Artificial Neural Networks
Examination, June 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

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.*

D. none of the above answers.

3 Pattern Recognition

What is a *pattern vector*?

A. A vector of weights $\mathbf{w} = [w_1, w_2, ..., w_n]^T$ in a neural network.

B. A vector of measured features $\mathbf{x} = [x_1, x_2, ..., x_n]^T$ of an input example.

C. A vector of outputs $\mathbf{y} = [y_1, y_2, ..., y_n]^T$ of a classifier.

4 Classical Pattern Recognition

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

A. A line.

B. A curve.

C. A plane.

D. A hyperplane.

E. A discriminant.

5 Classical Pattern Recognition

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

A. A line.

B. A curve.

C. A plane.

D. A hypercurve.

E. A discriminant.

6 Classical Pattern Recognition

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

$$\text{Class 1:} \quad \begin{bmatrix} -1.0 \\ -0.5 \end{bmatrix}, \begin{bmatrix} -1.0 \\ -1.5 \end{bmatrix} \qquad \text{Class 2:} \quad \begin{bmatrix} 1.0 \\ 0.5 \end{bmatrix}, \begin{bmatrix} 1.0 \\ -0.5 \end{bmatrix} \qquad \text{Class 3:} \quad \begin{bmatrix} -1.0 \\ 0.5 \end{bmatrix}, \begin{bmatrix} -1.0 \\ 1.5 \end{bmatrix}$$

Then classify the test vector $[0.5, -1]^T$ 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. $d_1(\mathbf{x}) = -0.5$, $d_2(\mathbf{x}) = -0.5$, $d_3(\mathbf{x}) = -1.5$.

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

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

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

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

Give the equation of the decision line between classes 1 and 3 for the minimum distance classifier in question 6.

A. $x_2 = 0$

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

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

10 Training and Testing

Is the following statement true or false? "An outlier is an input pattern that is very different from the typical patterns of the same class".

A. TRUE.

B. FALSE.

11 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 in the test set.

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

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

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

12 Biological Neurons

Is the following statement true or false? "In the human brain, roughly 10% of the neurons are used for input and output. The remaining 90% are used for information processing."

A. TRUE.

B. FALSE.

13 Biological Neurons

Which of the following statements is the best description of supervised learning?

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 parallel computing devices consisting of many interconnected simple processors."

A. TRUE.

B. FALSE.

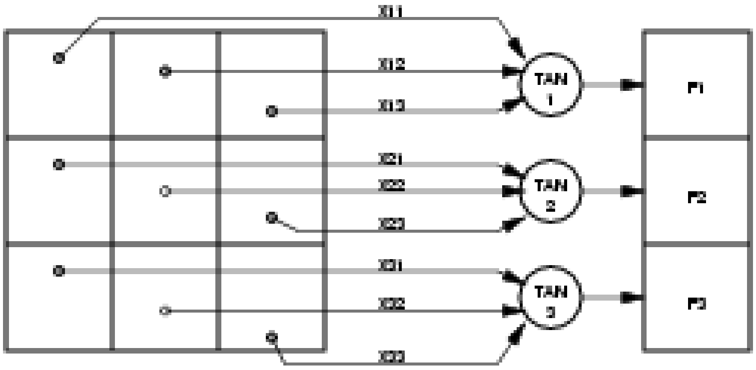
15 Artificial Neural Networks

Is the following statement true or false? "Knowledge is acquired by a neural network from its environment through a learning process, and this knowledge is stored in the connections strengths (neurons) between processing units (weights)."

A. TRUE.

B. FALSE.

16 Artificial Neural Networks



What is this network?

A. A single-layer feedforward neural network.

B. An auto-associative neural network.

C. A multi-layer feedforward neural network.

D. A Hopfield neural network.

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. A Fast Fourier Transform.

D. Principal Components Analysis (PCA).

E. Deleting outliers from the training set.

18 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.

19 Artificial Neural Networks

Which of the following neural networks uses supervised learning?

A. Simple recurrent network.

B. Self-organizing feature map.

C. Hopfield network.

D. All of the above answers.

E. None of the above answers.

20 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.

21 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.

