# Artificial Neural Networks
# Examination, March 2001

## Instructions

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 and large enough to be read by an average examiner!

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

# Questions

## 1 Classical Pattern Recognition

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

A. *patterns*.

B. *classes*.

C. *features*.

## 2 Classical 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 Classical Pattern Recognition

What is a *pattern vector*?

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

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

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

## 4 Classical Pattern Recognition

What is a *decision function*?

A. A function $d_{ij}(\mathbf{x})$ which gives the correct answer in a pattern recognition problem with $M$ classes.

B. A function $d_j(\mathbf{x})$ which is defined for each class in a classification problem, so that an input pattern $\mathbf{x}$ is assigned to class $\omega_k$ if $d_k(\mathbf{x}) = \max\{d_1(\mathbf{x}), ..., d_M(\mathbf{x})\}$.

C. A function $d_j(\mathbf{x})$ which specifies the class conditional probability $P(\mathbf{x}|\omega_k)$ of an input example $\mathbf{x}$ given that it belongs to class $\omega_j$.

## 5 Classical Pattern Recognition

Which of the following statements is NOT true for a minimum distance classifier (MDC)?

A. The MDC is specified completely by the prototype vectors for all $M$ classes in a classification problem.

B. The MDC minimizes the average loss of misclassifying the input patterns.

C. The MDC is a special case of the Bayes optimal classifier.

D. The MDC makes a decision boundary between two classes. This boundary is a line, plane or hyperplane where the two decision functions are equal.

# 6 Classical Pattern Recognition

Which of the following statements is NOT true for a Bayes optimal classifier (BOC)?

A. The BOC considers the prior probabilities of each class, and these probabilities can be estimated from a random sample.

B. The decision function for the BOC is given by $d_j(\mathbf{x}) = \mathbf{x}^T\mathbf{m}_j - \frac{1}{2}\mathbf{m}_j^T\mathbf{m}_j$ where $\mathbf{m}_j$ specifies the mean vector of each class $\omega_j$.

C. If the data for each class is normally distributed, then the BOC can be statistically optimal.

D. If the data for each class is normally distributed, then the BOC is specified completely by the mean vectors and covariance matrices of each class.

E. The BOC makes a decision boundary between two classes. If the covariance matrices of the two classes are the same, then this boundary is a line, plane or hyperplane where the two decision functions are equal.

# 7 Classical Pattern Recognition

How should you train a minimum distance classifier if you assume that the classes are normally distributed?

A. Randomly pick some of the training data for training (e.g., 70%) and use the rest for testing.

B. Calculate the maximum likelihood estimates for the mean vectors of each class using the training data.

C. Save one of the pattern vectors for testing, and use the rest for training. Repeat for all of the vectors in the training set.

D. Calculate the maximum likelihood estimates for the mean vectors and covariance matrices of each class using the training data.

# 8 Classical Pattern Recognition

What is a *hyperplane*?

A. A very fast aeroplane.

B. A planar (flat) surface in high-dimensional space.

C. Any high-dimensional surface.

# 9 Training and Testing

An artificial neural network may be trained on one data set and tested on a second data set. The system designer can then experiment with different numbers of hidden layers, different numbers of hidden units, etc. 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 provides a better (unbiased) estimate of the true generalization error.

B. The error on the third data set is used to choose between lots of different possible systems.

C. It's not important – testing on the second data set indicates the generalization performance of the system.

## 10 Training and Testing

What is an *outlier*?

A. An input pattern which produces a classification error.

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

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

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

## 11 Training and Testing

Which one of the following statements is the best description of *hold-one-out* training?

A. Randomly pick some of the training data for training (e.g., 70%) and use the rest for testing.

B. Calculate the maximum likelihood estimates for the mean vectors of each class using the training data.

C. Save one of the pattern vectors for testing, and use the rest for training. Repeat for all of the vectors in the training set.

D. Calculate the maximum likelihood estimates for the mean vectors and covariance matrices of each class using the training data.

## 12 Biological Neurons

Is the following statement true or false? "In the human brain, about 70% of the neurons are input (afferent) and output (efferent). The remaining 30% are used for information processing."

