What are the characteristics of the back-propagation algorithm?

a. Does not require the derivatives of activation functions to be known at network design time.
b. It is crossing easy plateaus in the error function landscape.
c. Is guaranteed to find the global minimum of the error function, not only a local minimum.
d. None of the above.

). The softmax function:

a. It incorporates the cross entropy function.
b. Transforms the output scores for the classes in probabilities.
c. Is used in regressions.
d. None of the above.

. Select the correct statements for Cross-Entropy loss:

a. Is used in regressions
b. Is never used when we apply the softmax transformation to the network's output.
c. Is the difference between two probability distributions for a provided set of occurrences or random variables.
d. None of the above.

What is the proper encoding for an ndividual in ANNs?

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

at is the feature of the database in ining with indirect experience:

a. It is based on independent data with annotated content.
b. It comes in pairs (in/out).
c. It is based on useful feedback for some objective function.
d. None of the above.

14. Select the correct statement supervised learning:

a. The aim is to provide an arbitrary output for a new input.
b. We search for a known function that maps the input attributes to the outputs.
c. The training data comes in an unpaired format: only attributes or only output.
d. None of the above

15. How does a ConvNet figure out what is in an image?

a. Automatically detecting the weights for the kernels during training.
b. There are usually three stages: several convolutions, a decomposing stage, and a flattening stage.
c. By decomposing the features.
d. None of the above.

16. What elements determine the new velocity of a particle in a PSO algorithm? (check all correct ones)

a. the current position of the weakest particle
b. inertia, social coefficient
c. the old velocity of the best particle
d. None of the above.

17. The universal approximation theorem states that:

a. There should be enough neurons on the hidden layer in order to do the approximation.
b. Any function can be approximated with a proper neural network.
c. The conditions to approximate a function include the continuity of that function.
d. None of the above.

18. How does the artificial neuron process the information?

a. Based on the activation function.
b. Based on back propagation.
c. Based on the error.
d. None of the above.

All questions are worth one point. The mark is the number of correct questions. Each question has only one correct option!

1. The back-propagation algorithm:
   a. Is a training algorithm for ANNs.
   b. Can only be applied to shallow neural networks and is not suitable for deep learning architectures.
   c. Guarantees finding the optimal set of weights and biases in a finite number of iterations.
   d. None of the above.

2. What is a tensor?
   a. An black and white image with multi-channels
   b. A mathematical object that contains a one dimensional array of values.
   c. Generalizations of scalars, vectors, and matrices to an arbitrary number of indices.
   d. None of the above.

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

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

5. Which factor is the primary consideration when selecting an appropriate learning algorithm?
   a. Alignment with the desired data.
   b. Ability to predict cluster membership.
   c. Minimization of error through a cost function or loss function.
   d. Computational complexity of the target objective.

6. What is standardization?
   a. The process by which raw values are transformed into z-scores.
   b. A data transformation that introduces the scale effect.
   c. The operation that transforms continuous values into discrete ones.
   d. None of the above.

7. What crossover method(s) are correct for a binary representation in an GA?
   a. Insertion mutation
   b. There is no crossover for this representation
   c. Average crossover
   d. Uniform

8. In computer vision, we apply a filter over an image:
   a. In order to preprocess the input by subtracting some features from the initial image.
   b. Moving the kernel and adding to the part of the image that the kernel is hovering over.
   c. By using a convolution operation with a kernel
   d. None of the above.

30. The harmonic mean between precision and recall is:
    a. A measure for distance in certain clustering algorithms.
    b. The F1 score, which combines precision and recall into a single value.
    c. A metric used to evaluate the trade-off between precision and recall in deterministic algorithms.
    d. None of the above.

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

32. We can implement the infinite summation as a sum over a finite number of array elements:
    a. Such implementation is impossible in practice.
    b. In practice, we have two tensors: the input and the padding.
    c. By using a convolution operation.
    d. The input, the padding, and the kernel contain random numbers everywhere in the beginning.

33. What are the differences and similarities between the perceptron's rule and the delta rule?
    a. The perceptron rule and delta rule always converge to the global optimum and are not affected by local minima.
    b. The perceptron rule and delta rule do not require any iterative optimization process and can achieve optimal weights in a single pass.
    c. They both start with some random weights.
    d. None of the above.

34. For Unsupervised Learning, choose the appropriate statement:
    a. The training data comes in pairs (attributes, outputs).
    b. It finds an unknown function that groups the training data into several classes.
    c. The goal is to find a model or structure inside the data that is useful.
    d. None of those things.

35. The ReLu function
    a. Provides sparsity since y = 0 when x > 0
    b. Does not correct the problems that occur at sigmoid function
    c. It is a linear activation function
    d. Does not have a vanishing gradient when x > 0

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

37. In computer vision, we apply a filter over an image:
    a. By using a convolution operation with a kernel
    b. Moving the kernel and adding to the part of the image that the kernel is hovering over.
    c. In order to preprocess the input by subtracting some features from the initial image.
    d. None of the above.

38. Choose the correct answer.
    a. Backpropagation is insensitive to the choice of activation functions and can perform equally well with any activation function.
    b. The gradient descent is based on the error associated with the entire set of train data.
    c. Adding more training data will always result in better

generalization
for the ANN.
d. None of the above

39. What is standardizati
a. A data
intro
b.

Artificial Intelligence

Date: _____
AI-S1.2

Name: Marti Nicolae-Vlad
Grupa: 924

All questions are worth one point. The mark is the number of points obtained divided by the number of questions. Each question has only one correct option!