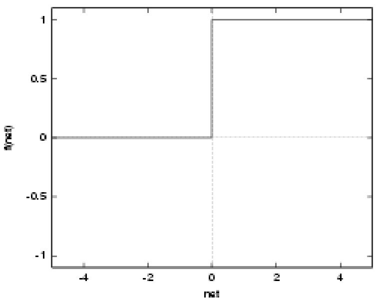
D. The Delta Rule is defined for linear activation functions, but the Perceptron Learning Rule is defined for step activation functions.

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

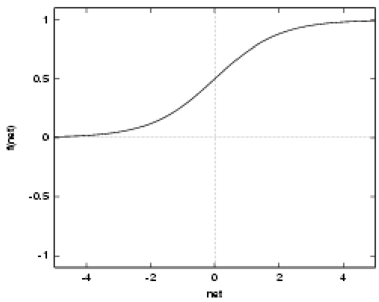
22 Artificial Neural Networks

Identify each of the following activation functions.

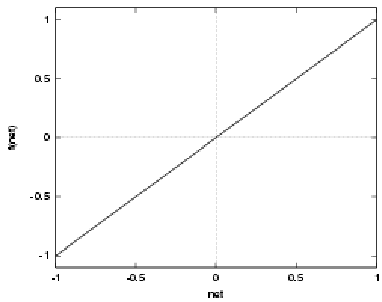
(i)



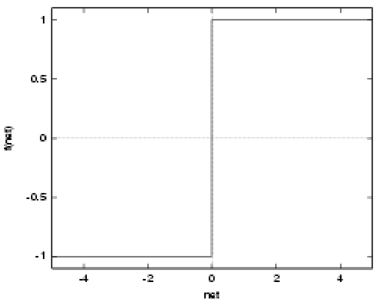
(ii)



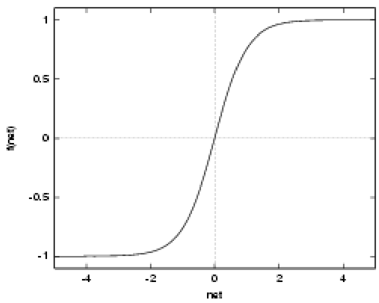
(iii)



(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`].

23 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 1. Does the perceptron give the correct answer (that is, is the actual output the same as the desired output)?

A. Yes.

B. No.

24 Perceptrons

The perceptron in question 23 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. $w_1 = 0.5, w_2 = -0.2, \theta = 0.3$.

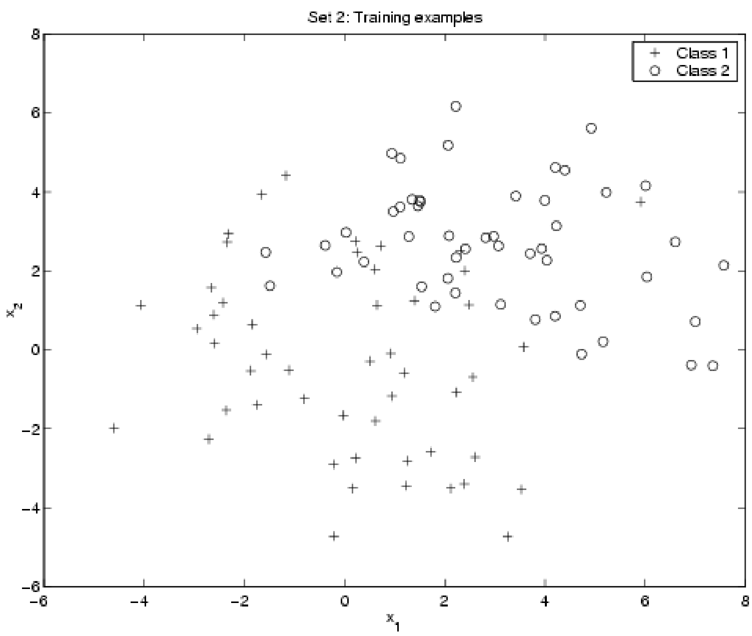
B. $w_1 = 0.5, w_2 = -0.3, \theta = 0.2$.

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

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

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

25 Perceptrons



Is the following statement true or false? "A perceptron is trained on the data shown above, which has two classes (the two classes are shown by the symbols '+' and 'o' respectively). After many epochs of training, the perceptron will converge and the decision line will reach a steady state."

