

Artificial Intelligence

Date: _____
AI-S 2.1

Name:
Grupa:

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. Choose the right compatibility between the output layer's activation function and error function:
 - a. Mean square error with the linear function.
 - b. Binary cross entropy with logistic error with the sigmoid activation function.**
 - c. Cross-entropy error with arctangent.
 - d. None of the above.
2. On a ConvNet the feature learning:
 - a. Will minimize the loss function by extracting those features that are most useful for classifying the images.**
 - b. 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.
 - c. Is performed before training the conv-layers.
 - d. None of the above.
3. 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.
4. What is standardization?
 - a. The process of changing discrete values into continuous values.
 - b. A change to the data that adds the scale effect.
 - c. The process by which raw values are transformed into z-scores.**
5. What crossover method(s) are correct for a binary representation in a GA?
 - a. Insertion mutation
 - b. Average crossover
 - c. There is no crossover for this representation**
 - d. Uniform**
6. The universal approximation theorem states that:
 - a. The conditions to approximate a function include the continuity of that function.**
 - b. There should be enough neurons on the hidden layer in order to do the approximation.
 - c. Any function can be approximated with a proper neural network.
 - d. None of the above.
7. When choosing an appropriate learning algorithm, what is the primary factor to consider?
 - a. Computational complexity of the target objective.
 - b. Alignment with the desired data.
 - c. Ability to predict cluster membership.
 - d. Minimization of error through a cost function or loss function.**
8. Select the correct combination:
 - a. Output type: Binary, Output Distribution: Bernoulli, Output Layer: Softmax, Cost Function: Binary Cross Entropy
 - b. Output type: Discrete, Output Distribution: Multinoulli, Output Layer: arctangent, Cost Function: Cross Entropy.

- c. Output type: Continuous, Output Distribution: Gaussian, Output Layer: Linear, Cost Function: MSE.
d. None of the above.
9. 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 emphasize the features.
c. We make the representation dependent on small translations of the input.
d. We handle inputs of different types.
10. The induction phase of the process of building a DT is:
a. It eliminates the branches that reflect noise or exceptions.
b. It labels the new data with the build rules.
c. Based on the training data.
d. Works bottom to bottom or top to top
11. What attribute can be attributed to the database used in training with indirect experience?
a. It encompasses paired input/output data.
b. It is based on independent data with annotated content.
c. It is based on useful feedback for some objective function.
d. None of the above.
12. What 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.
13. What is the relationship typically expected between the training and testing data?
a. The test data should be representative of real-life experiences, while the training
- data can be based on our experience.
b. The two sets must overlap.
c. They should adhere to the same distribution.
d. None of the above.
14. Choose the correct answer.
a. Backpropagation is insensitive to the choice of activation functions and can perform equally well with any activation function.
b. In order to determine the steepest slope on gradient descent, we derive E based on w.
c. Adding more training data will always result in better generalization and performance of the ANN.
d. None of the above.
15. Select the correct statements for Cross-Entropy loss:
a. Is the difference between two probability distributions for a provided set of occurrences or random variables.
b. Is never used when we apply the softmax transformation to the network's output.
c. Is used in regressions
d. None of the above.
16. What is the difference between Particle Swarm Optimization (PSO) and Genetic algorithms (GA)?
a. The particles have a memory, while the individuals don't.
b. PSO runs free until it converges to the solution, while GA never reaches the solution.
c. GA has particles, and PSO has individuals.
d. GA uses a fitness function, and PSO doesn't.
17. What are the main advantages of Deep Convolutional Neural Networks?
a. A ConvNet captures the feature gradient dependencies in a time series

