delta_k y_j(t)$
- E.  $\Delta^p w_{kj}(t+1) = \eta \delta_k y_j + \alpha f'(H_j) y_j(t)$

where  $\Delta w_{kj}^p(t)$  is the change to the weights from unit  $j$  to unit  $k$  at time  $t$  when a training pattern  $p$  is presented to the network,  $\eta$  is the learning rate,  $\alpha$  is the momentum coefficient,  $\delta_k$  is the error term for unit  $k$ , and  $y_j$  is the  $j^{th}$  input to unit  $k$ .

### 35 Multi-Layer Feedforward Networks

The following figure shows part of the neural network described in question 30. A new input pattern is presented to the network and training proceeds as follows. The actual outputs of the network are given by  $\mathbf{z} = [0.15, 0.36, 0.57]^T$  and the corresponding target outputs are given by  $\mathbf{t} = [1.00, 1.00, 1.00]^T$ . The weights  $w_{12}$ ,  $w_{22}$  and  $w_{32}$  are also shown below.



For the output units, the derivative of the sigmoid function 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.  $\delta_{\text{output}_1} = -0.4225$ ,  $\delta_{\text{output}_2} = 0.1056$ , and  $\delta_{\text{output}_3} = -0.1849$ .
- B.  $\delta_{\text{output}_1} = -0.1084$ ,  $\delta_{\text{output}_2} = -0.1475$ , and  $\delta_{\text{output}_3} = -0.1054$ .
- C.  $\delta_{\text{output}_1} = 0.1084$ ,  $\delta_{\text{output}_2} = 0.1475$ , and  $\delta_{\text{output}_3} = 0.1054$ .
- D.  $\delta_{\text{output}_1} = 0.4225$ ,  $\delta_{\text{output}_2} = -0.1056$ , and  $\delta_{\text{output}_3} = 0.1849$ .

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

### 36 Multi-Layer Feedforward Networks

For the hidden units, the derivative of the sigmoid function 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  $y_2 = 0.50$ ?

- A.  $\delta_{\text{hidden.2}} = 0.2388$
- B.  $\delta_{\text{hidden.2}} = 0.1475$
- C.  $\delta_{\text{hidden.2}} = 0.0534$
- D.  $\delta_{\text{hidden.2}} = -0.1233$
- E.  $\delta_{\text{hidden.2}} = -0.1672$

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

### 37 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 during training.
- B. Add random noise to the training examples at each epoch of training.
- C. Repeat the training process several times with new random weights, and pick the network that gives the best generalization performance.
- D. Train and test a committee or ensemble of networks, then use the average of the network outputs.
- E. After each epoch of training, test the network with an independent data set and stop the training when the generalization error starts to increase.

### 38 Multi-Layer Feedforward Networks

Which of the following statements is NOT true for an autoassociative feedforward network with a single hidden layer of neurons?

- A. During training, the target output vector is the same as the input vector.
- B. The trained network can be split into two machines: the first layer of weights compresses the input pattern (decoder), and the second layer of weights reconstructs the full pattern (encoder).
- C. After training, the hidden units give a representation that is equivalent to the principal components of the training data, removing redundant parts of the input data.
- D. The data compression can be useful for pre-processing the data for training another neural network.
- E. It is important to use smooth, non-linear activation functions in the hidden units.

### 39 Unsupervised Learning

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

- A. There is no target output in HCL.
- B. The neurons have unipolar step functions.
- C. During testing, HCL is equivalent to a minimum distance 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  $k$  are adapted by  $\Delta \mathbf{w}_k = \eta (\mathbf{x} - \mathbf{w}_k)$ , where  $\mathbf{x}$  is the input vector.

### 40 Unsupervised Learning

Which of the following statements is NOT true for a *self-organising feature map* (SOM)?

- A. There are two phases during training: soft competitive learning and fine tuning.
- B. The topological mapping in a well trained SOM preserves the probability distribution of the training data.
- C. The units are arranged in a regular geometric structure such as a 2-dimensional square or a 1-dimensional ring.
- D. The number of cluster units can be increased by adding new units if the distribution of the input data changes over time.
- E. The learning rate for the neighbours is a decreasing function of the distance from the winning unit in the geometric structure.