1. What are the correct statements about Decision Trees?
   a. The decision nodes are located at the terminal levels of the tree, while the result nodes are at the internal levels.
   b. Each leaf of the tree corresponds to a specific attribute or feature.
   c. They are used to divide a collection of articles into smaller sets by successively applying decision rules.
   d. Decision trees contain four types of nodes: decision nodes, hazard nodes, class nodes, and result nodes.

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

3. Clustering is:
   a. A one step process: testing.
   b. Another name for unsupervised learning.
   c. Using a labeled database.
   d. None of the above.

4. What is the feature of the database in training with indirect experience:
   a. It is based on useful feedback for some objective function.
   b. It comes in pairs (in/out).
   c. It is based on independent data with annotated content.
   d. None of the above.

5. An ANN has a structure of 26:15:10:2 with a sigmoid activation function. How many weights will have the first neuron from the first hidden layer?
   a. 2
   b. 15
   c. 26
   d. 10

6. When constructing a decision tree, the attribute selection can be:
   a. In preorder
   b. Random.
   c. Based on the top parent
   d. None of the above.

7. The induction phase of 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 to bottom or top to top

8. The back-propagation algorithm:
   a. Is a training algorithm for ANNs.
   b. Guarantees finding the optimal set of weights and biases in a finite number of iterations.
   c. Can only be applied to shallow neural networks and is not suitable for deep learning architectures.
   d. None of the above.

9. Select the correct statements for Cross-Entropy loss:
   a. Is never used when we apply the softmax transformation to the network's output.
   b. Is used in regressions
   c. Is the difference between two probability distributions for a provided set of occurrences or random variables.
   d. None of the above.

10. What are the characteristics of the back-propagation algorithm?
    a. It is crossing easy plateaus in the error function landscape.
    b. Does not require the derivatives of activation functions to be known at network design time.
    c. Is guaranteed to find the global minimum of the error function, not only the local minimum.
    d. None of the above.

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

12. On a ConvNet the feature learning:
    a. Allows a suite of tens or even hundreds of other small filters to be erased in order to detect more complex features in the image.
    b. Is performed before training the conv-layers.
    c. Will minimize the loss function by extracting the features that are most useful for classifying the images.
    d. None of the above.

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

14. The universal approximation theorem states that:
    a. Any function can be approximated with a proper neural network.
    b. There should be enough neurons on the hidden layer in order to do the approximation.
    c. The conditions to approximate a function include the continuity of that function.
    d. None of the above.

15. What is a tensor?
    a. A generalization of scalars, vectors, and matrices to an arbitrary number of indices.
    b. A mathematical object that contains a one dimensional array of values.
    c. An black and white image with multi-channels
    d. None of the above.

16. How does a ConvNet figure out what is in an image?
    a. Automatically detecting the weights for the kernels during training.
    b. By decomposing the features.
    c. There are usually three stages: several convolutions, a decomposing stage, a flatten stage.
    d. None of the above.

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

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

48. The limited model capacity of ANNs is overcome by:
    a. Reducing the number of artificial neurons.
    b. Adding nonlinearity to the model.
    c. Adding more layers at the output level.
    d. None of the above.

49. Select the correct statement for supervised learning:
    a. The aim is to provide an arbitrary output for a new input.
    b. The training data comes in an unpaired format: only attributes or only output.
    c. We search for a known function that maps the input attributes to the outputs.
    d. None of the above

50. How are the neurons connected into a feed forward ANN?
    a. Through a backward signal.
    b. Through an output with a neuron from the same layer.
    c. They are not connected.
    d. Through weighted links.

51. Which of the following statements is true when we apply a max pooling transformation over a tensor:
    a. We return the maximum value from the portion of the image covered by the kernel.
    b. We handle inputs of different types.
    c. We emphasize the features.
    d. We make the representation dependent on small translations of the input.

he L1 loss is:

a. It computes the average of the sum of absolute differences between actual values and predicted ones
b. Used for classification problems
c. Is also called the softmax loss
d. Is never used when the distribution has outliers

20. What is the difference between Particle Swarm Optimization (PSO) and Genetic
21. Algorithms (GA)?

a. PSO runs free until it converges to the solution, while GA never reach the solution
b. The particles have a memory, while the individuals don't.
c. GA has particles, and PSO has individuals.
d. GA uses a fitness function, and PSO doesn't.

22. How does the artificial neuron process the information?

a. Based on back propagation.
b. Based on the activation function.
c. Based on the error.
d. None of the above.

23. 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 an Artificial Cortex
b. A ConvNet captures the feature gradient dependencies in a time series
c. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms.
d. None of the above.

24. What can be used when comparing two algorithms?

a. The divergence of accuracy.
b. Overconfidence intervals.
c. Performance measures.
d. None of the above.

25. What are the advantages of going in depth in an ANN?

a. To avoid overfitting.
b. To speed up the network's evaluation.
c. We avoid underfitting the model.
d. None of the above.

26. How is the objective of a machine learning algorithm typically represented?

a. A database table.
b. Numeric functions.
c. Distributions of probability
d. A set of non - symbolic rules.

27. In order to overcome the perceptron's limits, we can:

a. Installing additional RAM directly enhances the neuron's processing power.
b. These limits can't be overcome.
c. Use neurons with a continuous threshold.
d. Applying the glitter property to the perceptron's activation function improves its ability to learn.

28. The vanishing gradients during backpropagation is:

a. An advantage in the training process that leads to faster convergence.
b. Since the derivative is zero in this case, on most of the domain, it does not affect us.
c. A problem tipical when the network has too many hidden layers
d. None of the above.

29. What problems can be solved with machine learning?

a. Ethical and moral considerations
b. Planning and classifications.
c. Creative problems that require innovation
d. None of the above.