A. TRUE.

B. FALSE.

## 13 Biological Neurons

Which of the following statements are true for typical neurons in the human brain?

A. Electrical potential is summed in the neuron.

B. When the potential is bigger than a threshold, the neuron fires a pulse through the axon.

C. The neurons are connected to each other by axons, synapses and dendrites.

D. All of the above answers.

## 14 Artificial Neural Networks

Which of the following answers is NOT a general characteristic of artificial neural networks?

A. Learning.

B. Fast processor speed.

C. Generalization.

D. Robust performance.

E. Parallel processing.

## 15 Hebbian Learning

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

A. $\triangle \mathbf{w}_k = \eta y_k \mathbf{x}$

B. $\triangle \mathbf{w}_k = \eta \left(d_k - y_k\right) \mathbf{x}$

C. $\triangle \mathbf{w}_k = \eta \left(\mathbf{x} - \mathbf{w}_k\right)$

D. $\triangle \mathbf{w}_j = \eta_j \left(\mathbf{x} - \mathbf{w}_j\right)$, 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 Perceptrons

Each of the inputs to the perceptron is multiplied by a number to give it a weight. These weights allow the strength of the different connections to be changed so that the perceptron can learn.

A. TRUE.

B. FALSE.

## 17 Perceptrons

A perceptron adds up all the weighted inputs it receives. If the sum exceeds a certain value, then the perceptron outputs a 1, otherwise it just outputs a 0.

A. TRUE.

B. FALSE.

C. Sometimes - it can also output continuous values inbetween 0 and 1.

## 18 Perceptrons

The name for the function in question 17 is

A. Unipolar step function.

B. Bipolar step function.

C. Sigmoid function.

D. Logistic function.

E. Perceptron function.

## 19 Perceptrons

Perceptrons can be used for many different tasks, e.g., to recognise letters of the alphabet. Can a perceptron find a good solution in any pattern recognition task?

A. Yes.

B. No.

## 20 Perceptrons

Which of the following equations is the best description of the *Perceptron Learning Rule*?

A. $\triangle \mathbf{w}_k = \eta y_k \mathbf{x}$

B. $\triangle \mathbf{w}_k = \eta \left(d_k - y_k\right) \mathbf{x}$

C. $\triangle \mathbf{w}_k = \eta \left(\mathbf{x} - \mathbf{w}_k\right)$

D. $\triangle \mathbf{w}_j = \eta_j \left(\mathbf{x} - \mathbf{w}_j\right)$, 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 $k$.

## 21 Perceptrons

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

A. TRUE.

B. FALSE.

## 22 Perceptrons

Why is the XOR problem interesting to neural network researchers?

A. Because it cannot be encoded in a way that allows you to use a neural network.

B. Because it is the simplest linearly inseparable problem that exists.

C. Because it is a complex binary function that cannot be solved by a neural network.

## 23 Perceptrons

A perceptron has input weights $w_1 = 3.1$ and $w_2 = 1.9$ and a threshold value $T = 0.4$. What output does it give for the input $x_1 = 1.2$, $x_2 = 2.3$?

A. $3.1 \times 1.2 + 1.9 \times 2.3 = 3.72 + 4.37 = 8.09$.

B. $3.1 \times 1.2 + 1.9 \times 2.3 = 3.72 + 4.37 = 8.09$. Now subtract the threshold: $8.09 - 0.4 = 7.69$.

C. $3.1 \times 1.2 + 1.9 \times 2.3 = 3.72 + 4.37 = 8.09$. This is more than the threshold, so output 0.

D. $3.1 \times 1.2 + 1.9 \times 2.3 = 3.72 + 4.37 = 8.09$. This is more than the threshold, so output +1.

E. 1.2 and 2.3 are both greater than 0.4, so both inputs are replaced by +1. The weighted sum is therefore $3.1 + 1.9 = 5.0$, which is the answer.

