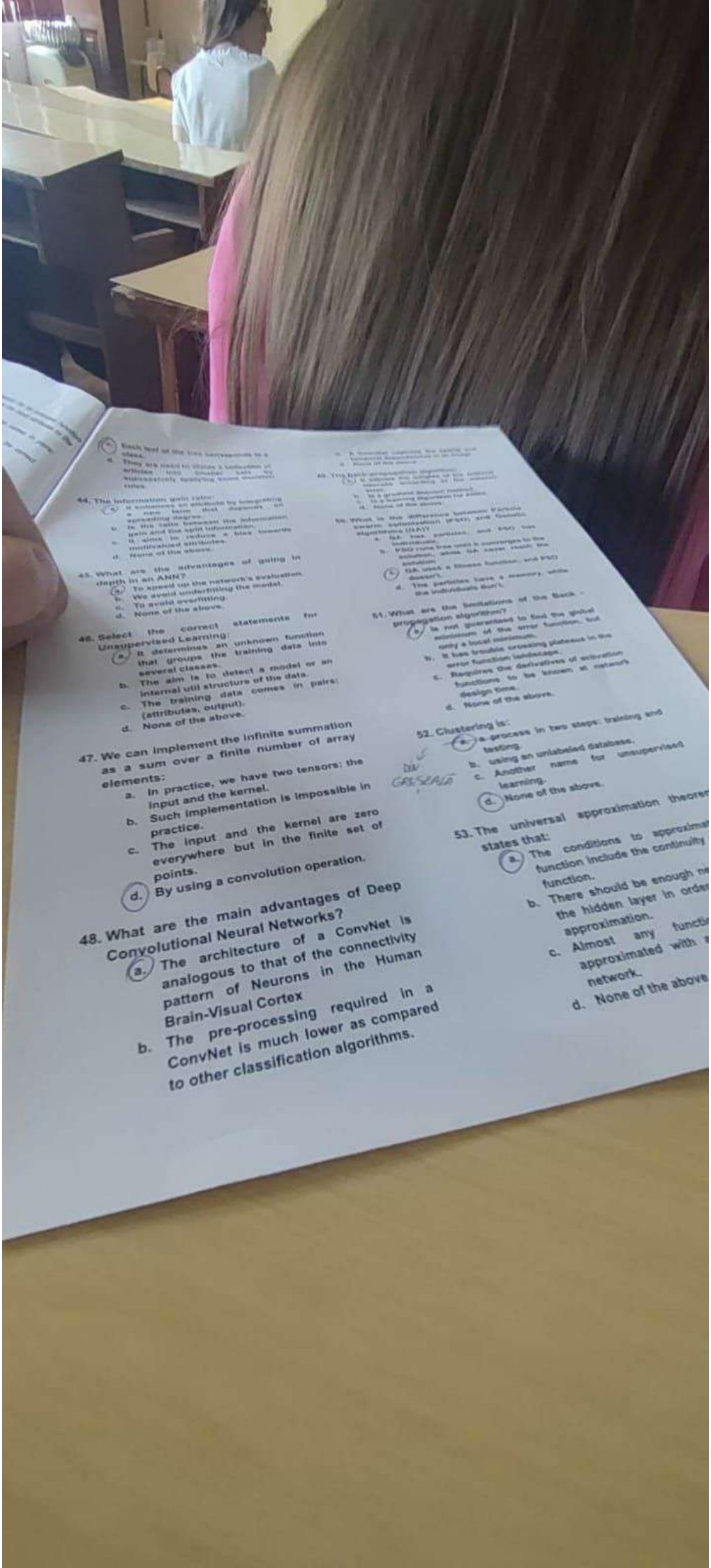

Each leaf of the tree corresponds to a class.
They are used to divide a collection of articles into smaller sets by successively applying some decision rules.

44. The information gain ratio:
a. It evaluates an attribute by integrating a new term that depends on spreading degree.
b. Is the ratio between the information gain and the split information.
c. It aims to reduce a bias towards multivalued attributes.
d. None of the above.

45. What are the advantages of going in depth in an ANN?
a. To speed up the network's evaluation.
b. We avoid underfitting the model.
c. To avoid overfitting.
d. None of the above.

46. Select the correct statements for Unsupervised Learning:
a. It determines an unknown function that groups the training data into several classes.
b. The aim is to detect a model or an internal util structure of the data.
c. The training data comes in pairs: (attributes, output).
d. None of the above.

47. We can implement the infinite summation as a sum over a finite number of array elements:
a. In practice, we have two tensors: the input and the kernel.
b. Such implementation is impossible in practice.
c. The input and the kernel are zero everywhere but in the finite set of points.
d. By using a convolution operation.

48. What are the main advantages of Deep Convolutional Neural Networks?
a. The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain-Visual Cortex
b. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms.

49. The back-propagation algorithm:
a. It allows the weights of the network to be updated in order to minimize the error.
b. Is a gradient descent method.
c. Is a learning algorithm for ANNs.
d. None of the above.

