



Artificial Intelligence Course - Multiple Choice Questions

1. **In order to overcome the perceptron's 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
2. **What problems can be solved with machine learning?**
 - a) regressions
 - b) ???
 - c) plannings and classifications
 - d) none of the above
3. **The objective of a machine learning algorithm can be represented as:**
 - a) numeric functions
 - b) probabilistic functions
 - c) a set of symbolic rules
 - d) none of the above
4. **The induction phase for the process of building a decision tree is:**
 - a) it labels the new data with the build rules
 - b) it eliminates the branches that reflect noise & exceptions
 - c) based on the training data
 - d) works bottom-up or top-down

5. What is a tensor?

- a) an image with multiple
- b) a mathematical object
- c) generalizations of scalars, vectors, and matrices in an arbitrary number of indices
- d) none of the above

6. 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: BCE
- d) none of the above

7. 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.
- b) The perceptron's rule is based on gradient descent.
- c) They both start with some random weights.
- d) Perceptron's rule recomputes the weights based on while the delta rule.

8. 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

9. The limited model capacity of an ANN can be 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
10. **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 necessarily form 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
11. **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
12. **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
13. **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

14. How does a 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

15. 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

16. 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 data test should be disjunct sets
- d) none of the above

17. The harmonic mean between the precision and the recall is:

- a) a metric for distance in a reinforcement learning algo
- b) the F1 score
- c) a statistical metric used to evaluate performance in a supervised learning process
- d) none of the above

18. 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

19. **How are the neurons connected in a feed forward ANN?**

- a) through a backward signal
- b) through an output with a neuron -(missing word)- on the name layer
- c) they are not connected
- d) through weighted links

20. **What sort of problems can a perceptron solve?**

- a) linear separation of elements from the domain
- b) XOR problem
- c) depends on the structure
- d) none of the above

21. **The vanishing gradients during backpropagation are:**

- a) not affecting us since the derivative is zero in this case on most of the domain
- b) are advantage in the training process that leads to faster convergence
- c) a typical problem when the network has too many hidden layers
- d) none of the above

22. **The activation logistic function:**

- a) is a linear function
- b) suffers from a vanishing gradient
- c) has limitations regarding the output domain
- d) none of the above

23. **An ANN with a structure of 226:15:10:2 with a sigmoid activation function. How many weights will the first neuron from the first hidden layer have?**

- a) 2
- b) 226
- c) 10
- d) 15

24. **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

25. **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

26. **On a ConvNet, 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

27. **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

28. **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

29. **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.

30. **The activation logistic function:**

- a) Suffers from vanishing gradient.
- b) is a linear function.
- c) has limitations regarding the input domain.
- d) none of the above.

31. **Clustering is:**

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

32. **What is a 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.
33. **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
34. **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.
35. **The induction phase of the process of building a Decision Tree is:**
- a) It labels the new data with the built 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.
36. **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.
37. **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.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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.

43. 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. A black and white image with multi-channels.
- d. None of the above.

44. 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.

45. 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.

46. 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.

47. The 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

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

a. PSO runs free until it converges to the solution, while GA never reaches 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

49. How does the artificial neuron process the information?

a. Based on backpropagation

b. Based on the activation function

c. Based on the error

d. None of the above

50. 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 preprocessing required in a ConvNet is much lower as compared to other classification algorithms

d. None of the above

51. **What can be used when comparing two algorithms?**
- a. The divergence of accuracy
 - b. Overconfidence intervals
 - c. Performance measures
 - d. None of the above
52. **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
53. **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
54. **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
55. **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 typical when the network has too many hidden layers
 - d. None of the above

56. **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
57. **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.
58. **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.
59. **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.
60. **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.

31. **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.

32. **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 $x < 0$

33. **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.

34. **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.

35. **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 and performance for the ANN.
- d. None of the above.

36. **What is standardization?**

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

37. **What crossover method(s) are correct for a binary representation in a GA?**

- a. Uniform
- b. There is no crossover for this representation
- c. Average crossover
- d. Insertion mutation

38. **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 an induction formula.
- c. Based on the difference between the real output of the network and the desired output.
- d. None of the above.

39. **Which factor is the primary consideration when selecting an appropriate learning algorithm?**

- a. Ability to predict cluster membership.

- b. Minimization of error through a cost function or loss function.
 - c. Alignment with the desired data.
 - d. Computational complexity of the target objective.
70. **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.
71. **In an artificial neuron, the transfer function:**
- a. Represents the equation of a hyperplane.
 - b. Utilizes entanglement to calculate the output.
 - c. Requires the neuron to perform complex mathematical operations with imaginary numbers.
 - d. Is the sigmoid function.
72. **The Perceptron's algorithm:**
- a. It changes the weights based on the inverse error associated with a train data instance.
 - b. Is based on maximizing the error for a given set of train data.
 - c. The error is the difference between what the real output y is and what the perceptron's output o is for a given input.
 - d. None of the above.
73. **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.

74. **What sort of problems can a perceptron solve?**
- a. Linear separations of elements from the domain.
 - b. It depends on the structure.
 - c. XOR problem.
 - d. None of the above.
75. **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.
76. **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.
77. **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.
78. **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.