### 41 Unsupervised Learning

An input vector  $\mathbf{x}$  and two prototype vectors  $\mathbf{p}_1$  and  $\mathbf{p}_2$  are given by

$$\begin{aligned}\mathbf{x} &= [-1.40, 2.30, 0.20]^T \\ \mathbf{p}_1 &= [-1.00, 2.20, 0.10]^T \\ \mathbf{p}_2 &= [-4.00, 7.00, 0.60]^T\end{aligned}$$

Which prototype is nearest to  $\mathbf{x}$  in terms of squared Euclidean distance?

- A.  $\mathbf{p}_1$
- B.  $\mathbf{p}_2$

### 42 Unsupervised Learning

Adapt the weight vector of the winning prototype in question 41 according to the SOM learning algorithm with a learning rate of 0.5. What is the new weight vector?

- A.  $\mathbf{p}_{\text{winner}} = [-0.80, 2.15, 0.05]^T$
- B.  $\mathbf{p}_{\text{winner}} = [-1.20, 2.25, 0.15]^T$
- C.  $\mathbf{p}_{\text{winner}} = [-2.70, 4.65, 0.40]^T$
- D.  $\mathbf{p}_{\text{winner}} = [-5.30, 9.35, 0.80]^T$

### 43 Unsupervised Learning

A SOM can be used for fault detection in an industrial process. The network is trained using only examples of the “normal” state of the process. During testing, a new input pattern  $\mathbf{x}$  is presented to the network. Which method is used to decide whether there is a possible fault in the process?

- A. There may be a fault if the squared Euclidean distance from the input pattern  $\mathbf{x}$  to the best matching unit in the SOM is high.
- B. There may be a fault if the squared Euclidean distance from the input pattern  $\mathbf{x}$  to the best matching unit in the SOM is low.

### 44 Unsupervised Learning

Is the following statement true or false? “An ensemble of SOMs (ESOM) can be used as a classifier. One SOM is trained for each class in the training data. During testing, a new input pattern  $\mathbf{x}$  is presented to all of the SOMs, and the SOM with the highest distance to its best matching unit will be the winner.”

- A. TRUE.
- B. FALSE.

### 45 Associative Memory

How many hidden layers are there in an autoassociative Hopfield network?

- A. None (0).
- B. One (1).
- C. Two (2).

### 46 Associative Memory

A Hopfield network has 20 units. How many adjustable parameters does this network contain?

- A. 95
- B. 190
- C. 200
- D. 380
- E. 400

## 47 Associative Memory

Calculate the weight matrix for a Hopfield network to store the pattern  $[1 \quad -1 \quad 1 \quad -1]$ .

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

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

## 48 Associative Memory

Calculate the weight matrix for a Hopfield network to store two patterns  $[1 \quad -1 \quad 1 \quad -1]$  and  $[-1 \quad -1 \quad -1 \quad 1]$ .

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

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

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

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

## 49 Associative Memory

The trained Hopfield network in question 48 is used for recall. A pattern  $[-1 \quad -1 \quad 1 \quad -1]$  is presented to the network, then the nodes of the network are updated until a steady state is reached. What is the final state of the network?

A.  $[-1 \quad -1 \quad 1 \quad -1]$

B.  $[1 \quad -1 \quad 1 \quad -1]$

C.  $[-1 \quad -1 \quad -1 \quad 1]$

D.  $[-1 \quad 1 \quad -1 \quad 1]$

## 50 Genetic Algorithms

Is the following statement true or false? “A genetic algorithm will always increase the overall fitness of the population.”

- A. TRUE.
- B. FALSE.

## 51 Genetic Algorithms

Which genetic operator can introduce new genetic information into the population?

- A. Crossover.
- B. Elitism.
- C. Mutation.
- D. Reproduction.
- E. Sharing.

## 52 Genetic Algorithms

Which genetic operator is used to decrease the fitness of individuals that are too similar to other members of the population?

- A. Crossover.
- B. Elitism.
- C. Mutation.
- D. Reproduction.
- E. Sharing.