- b. The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in an Artificial Cortex
- c. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms.
- d. None of the above.
18. What are the characteristics of the back-propagation algorithm?
- Does not require the derivatives of activation functions to be known at network design time.
 - It is crossing easy plateaus in the error function landscape.
 - Is guaranteed to find the global minimum of the error function, not only a local minimum.
 - None of the above.
19. What is the proper encoding for an individual in ANNs?
- A computer program that learns to classify and performs regressions.
 - A set of weights used to propagate a signal.
 - An array of bits that encode proper information related to the solution.
 - There are no individuals in ANN.
20. The ReLu function:
- Provides sparsity since $y = 0$ when $x < 0$
 - Has a vanishing gradient when $x > 0$
 - It is a linear activation function
 - Does not correct the problems that occur at sigmoid function
21. The information gain ratio:
- It enhances an attribute by integrating a new term that depends on spreading degree.
 - It aims to reduce a bias towards multivalued attributes.
 - Is the ratio between the information gain and the split information.
 - None of the above.
22. The vanishing gradients during backpropagation:
- Since the derivative is zero in this case, on most of the domain, it does not affect us.
 - An advantage in the training process that leads to faster convergence.
 - A problem typical when the network has too many hidden layers
 - None of the above.
23. What is a tensor?
- A mathematical object that contains a one dimensional array of values.
 - A generalization of scalars, vectors, and matrices to an arbitrary number of indices.
 - An black and white image with multi-channels
 - None of the above.
24. What are the advantages of going in depth in an ANN?
- We avoid underfitting the model.
 - To avoid overfitting.
 - To speed up the network's evaluation.
 - None of the above.
25. Which statement about supervised learning is correct?
- The aim is to provide an arbitrary output for a new input.
 - The training data comes in an unpaired format: only attributes or only output.
 - We search for a known function that maps the input attributes to the outputs.
 - None of the above.
26. How does the artificial neuron process the information?
- Based on the activation function.
 - Based on the error.
 - Based on back propagation.
 - None of the above.

27. Clustering means:

- a. a one step process - testing.
- b. Another name for unsupervised learning.
- c. using a labeled database.
- d. None of the above.

28. The harmonic mean between precision and recall is:

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

29. The L1 loss is:

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

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

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

31. The back-propagation algorithm:

- a. Can only be applied to shallow neural networks and is not suitable for deep learning architectures.
- b. Is a gradient descent method.
- c. Guarantees finding the optimal set of weights and biases in a finite number of iterations.
- d. None of the above.

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

a. Installing additional RAM or enhances the neuron's processing power.

- b. Increase the number of neurons.
- c. These limits can't be overcome.
- d. Applying the glitter property to the perceptron's activation function improves its ability to learn.

33. An ANN has a structure of 14:15:10:2 with a sigmoid activation function. How many weights will have the first neuron from the first hidden layer?

- a. 10
- b. 15
- c. 2
- d. 14

34. What kinds of problems can machine learning solve?

- a. Regressions.
- b. Creative problem-solving and innovation
- c. Moral and ethical concerns
- d. None of the above.

35. 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 difference between the real output of the network and the desired output.
- b. Based on the output of the hidden layer.
- c. Based on an induction formula.
- d. None of the above.

36. In an artificial neuron, the transfer function:

- a. Is the sigmoid function.
- b. Involves crossovers to compute the output.
- c. Represents the equation of a hyperplane.
- d. Relies on the neuron's ability to perform magical transformations.

37. When making a decision tree, the choice of attributes can be:

- a. Based on the gained information.
- b. Based on the top parent

- c. In inorder
d. None of the above.
38. Which option is commonly employed when comparing two algorithms?
a. Performance measures.
b. Overconfidence intervals.
c. The divergence of accuracy.
d. None of the above.
39. limited model capacity of ANNs is overcome by:
a. Reducing the number of artificial neurons.
b. Adding more layers at the output level.
c. Adding nonlinearity to the model.
d. None of the above.
40. For Unsupervised Learning, choose the appropriate statement:
a. The goal is to find a model or structure within the data that is useful.
b. The data for training come in pairs: (attributes, output).
c. It finds an unknown function that groups the training data into different classes.
d. None of the above.
41. The activation logistic function:
a. Has limitations regarding the input domain.
b. Is a linear function.
c. Suffers from a vanishing gradient.
d. None of the above.
42. The Perceptron's algorithm:
a. Is based on maximizing the error for a given set of train data.
b. The error is the difference between what the real output y is and what the perceptron's output \hat{y} is for a given input.
c. It changes the weights based on the inverse error associated with a train data instance.
d. None of the above.
43. In computer vision, we apply a filter to an image:
- a. In order to preprocess the input by subtracting some features from the initial image.
b. By using a convolution operation with a kernel
c. Moving the kernel and adding to the part of the image that the kernel is hovering over.
d. None of the above.
44. What are the correct statements about Decision Trees?
a. Each leaf of the tree corresponds to a specific attribute or feature.
b. The decision nodes are located at the internal levels of the tree, while the result nodes are at the terminal levels.
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.
45. Which representation is commonly used to define the objective in a machine learning algorithm?
a. Distributions of probability
b. A database table.
c. A set of non-symbolic rules.
d. Numeric functions.
46. The softmax function:
a. It incorporates the cross entropy function.
b. Is used in regressions.
c. Transforms probabilities the output scores for the classes.
d. None of the above.
47. How are the neurons connected into a feed forward ANN?
a. They are not connected.
b. Through a backward signal.
c. Through weighted links.
d. Through an output with a neuron from the same layer.