## 24 Perceptrons

Which of the following 2 input Boolean logic functions are linearly inseparable (that is, NOT linearly separable)? (i) AND, (ii) OR, (iii) XOR, (iv) NAND, (v) NOT XOR.
Truth tables for these functions are provided below, showing the inputs $x_1$ and $x_2$ together with the respective desired outputs $d$.

(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), (ii), (iii), (iv) and (v)

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

C. (ii) and (iii)

D. (iii) only

E. (iii) and (v)

## 25 Perceptrons

If a perceptron is given a training set which is NOT linearly separable, then what will happen during training?

A. The weights of the network will reach a steady state, where the decision boundary gives the best possible separation of the data.

B. The weights of the network will reach a steady state, but the decision line may give a sub-optimal (almost optimal) separation of the data.

C. The weights of the network will not reach a steady state, and the decision line will not stop moving.

D. The weights of the network will not reach a steady state, therefore a decision line cannot exist.

## 26 Associative Memory

What is an *associative memory* (or *content addressable memory*)?

A. A neural network which can solve linearly inseparable pattern recognition problems such as face recognition.

B. A neural network which can store a set of patterns, and recall the nearest stored pattern when an incomplete pattern is presented to the network.

C. A neural network which can solve classification problems by learning to associate pattern vectors with the desired outputs, provided that it doesn't get stuck in a local minimum.

D. A neural network where the desired outputs are equal to the inputs, and a set of patterns are stored in the hidden layer.

E. A neural network which can store a set of patterns, and interpolate between the stored patterns when noisy patterns are presented to the network.

## 27 Associative Memory

Which of the following statements does NOT describe a Hopfield network?

A. A Hopfield network can be trained using unsupervised learning.

B. A Hopfield network is a completely connected network.

C. A Hopfield network minimizes an energy function during recall.

D. A Hopfield network uses soft competitive learning.

E. The units in a Hopfield network can be updated randomly during recall.

## 28 Associative Memory

Why was Hopfield's energy analysis so important?

A. This analysis proved that a Hopfield network cannot get stuck in a local minimum.

B. This analysis proved that the learning rule in a Hopfield network will always converge.

C. This analysis proved that the Hopfield network is guaranteed to reach a steady state called an "attractor" during recall.

D. This analysis defined a quantity called "energy" which cannot decrease when the network is updated.

## 29 Multi-Layer Feedforward Networks

The sigmoid function is

A. S-shaped.

B. Z-shaped.

C. A step function.

D. U-shaped.

## 30 Multi-Layer Feedforward Networks

Why is the sigmoid function important to the success of multi-layer feedforward networks?

A. It allows the perceptrons to reconfigure their arrangement and therefore to solve the problem.

B. It allows the thresholding to be carried out more quickly.

C. It allows the network to learn.

## 31 Multi-Layer Feedforward Networks

What are *hidden layers*?

A. Layers of units that have no direct connections to any other units.

B. Layers of units that have no direct connection to the input or the output.

C. Layers of units that do not contribute towards the output.

D. None of the above answers.

## 32 Multi-Layer Feedforward Networks

A multi-layer feedforward network with "`logsig`" activation functions can solve the XOR problem satisfactorily: this is because each unit can linearly separate one part of the space, and they can then combine their results.

A. True - these networks can do this, but they are unable to learn to do it - they have to be coded by hand.

B. True - this usually works, and these networks can learn to classify even complex problems.

C. False - only a network with sigmoid activation functions can learn to solve the XOR problem.

D. False - just having a single layer network is enough.

## 33 Multi-Layer Feedforward Networks

What is *back-propagation*?

A. It is the transfer of error back through the network to adjust the inputs.

B. It is the transfer of error back through the network to allow the weights to be adjusted.

C. It is the transfer of error back through the network using a set of recurrent connections.

D. It is the transfer of outputs back from the hidden layer to the input layer using a set of recurrent connections.

## 34 Multi-Layer Feedforward Networks

A multi-layer network should have the same number of units in the input layer and the output layer.

A. TRUE.

B. FALSE.

## 35 Multi-Layer Feedforward Networks

The number of connections in a feedforward network is proportional to the number of units.