## 53 Genetic Algorithms

Which genetic operator is used to combine the genetic information from two parents to produce two offspring?

- A. Crossover.
- B. Elitism.
- C. Mutation.
- D. Reproduction.
- E. Sharing.

## 54 Laborations

What does the following command in MATLAB do?

```
>> y = sim(net, x);
```

- A. Initialize a neural network `net` using the input vectors `x`.
- B. Train a neural network `net` using the input vectors `x`.
- C. Test a neural network `net` using the input vectors `x`.
- D. Calculate the generalization error of a neural network `net` using the input vectors `x`.

## 55 Applications

Is the following statement true or false? “Learning produces changes within an agent that over time enables it to perform more effectively within its environment.”

- A. TRUE.
- B. FALSE.

## 56 Applications

Which type of learning system is used for adaptive noise cancellation in telephone circuits?

- 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.

## 57 Applications

Which type of learning system can be used for path planning in mobile robots?

- 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.

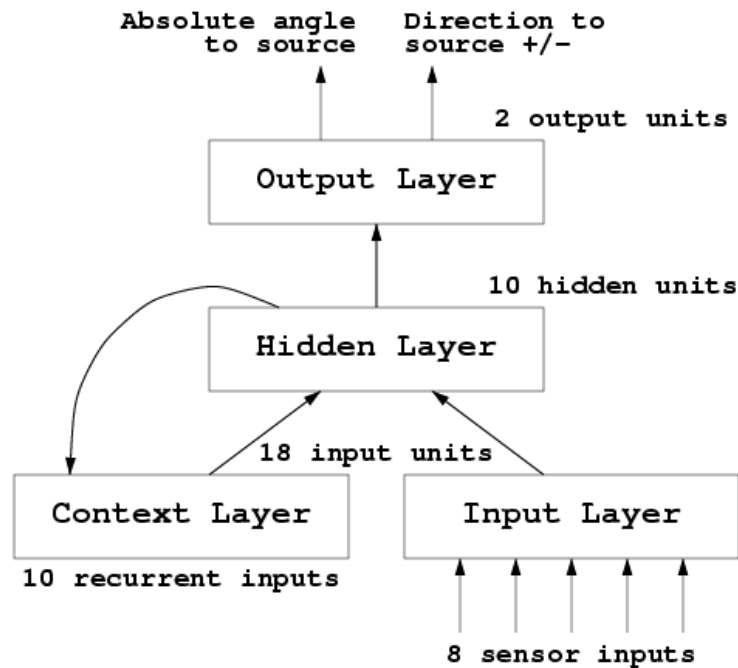
## 58 Applications

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.

## 59 Applications

The following schematic diagram shows the neural network that was used to train a mobile robot with an electronic nose to estimate the direction to an odour source.



How many weights does this network contain?

- A. 100.
- B. 120.
- C. 180.
- D. 200.

## 60 Applications

What is the usual sequence of events for training the neural network in question 59?

- (i) Calculate the activation of the units in the output layer.
- (ii) Calculate the activation of the units in the hidden layer.
- (iii) Calculate the activation of the units in the context layer.
- (iv) Copy the outputs of the hidden layer to the context layer.
- (v) Copy the outputs of the output layer to the context layer.
- (vi) Update the weights from the hidden units to the output units.
- (vii) Update the weights from the input units to the hidden units.
- (viii) Update the weights from the hidden units to the context units.

- A. (i), (ii), (vi), (iv), (vii).
- B. (ii), (i), (vi), (vii), (iv).
- C. (ii), (i), (vi), (vii), (viii).
- D. (ii), (i), (vi), (vii), (v).
- E. (iii), (ii), (i), (vii), (vi).

## Answers

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

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

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

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

41.		A	+1	-1	
42.		B	+1	$-\frac{1}{3}$	
43.		A	+1	-1	
44.		B	+1	-1	
45.		A	+1	$-\frac{1}{2}$	
46.		B	+1	$-\frac{1}{4}$	
47.		D	+1	$-\frac{1}{3}$	
48.		D	+1	$-\frac{1}{3}$	
49.		B	+1	$-\frac{1}{3}$	
50.		B	+1	-1	

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