48. We can implement the infinite summation as a sum over a finite number of array elements:

- a. Such implementation is impossible in practice, and the elements:
- b. The input, the padding, and the kernel contain random numbers.
- c. In practice, we have two tensors: everywhere, we have two tensors.
- d. By using a convolution operation.

49. What are the differences and similarities between the perceptron's rule and the delta rule?

- a. In the delta's rule, the model's quality is established based on all
- b. By using a convolution operation.
- c. In practice, we have two tensors: everywhere, we have two tensors.
- d. By using a convolution operation.

data, while the perceptron rule is based on one entry at a time, the perceptron rule and delta rule do not require any iterative optimization process. Weights in a single pass, the perceptron rule and delta rule achieve optimality and can be used interchangeably in any scenario.

- a. The values of the derivative.
- b. The vanishing gradient.
- c. The identical and can be used interchangeably in any scenario.
- d. None of the above.

50. The "dying ReLU" problem refers to:

- a. The values of the function.
- b. The vanishing gradient.
- c. The values of the derivative.
- d. None of the above.

Artificial Intelligence

Date: _____
AI-S1.1

Name:
Grupa:

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. 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
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.

9. 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.

10. 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.

11. 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.

12. What is the proper encoding for an individual 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.

13. What is the feature of the database in training 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 statements for supervised learning:
- a. The aim is to provide an 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.

not are the correct statements about decision trees?
a. The terminal levels of the tree.
b. Decision trees contain four types of nodes; decision nodes, class nodes, nodes, class nodes, and leaves.
c. Each leaf of the tree contains a specific value.
d. They are called decision trees.

e. It is a decision tree.

14. Select the correct statement
a. Supervised learning:
b. The aim is to provide an arbitrary output for a new input.
c. We search for a known function that maps the input attributes to the output.
d. The training data comes in an unpaired format only attributes or only output.
15. Does a ConvNet figure out what is implicitly for the kernels during the three stages: a. Detecting the features b. Classifying the features c. Generating the output
16. 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. Decision trees contain four types of nodes: decision nodes, hazard nodes, class nodes, and result nodes.
c. Each leaf of the tree corresponds to a specific attribute or feature.
d. They are used to divide a collection of articles into smaller sets by successively applying decision rules.
17. What can be used when comparing two algorithms?
a. Overconfidence intervals.
b. The divergence of accuracy.
c. Performance measures.
d. None of the above.
18. 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. 10
b. 26
c. 15
d. 2
19. What problems can be solved with machine learning?
a. Ethical and moral considerations
b. Creative problems that require innovation
c. Planning and classifications.
d. None of the above.
20. The "dying ReLU" problem refers to:
a. The vanishing gradient.
b. The values of the function.
c. The values of the derivative.
d. None of the above.
21. The ReLU function
a. Provides sparsity since $y = 0$ when $x > 0$
b. Does not correct the problems that occur at sigmoid function
22. It is a linear activation function
a. Does not have a vanishing gradient when $x > 0$
23. How is the objective of a machine learning algorithm typically represented?
a. A set of non-symbolic rules.
b. Distributions of probability
c. A database table.
d. Numeric functions.
24. In an artificial neuron, the transfer function:
a. Utilizes entanglement to calculate the output.
b. Represents the equation of a hyperplane.
c. Is the sigmoid function.
d. Requires the neuron to perform complex mathematical operations with imaginary numbers.
25. When constructing a decision tree, the attribute selection can be:
a. Based on the top parent
b. In preorder
c. Random.
d. None of the above.
26. 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, not being 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.
27. The induction phase in the process of building a DT:
a. Eliminates the branches that reflect noise or exceptions.
b. Labels the new data with the build rules.
c. Is based on the training data.