50. What is the difference between Particle swarm optimization (PSO) and Genetic algorithms (GA)?
a. GA has particles, and PSO has individuals.
b. PSO runs less once it converges to the solution, while GA never reach the solution.
c. GA uses a fitness function, and PSO doesn't.
d. The particles have a memory, while the individuals don't.

51. What are the limitations of the Back - propagation algorithm?
a. Is not guaranteed to find the global minimum of the error function, but only a local minimum.
b. It has trouble crossing plateaus in the error function landscape.
c. Requires the derivatives of activation functions to be known at network's design time.
d. None of the above.

52. Clustering is:
a. a process in two steps: training and testing.
b. using an unlabeled database.
c. Another name for unsupervised learning.
d. None of the above.

53. The universal approximation theorem states that:
a. The conditions to approximate function include the continuity function.
b. There should be enough ne the hidden layer in order approximation.
c. Almost any functi approximated with network.
d. None of the above

c. The error gradient in an ANN is computed based on the neuron's activation function.
d. The gradient descent is based on the error associated with the entire set of train data.

33. The main principle of selecting a proper learning algorithm is:
a. According to the desired data.
b. Computational complexity of the target objective.
c. Ability to predict a cluster membership.
d. Error minimization (cost function - loss function).

34. What sort of problems can a perceptron solve?
a. Linear separations of elements from the domain.
b. XOR problem.
c. It depends on the structure.
d. None of the above.

35. The vanishing gradients during backpropagation are:
a. Not affecting us, since the derivative is zero in this case, on most of the domain.
b. An advantage in the training process that leads to faster convergence.
c. A tipical problem when the network has too many hidden layers
d. None of the above.

The activation logistic function:
b. Is a linear function.
b. Suffers from a vanishing gradient.
c. Has limitations regarding the output domain.
None of the above.

ter vision, we apply a filter to an
r to extract some features
a convolution operation with
kernel and performing a
rom the portion P of the
which the kernel is

nts for

a. We search for an unknown f that maps the input attributes outputs.
b. The training data comes in attributes and output.
c. The aim is to provide the co output for a new input.
d. None of the above

DR CASSIAS

39. An ANN has a structure of 226:15:10: with a sigmoid activation function. How many weights will have the first neuron from the first hidden layer?
a. 2
b. 226
c. 10
d. 15

40. In an artificial neuron, the transfer function:
a. Is the equation of a hyperplane.
b. Its nature limits the solving capacity of the neuron.
c. Is the inner product of the input vector with the weight vector.
d. It can be the sigmoid function.

41. The indirect experience when choosing the training database is:
a. In pairs (in / out).
b. Useful feedback for the objective function.
c. Based on independent data with annotated content.
d. None of the above.

42. On a ConvNet the feature learning:
a. will minimize the loss function by extracting those who are most useful for classifying the images.
b. allows a suite of tens or even hundreds of other small filters to be designed in order to detect more complex features in the image.
c. is performed before training the conv-layers.
d. None of the above.

43. What are the correct statements about decision trees?
a. DT contains three types of nodes: decision nodes, hazard nodes, and result nodes.
b. The decision nodes are on the leafs while the result nodes are in the tree levels

What crossover method(s) are correct for a binary representation in a GA?
a. There is no crossover for this representation
b. With n-cutting points
c. Uniform
d. Strong mutation

22. The "dying ReLU" problem refers to:
a. The values of the derivative.
b. The vanishing gradient.
c. The values of the function.
d. None of the above.

23. Choose the correct compatibility between the error function and the activation function from the output layer:
a. Cross entropy error with softmax.
b. Binary cross entropy with logistic error with the sigmoid activation function.
c. Mean square error with the linear function.
d. None of the above.

24. How does ConvNet compute an image?
a. In such a network, the filters' results are never combined.
b. There is automation in detecting the weights for the kernels.
c. There are usually three stages: several convolutions, a detector stage, and a pooling stage.
d. None of the above.

25. Select the correct statements for Cross-Entropy loss:
a. Is used in classifications
b. Is the difference between two probability distributions for a provided set of occurrences or random variables.
c. Is never used after the softmax transformation after output.
d. None of the above.

26. What are the properties of the training and testing data?
a. They have to respect the same distribution law.
b. The test data should be based on real experiences, and the training should be based on theoretical experiences.
c. If possible, the training and the test data should be disjunct sets.
d. None of the above.

27. The ReLu function
a. Does not have a vanishing gradient when x > 0
b. It is a non-linear activation function
c. Corrects the problems that occur as sigmoid function
d. Provides sparsity since y = 0 when x < 0

28. Which of the following statements is true when we apply a max pooling transformation over a tensor:
a. We emphasize the features.
b. We handle inputs of varying sizes.
c. We make the representation approximately invariant to small translations of the input.
d. We return the maximum value from the portion of the image covered by the kernel.

29. The harmonic mean between the precision and the recall is:
a. A metric for distance in a reinforcement learning algorithm.
b. The F1 score.
c. A statistical metric used to evaluate performance in a supervised learning process.
d. None of the above.