A. TRUE.

B. FALSE.

## 36 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 $j$.

## 37 Multi-Layer Feedforward Networks

Why is the error function called "the error function"?

A. It just is.

B. Because it measures the percentage of incorrectly classified vectors.

C. Because it measures the error in the output of the network - large values are close to being correct, while small values are very wrong.

D. Because it measures the error in the output of the network - small values are close to being correct, while large values are very wrong

## 38 Multi-Layer Feedforward Networks

How can learning in a multi-layer feedforward (MLFF) network be explained using the error function?

A. It can't - the error function is irrelevant.

B. The network learns by adjusting its weights to reduce the error each time.

C. The network reduces the error by adjusting the target patterns each time.

D. The networks learns by minimizing the value of the energy function, until a local or global minimum is reached.

## 39 Multi-Layer Feedforward Networks

Which of the following answers is NOT a strategy for dealing with local minima when training a multi-layer feedforward (MLFF) network?

A. Add extra random noise to the weights or input vector.

B. Add a momentum term to the weight update equation in the Generalized Delta Rule.

C. Train a number of different networks, and select the one with the highest sum squared error on an independent data set.

D. Train a number of different networks, and take the average output of all the networks.

## 40 Multi-Layer Feedforward Networks

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

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

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

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

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

## 41 Multi-Layer Feedforward Networks

Which of the following statements is the best description of *early stopping*?

A. Train the network until a local minimum in the error function is reached.

B. Simulate the network on a test data set after every epoch of training. Stop training when the generalization error starts to increase.

C. Add a momentum term to the weight update in the Generalized Delta Rule, so that training converges more quickly.

D. Use a faster version of backpropagation, such as the 'Quickprop' algorithm.

## 42 Multi-Layer Feedforward Networks

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

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

B. The network cannot predict the correct outputs for test examples which are outside the range of the training examples.

C. The network does not contain enough adjustable parameters (e.g., hidden units) to find a good approximation to the unknown function which generated the training data.

D. The network cannot predict the correct outputs for test examples which lie inbetween some of the training examples.

## 43 Multi-Layer Feedforward Networks

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

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

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

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

## 44 Recurrent Artificial Neural Networks

Which of the following statements is NOT true for Elman's *simple recurrent network* (SRN)?

A. The hidden units receive inputs which depend on previous computations in the hidden layer.

B. The sequence (order) of the examples in the training set is important.

C. SRNs are based on a computational model of biological vision.

D. SRNs can be trained using backpropagation.

E. SRNs can learn to approximate functions of time or space.

## 45 Self-Organization

What is a *dead unit*?

A. A unit which does not get updated by any of its neighbours during training.

B. A unit which does not respond maximally to any of the training patterns.

C. The unit which produces the biggest sum-squared error.

## 46 Self-Organization

Why is soft competitive learning (SCL) different from hard competitive learning (HCL)?

A. In HCL, there may be dead units after training.

B. In SCL, a Fast Fourier Transform can be used for feature extraction.

C. In HCL, only the winner is adapted during training.

D. In SCL, the units are picked randomly during testing.

E. In HCL, the neighbours of the winner are updated during training.

## 47 Self-Organization

Which of the following equations is the best description of *hard competitive learning*?

A. $\triangle \mathbf{w}_k = \eta y_k \mathbf{x}$

B. $\triangle \mathbf{w}_k = \eta \left( d_k - y_k \right) \mathbf{x}$

C. $\triangle \mathbf{w}_k = \eta \left( \mathbf{x} - \mathbf{w}_k \right)$

D. $\triangle \mathbf{w}_j = \eta_j \left( \mathbf{x} - \mathbf{w}_j \right)$, 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 $k$.

## 48 Self-Organization

What is the *neighbourhood* in the self-organizing feature map?

A. A group of neurons next to the winning unit.

B. A group of neurons which are not updated during training.

C. A group of input patterns which are very similar.

D. A group of neurons which are not dead units.

## 49 Self-Organization

What is the *topological mapping* in the self-organizing feature map?

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

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

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

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

## 50 Genetic Algorithms

Which of the following operations are standard operations in a genetic algorithm?
(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).