A. TRUE.

B. FALSE.

26 Perceptrons

Is the following statement true or false? "The Perceptron Convergence Theorem states that for any data set which is linearly separable, the Perceptron Learning Rule is guaranteed to find a solution in a finite number of steps."

A. TRUE.

B. FALSE.

27 Perceptrons

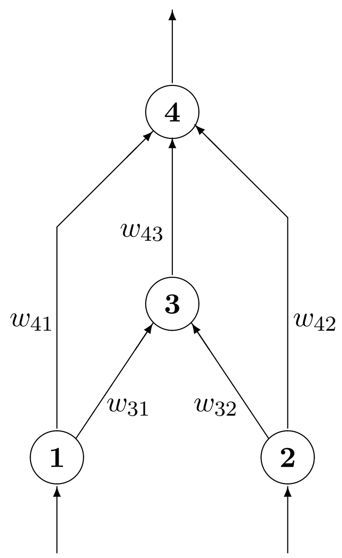
Is the following statement true or false? "The XOR problem can be solved by a multi-layer perceptron, but a multi-layer perceptron with `tansig` activation functions cannot learn to do this."

A. TRUE.

B. FALSE.

28 Perceptrons

The following network is a multi-layer perceptron, where all of the units have binary inputs (0 or 1) 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 $w_{43} = 3$. 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.

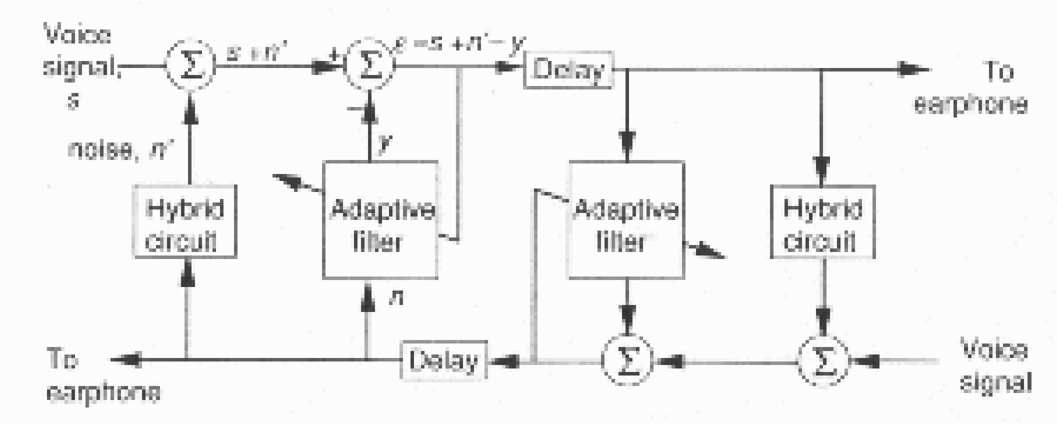
B. OR.

C. XOR.

D. All of the above answers.

E. None of the above answers.

29 Adaline



The Adaline neural network can be used as an adaptive filter for echo cancellation in telephone circuits. For the telephone circuit given in the above figure, which one of the following signals carries the corrected message sent from the human speaker on the left to the human listener on the right? (Assume that the person on the left transmits an outgoing voice signal and receives an incoming voice signal from the person on the right.)

A. The outgoing voice signal, s.

B. The delayed incoming voice signal, n.

C. The contaminated outgoing signal, $s + n'$.

D. The output of the adaptive filter, y.

E. The error of the adaptive filter, $\varepsilon = s + n' - y$.

30 Multi-Layer Feedforward Networks

What is the *credit assignment problem* in the training of multi-layer feedforward networks?