30. The limited model capacity of ANNs is overcome by:
a. Reducing the number of artificial neurons.
b. Adding more layers at the output level.
c. Adding nonlinearity to the model.
d. None of the above.
31. What are the main advantages of Deep Convolutional Neural Networks?
a. A ConvNet captures the feature gradient dependencies in a time series.
b. The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in an Artificial Cortex.
c. Pre-processing required in a ConvNet is much lower as compared to other classification algorithms.
d. None of the above.
32. What is the difference between Particle Swarm Optimization (PSO) and Genetic Algorithms (GA)?
a. GA uses a fitness function, and PSO doesn't.
b. GA has particles, and PSO has individuals.
c. The particles have a memory, while the individuals don't.
d. PSO runs free until it converges to the solution, while GA never reaches the solution.
33. What sort of problems can a perceptron solve?
a. It depends on the structure.
b. Linear separations of elements from the domain.
c. XOR problem.
d. None of the above.
34. The activation logistic function:
a. Has limitations regarding the input domain.
b. Suffers from a vanishing gradient.
35. What are the advantages of going depth in an ANN?
a. To speed up the network evaluation.
b. We avoid underfitting the model.
c. To avoid overfitting.
d. None of the above.
36. In order to overcome the perceptron's limits:
a. We can't overcome these limits.
b. We can install additional RAM to directly enhance the neuron's processing power.
c. We can use neurons with a continuous threshold.
d. We can apply the glitter property to the perceptron's activation function to improve its ability to learn.
37. The L1 loss:
a. Is never used when the distribution has outliers.
b. It computes the average of the sum of absolute differences between actual values and predicted ones.
c. Used for classification problems
d. Is also called the softmax loss
38. The Perceptron's algorithm:
a. It changes the weights based on the inverse error associated with a train data instance.
b. The error is the difference between what the real output y is and what the perceptron's output o is for a given input.
c. Is based on maximizing the error for a given set of train data.
d. None of the above.
39. How are the neurons connected into a feed forward ANN?
a. Through an output with a neuron from the same layer.
b. Through a backward signal.
c. Through weighted links.
40. Clustering is:
a. Using a labeled database.
b. A one step process finding learning rules for unstructured data.
c. None of the above.
41. We can implement the initial connection elements:
a. a sum over a wide number of any learned context neurons.
b. The speed, the padding and the everywhere in the beginning.
c. A matrix to reduce the error in the learning optimization.
d. A vector and to reduce the error in the learning optimization.

35. What are the advantages of going in depth in an ANN?
 a. To speed up the networks
 b. We avoid underfitting the model.
 c. To avoid overfitting.
 d. None of the above.
36. Is it a linear function.
 a. Yes
 b. No
37. Which of the following statements is true when we apply a max pooling transformation over a tensor:
 a. We make the representation dependent on small translations of the input.
 b. We emphasize the features.
 c. We return the maximum value from the portion of the image covered by the kernel.
 d. We handle inputs of different types.
38. We can implement the infinite summation as a sum over a finite number of array elements:
 a. The input, the padding, and the kernel contain random numbers everywhere in the beginning.
 b. In practice, we have two tensors: the input and the padding.
 c. Such implementation is impossible in practice.
 d. By using a convolution operation.
39. On a ConvNet the feature learning:
 a. Will minimize the loss function by extracting those features that are most useful for classifying the images.
 b. Is performed before training the conv-layers.
 c. 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.
 d. None of the above.
40. 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.
41. Clustering is:
 a. Using a labeled database.
 b. A one step process: testing.
 c. Another name for unsupervised learning.
 d. None of the above.
42. They are not connected.
 a. We emphasize the features.
 b. We return the maximum value from the portion of the image covered by the kernel.
 c. We handle inputs of different types.
43. The harmonic mean between precision and recall is:
 a. A measure for distance in certain clustering algorithms.
 b. A metric used to evaluate the trade-off between precision and recall in deterministic algorithms.
 c. The F1 score, which combines precision and recall into a single value.
 d. None of the above.
44. The vanishing gradients during backpropagation is:
 a. A problem typical when the network has too many hidden layers.
 b. An advantage in the training process that leads to faster convergence.
 c. Since the derivative is zero in this case, on most of the domain, it does not affect us.
 d. None of the above.
45. What is the relationship between the training and testing data?
 a. The test data should reflect real-life experiences, while the training data can be based on theoretical experiences.
 b. They should follow the same distribution.
 c. The two sets must overlap.
 d. None of the above.
46. Choose the correct answer.
 a. The gradient descent is based on the error associated with the entire set of train data.
 b. Adding more training data will always result in better generalization and performance for the ANN.