## 51 Genetic Algorithms

Which operation in a genetic algorithm can introduce new genetic information into a population?

A. Reproduction.

B. Elitism.

C. Mutation.

D. Ranking.

E. Cross-over.

## 52 Genetic Algorithms

The *weighted roulette wheel* is a technique used for

A. picking the best chromosome in the population.

B. randomly selecting the chromosomes in the reproduction operation.

C. crossing-over the fittest chromosomes.

D. ranking the chromosomes in the mutation process.

E. measuring the fitness of the chromosomes in the reproduction operation.

## 53 Applications

Which of the following neural network applications uses *unsupervised* learning?

A. The ALVINN autonomous vehicle.

B. Obstacle avoidance by a mobile robot.

C. The phonetic typewriter.

D. Stock market prediction.

E. Vegetation classification in satellite images.

## 54 Applications

Which of the following applications of multi-layer feedforward (MLFF) networks is NOT a good example of a regression problem?

A. Approximating a continuous function.

B. Wall following by a mobile robot.

C. Learning to steer the ALVINN autonomous vehicle.

D. Speech recognition.

## 55 Applications

A mobile robot can locate an odour source using an electronic nose. It does this by turning and "smelling" the air using a set of gas sensors. In this system, an artificial neural network can be used to learn the direction to the odour source. Why is a recurrent neural network better for this purpose than a non-recurrent neural network?

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

B. The recurrent network can learn a topological mapping in its hidden layer - this helps to overcome the problems of air turbulence, noise and delays in the sensors.

C. It is faster to train a recurrent network, so the system can learn to deal with the air turbulence problems in real-time.

D. The recurrent network has feedback connections which allow the network to consider the whole sequence of sensor readings, so the system is more robust.

## 56 Applications

"An *auto-associator* is a special type of multi-layer feedforward network which can be used for data compression. This is because (i) the target outputs are the same as the inputs, (ii) the number of hidden units is smaller than the number of input units, and (iii) the backpropagation algorithm tries to minimize the sum squared error on the training set. Therefore, the algorithm should find a hidden layer representation which encodes the most useful information in the input vector and ignores the redundant information."

A. TRUE

B. FALSE

## 57 Applications

Which technique was used to improve the training in the ALVINN autonomous vehicle?

A. Early stopping.

B. A Fast Fourier Transform was used to pre-process the image data.

C. Each input image was modified slightly and re-used several times in the training set.

D. The training signal was generated internally by another sensor (a laser range finder).

E. The trainer was the brother of Michael Schumacher.

## 58 Applications

How many hidden layers were there in Kohonen's *phonetic typewriter*?

A. None (0).

B. One (1).

C. Two (2).

## 59 Applications

How many hidden layers were there in the ALVINN autonomous vehicle?

A. None (0).

B. One (1).

C. Two (2).

## 60 Laborations

What does the following MATLAB function do?

```
>> net = newff(minmax(p), [4,2], {'tansig','logsig'});
```

A. Initialize a single-layer network with 4 input units, 2 output units and linear activation functions.

B. Initialize a multi-layer network with 4 hidden units, 2 output units and sigmoid activation functions.

C. Initialize a multi-layer network with non-linear activation functions and two hidden layers - the first hidden layer has 4 units and the second one has 2 units.

D. Initialize a multi-layer network with sigmoid activation functions, 4 hidden units and 2 recurrent connections back to the input layer.

## Answers

1.C
2.A
3.B
4.B
5.B
6.B
7.B
8.B
9.A
10.B

11.C
12.B
13.D
14.B
15.A
16.A
17.A or B
18.A
19.B
20.B

21.B
22.B
23.D
24.E
25.C
26.B
27.D
28.C
29.A
30.C

31.B
32.B
33.B
34.B
35.B
36.B
37.D
38.B
39.C
40.A

41.B
42.C
43.B
44.C
45.B
46.C
47.C
48.A
49.B
50.C

51.C
52.B
53.C
54.D
55.D
56.A
57.C
58.A
59.B
60.B