A. The problem of adjusting the weights for the output layer.

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

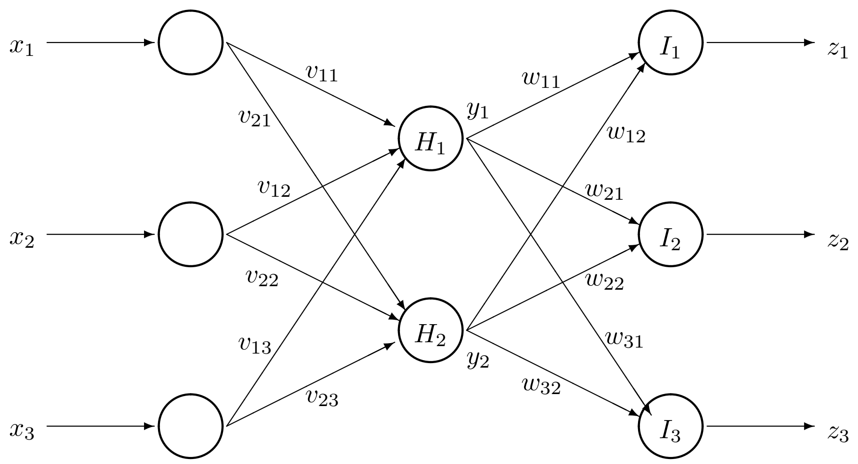
C. The problem of defining an error function for linearly inseparable problems.

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

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

31 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:

$$\mathbf{v}_1 = \begin{bmatrix} 0.4 \\ -0.6 \\ 1.9 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1.2 \\ 0.5 \\ -0.7 \end{bmatrix}, \quad \mathbf{w}_1 = \begin{bmatrix} 1.0 \\ -3.5 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 0.5 \\ -1.2 \end{bmatrix} \quad \text{and} \quad \mathbf{w}_3 = \begin{bmatrix} 0.3 \\ 0.6 \end{bmatrix}.$$

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 output y_1 of the first hidden neuron will be

A. -2.300

B. 0.091

C. 0.644

D. 0.993

E. 4.900

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

32 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_2 of the second output neuron?

A. 0.353

B. 0.387

C. 0.596

D. 0.662

E. 0.674

33 Multi-Layer Feedforward Networks

For the hidden units of the network in question 31, the generalized Delta rule can be written as

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

where $\triangle v_{ji}$ is the change to the weights from unit i to unit j, η is the learning rate, δ_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 δ_j?

A. $\delta_j = f'(H_j)(t_k - z_k)$.

B. $\delta_j = f'(I_k)(t_k - z_k)$.

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

D. $\delta_j = f'(I_k) \sum_k \delta_k w_{kj}$.

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

34 Multi-Layer Feedforward Networks

For the output units of the network in question 31, the generalized Delta rule can be written as

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

where $\triangle w_{kj}$ is the change to the weights from unit j to unit k, η is the learning rate, δ_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 δ_k?

A. $\delta_k = f'(H_j)(t_k - z_k)$.

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

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

D. $\delta_k = f'(I_k) \sum_k \delta_k w_{kj}$.

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

35 Multi-Layer Feedforward Networks

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

A. $\triangle^p w_{kj}(t+1) = \eta \delta_k y_j + \alpha f'(H_j) y_j(t)$

B. $\triangle^p w_{kj}(t+1) = \alpha \delta_k y_j(t)$

C. $\triangle^p w_{kj}(t+1) = \eta \delta_k y_j + \alpha \triangle^p w_{kj}(t)$

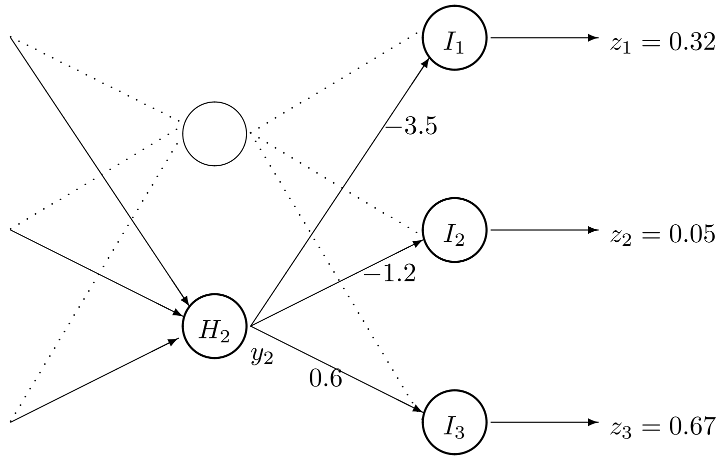
D. $\triangle^p w_{kj}(t+1) = \eta \delta_k y_j(t)$

E. $\triangle^p w_{kj}(t+1) = \eta \delta_k y_j + \alpha \delta_k y_j(t)$

where $\triangle w^p_{kj}(t)$ 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, η is the learning rate, α is the momentum coefficient, δ_k is the error term for unit k, and y_j is the j^{th} input to unit k.