30. When we compare two algorithms, we can use:
a. The divergence of the Accuracy.
b. Confidence intervals.
c. Performance measures.
d. None of the above.

31. How are the neurons connected in feed forward ANN?
a. Through a backward signal.
b. Through an output with a neu the same layer.
c. They are not connected.
d. Through weighted links.

32. Select the correct statements
a. In order to determine slope on gradient des E based on w.
b. When training an A weights in the steepest slope of for the entire set

b. The perceptron's rule is based on gradient descent.
c. They both start with some random weights.
d. On Preceptron's rule, we recompute the weights based on the model's quality, while on the delta rule, we don't.

12. The L1 Loss is:
a. It computes the average of the sum of absolute differences between actual values and predicted ones
b. Is also called the softmax loss
c. Is used especially when the distribution has outliers
d. used for regression problems

13. What is the proper encoding for an individual in ANNs?
a. There are no individuals in ANN
b. An array of bits that encode proper information related to the solution.
c. A set of weights used to propagate a signal.
d. A computer program that learns to classify and performs regressions.

14. In a Search Strategies algorithm with a tree-based structure, the evaluation function:
a. It is used for estimating the cost of a solution through a node (state).
b. It guides the search.
c. It is the sum between the cost of a solution path from the initial node (state) to node (state) n and the cost of a solution path from node (state) n to the final node (state)
d. None of the above

...t is standardization?
The process by which raw values are transformed into z scores.
...he operation that transforms ...inuous values in discrete ones.
...a transformation that removes ...ale effect.
...f the above

...owing problem: There is ...s printed with integer ...o 10. Select two subsets ...elements and do not ...ary) using a GA, in

such a way that they have the ... elements. What is a proper repr...
a. Vectors of M real num... 10
b. Vectors of M integers from ...
c. Permutation of size M
d. Binary

17. What elements determine the ... velocity of a particle in a PSO algorithm (check all correct ones)
a. the old velocity
b. inertia, social coefficient
c. the current position of the particle
d. None of the above.

18. The limited model capacity of ANNs is overcome by:
a. Adding more layers in depth.
b. Reducing the number of artificial neurons.
c. Adding nonlinearity to the model.
d. None of the above.

19. What is the fitness function for the following problem? "There is a set of M cards printed with integer numbers from -10 to 10. Select two subsets (that have no common elements and do not form necessary a partition) using a GA, in such a way that they have the same sum of elements."
a. The absolute value of the sum of the selected elements from the first set / the number of elements from the second set
b. The difference in absolute value between the sum of the subset' elements and their divisors
c. The difference in absolute value between the sums of elements from each subset
d. The number of elements

20. The Perceptron's algorithm:
a. Is based on error minimization associated with an instance of train data.
b. The error is the difference between the real output y and the output o computed by the perceptron for an input.
c. It modifies the weights based on errors associated with an instance of train data.
d. None of the above.

Artificial Intelligence

[redacted]

1. When constructing a decision tree, the attribute selection can be:
   a. In a preestablished order
   b. Based on the gained information
   c. Random
   d. Based on data frequency

2. The softmax function:
   a. Transforms in probabilities the output scores of the classes
   b. It interpolates the cross-entropy function
   c. Is used in regressions
   d. None of the above

3. What problems can be solved with machine learning?
   a. Regressions
   b. Predictions
   c. Plannings and classifications
   d. None of the above

4. In order to overcome the perceptrons' limits, we can:
   a. Use neurons with a continuous threshold
   b. Use kernel transformations on the domain
   c. These limits can't be overcome
   d. Increase the number of neurons

5. The objective of a machine learning algorithm can be represented as:
   a. Numeric functions
   b. Probabilistic functions
   c. A set of symbolic rules
   d. A table

6. The induction faze for the process of building a DT is:
   a. It labels the new data with the build rules
   b. It eliminates the branches that reflect noise or exceptions
   c. Based on the training data
   d. Works bottom-up or top-down

7. How does the artificial neuron process the information?
   a. Based on the error
   b. Based on the activation function
   c. Based on back propagation
   d. None of the above

8. Using a feed-forward ANN, we want to determine if a shape from a black-and-white image is a square or not. How is the error computed?
   a. Based on the output of the hidden layer
   b. Based on the difference between the real output of the network and the desired output
   c. Based on an induction formula
   d. None of the above

9. What is a tensor?
   a. An image with multiple channels
   b. A mathematical object
   c. Generalizations of scalars, vectors, and matrices to an arbitrary number of indices
   d. None of the above

10. Select the correct combination(s):
    a. Output type: Continuous. Output Distribution: Gaussian, Output Layer: Linear, Cost Function: MSE
    b. Output type: Discrete. Output Distribution: Multinoulli. Output Layer: Softmax, Cost Function: Cross Entropy
    c. Output type: Binary. Output Distribution: Bernoulli, Output Layer: Sigmoid, Cost Function: Binary Cross Entropy
    d. None of the above.

11. What are the differences and similarities between the perceptron's rule and the delta rule?
    a. In the delta rule, the model's quality is established based on all the data.