36 Multi-Layer Feedforward Networks

The following figure shows part of the neural network described in question 31. A new input pattern is presented to the network and training proceeds as follows. The actual outputs of the network are given by $\mathbf{z} = [0.32, 0.05, 0.67]^T$ 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. $\delta_{\texttt{output\_1}} = -0.2304$, $\delta_{\texttt{output\_2}} = 0.3402$, and $\delta_{\texttt{output\_3}} = -0.8476$.

B. $\delta_{\texttt{output\_1}} = 0.1084$, $\delta_{\texttt{output\_2}} = 0.1475$, and $\delta_{\texttt{output\_3}} = 0.1054$.

C. $\delta_{\texttt{output\_1}} = 0.1480$, $\delta_{\texttt{output\_2}} = 0.0451$, and $\delta_{\texttt{output\_3}} = 0.0730$.

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 31.)

37 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. $\delta_{\texttt{hidden\_2}} = -0.4219$

B. $\delta_{\texttt{hidden\_2}} = -0.1321$

C. $\delta_{\texttt{hidden\_2}} = -0.0677$

D. $\delta_{\texttt{hidden\_2}} = 0.0481$

E. $\delta_{\texttt{hidden\_2}} = 0.1231$

(Assume exactly the same weights, activation functions and bias values as described in question 31, and exactly the same output vectors $\mathbf{t}$ and $\mathbf{z}$ as described in the previous question.)

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 testing, the target output vector is the same as the input vector.

B. It is important to use smooth, non-decreasing activation functions in the hidden units.

C. The weights of the trained network can be used for pre-processing the data for training another neural network.

D. 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.

E. The trained network can be split into two machines: the first layer of weights compresses the input pattern (encoder), and the second layer of weights reconstructs the full pattern (decoder).

39 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.

40 Unsupervised Learning

A self-organizing map (SOM) has 8 input units, and 10 output units arranged in a ring. How many weights does this network have?

A. 8

B. 10

C. 80

D. 100

41 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.

42 Unsupervised Learning

Which of the following statements is NOT true for a self-organizing map?

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.

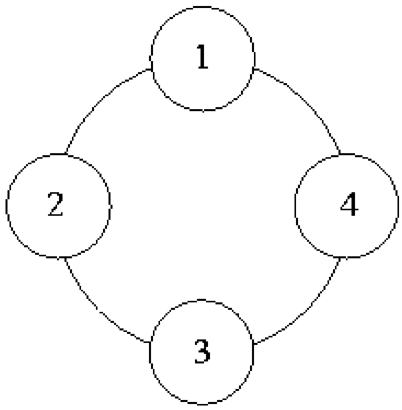
C. The learning rate is a function of the distance of the adapted units from the winning unit.

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 $j \neq k$.

43 Unsupervised Learning

A self-organizing 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.

44 Unsupervised Learning

Adapt the weight vector of the winning unit in question 43 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. $\mathbf{w}_{\texttt{winner}} = [0.30, 0.15, 2.70]^T$

45 Unsupervised Learning

Adapt the weight vectors of the neighbours of the winning unit in question 43 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 $\mathbf{w}_{\texttt{neighbour2}} = [1.32, -1.14, 4.20]^T$

46 Unsupervised Learning

Is the following statement true or false? "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. 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."

A. TRUE.

B. FALSE.

47 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 lowest distance to its best matching unit will be the winner."

A. TRUE.

B. FALSE.

48 Associative Memory

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

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

B. The weights must be positive — that is, $w_{ij} > 0$ for all units i and j.

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

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

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

49 Associative Memory

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

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

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

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

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

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

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

52 Genetic Algorithms

Which of the following operators are the standard ones that are used in nearly all genetic algorithms?
(i) Reproduction, (ii) Elitism, (iii) Mutation, (iv) Ranking, (v) Cross-over.

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

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

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

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

E. (i) and (v).

53 Genetic Algorithms

Is the following statement true or false? "Genetic programming is used for automatic programming (replacing a human programmer) in computer languages such as LISP, by applying the genetic operators in question 52 to a population of possible computer programs."

A. TRUE.

B. FALSE.

54 Applications

Which of the following applications is the most obvious example of a classification problem?

A. Exchange rate prediction (e.g., value of SEK compared to Euro).

B. Box pushing by a mobile robot.

C. Gesture recognition by an autonomous robot.

D. Steering of the ALVINN autonomous vehicle.

E. Locating a gas source with a mobile robot.

55 Applications

Which type of learning system was used for pattern recognition in the *phonetic typewriter*?

A. A linear feedforward network.

B. A multi-layer feedforward network.

C. A simple recurrent network.

D. A self-organizing map.

E. A genetic algorithm.

56 Applications

Which type of learning system was used for teaching a mobile robot to perform simple behaviours such as obstacle avoidance and wall following?

A. A linear feedforward network.

B. A multi-layer feedforward network.

C. A simple recurrent network.

D. A self-organizing map.

E. A genetic algorithm.

57 Applications

Which type of learning system was used to teach a mobile robot to perform route following by stimulus-response learning?

A. A linear feedforward network.

B. A multi-layer feedforward network.

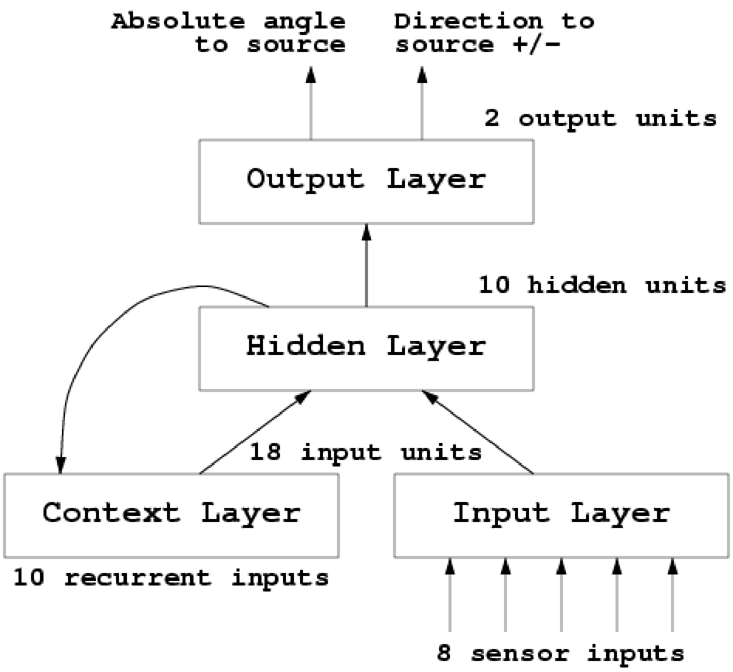
C. A simple recurrent network.

D. A self-organizing map.

E. A genetic algorithm.

58 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 a gas source.



How many weights does this network contain?

A. 100.

B. 200.

C. 300.

59 Applications

For the same application described in question 58, why is a recurrent neural network better for this purpose than a non-recurrent neural network?

A. It is faster to train a recurrent network, so the system can learn to locate the gas source in real-time.

B. A recurrent network always produces better generalization performance than a non-recurrent network, so the system is more robust.

C. The recurrent network has feedback connections which allow the network to take into account the sequence of sensor readings as the robot moves, so the system is more robust.

D. The recurrent network can learn a topological mapping in the weights from the hidden layer to the context layer – this helps to overcome the problems of air turbulence and noisy sensor readings.

60 Applications

What is the usual sequence of events for training the neural network in question 58?

(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. (ii), (i), (vi), (vii), (viii).

B. (iii), (ii), (i), (vii), (vi).

C. (ii), (i), (vi), (vii), (iv).

D. (i), (ii), (vi), (iv), (vii).

E. (ii), (i), (vi), (vii), (v).

Answers

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

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

21.		D	+1	$-\frac{1}{4}$	
22.		A	+1	$-\frac{1}{3}$	
23.		B	+1	-1	
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.		E	+1	$-\frac{1}{4}$	
30.		E	+1	$-\frac{1}{4}$	

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

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

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