

## Exam simulation 1

1. Nowadays, the usage of classical feature extraction and data analysis methods is outdated since the capability of the recent deep learning models and methods made them obsolete and not more present in the common practice.
  - a. True
  - b. False**
2. Artificial Intelligence can be applied to the following sectors.
  - a. Robotics
  - b. Information Extraction
  - c. All the above**
3. Artificial neural networks are capable to learn human biases.
  - a. False: the achievable complexity of the artificial neural networks is so far from the complexity of the human brain to make impossible to mimic this characteristic
  - b. False: human biases are not reproducible nor measurable
  - c. True**
4. Recent artificial intelligence models can solve analogy puzzles.
  - a. True**
  - b. False
5. Considering the “Data knowledge spectrum plot” discussed in class, the minimum amount of data required is in the following case.
  - a. No knowledge about the model generating the data is available
  - b. A statistical model of the process is available
  - c. A mathematical model of the process is available**
6. It is possible to think to the single datum in input to the neural network as a point in the “input space” of the model, even if the input is a single value, a N dimensional vector, or an image.
  - a. True**
  - b. False
7. It is correct to say the one of the key features of an intelligent artificial system is the capability to learn (even if only a limited sense) and/or get better in time.
  - a. True**
  - b. False
8. According to the Andries Engelbrecht definition of Computational intelligence what of the following is not included?
  - a. Artificial Neural Networks
  - b. Evolutionary Computing
  - c. Swarm Intelligence
  - d. Artificial immune system
  - e. Fuzzy Systems
  - f. All the above are included**
9. According to the class discussion of the Gestalt capability, what of the following sentences is more correct?
  - a. The Gestalt capability is a typical feature present by-design in the model of classical neural networks
  - b. The Gestalt capability is a typical feature present by-design in the model of deep learning neural networks

- c. The Gestalt capability is a typical human feature not well (yet) mimicked in current artificial networks**
- 10. The following activity: Data Selection, Data Filtering, Data Enhancing
  - a. Are part of the job of the artificial intelligent specialist in normal activities
  - b. Contribute to keep lower the complexity of the learning task
  - c. All the above**
  - d. Are part of the classical machine learning approaches and they are (correctly) no longer used in deep learning applications
- 11. The Mean Squared Error is typically present in what step of the design.
  - a. Representation
  - b. Evaluation**

## Exam simulation 2

1. Considering IoT devices as source of data for external intelligent systems (IS is not intended to be embedded into the IoT device), what kind of IoT devices can be really used?
  - a. Passive data IoT devices
  - b. Active data IoT devices
  - c. Dynamic data IoT devices
  - d. All of the above**
  - e. None of the above
2. Referring to the class discussion, the (correct) design practice for neural networks considers
  - a. Start with deep learning models since they are the cutting edge and most advanced technology that we have now
  - b. Start with deep learning models since they are the cutting edge and most advanced technology we have now, and then use classical method as reference
  - c. Start with simple neural networks before to consider deep learning models**
3. The missing values can also be occupied by computing mean, mode or median of the observed given values.
  - a. This is very unusual and not common in practice
  - b. This is a very simple and effective solution in case the learning method is not capable to deal with missing data**
  - c. This is not possible, since that is just descriptive statistics about the features, and cannot be used to fill missing data
4. Referring to the class discussion on data leakage what is the worst situation?
  - a. The unwanted leakage of data from test dataset to training data set**
  - b. The unwanted leakage of data from training dataset to test data set
  - c. None of the above since transferring data from test and/or training dataset is normal when the accuracy of the model is tested
5. An additional information can allow the model to learn or know something that it otherwise would not know and in turn invalidate the estimated performance of the model being constructed. This is called
  - a. Data leakage**
  - b. Data pre-processing
  - c. Data harmonization
  - d. Data wrangling
6. The degrees of freedom for a given problem are the number of independent problem variables which must be specified to uniquely determine a solution. Hence the #DoF is important to be considered
  - a. To design the number of vectors in the learning dataset.
  - b. To avoid overfitting problem in the model
  - c. All the above**
  - d. None of the above
7. About the cosine metrics it is possible to say that
  - a. Two vectors with the same orientation have a cosine similarity of 1
  - b. Two vectors oriented at 90° relative to each other have a similarity of 0
  - c. All of the above**

- d. None of the above
- 8. What similarity feature/features discussed in class offers/offers the property to allow a fast comparison based on a short 1D vector of elements or bits
  - a. phash
  - b. ahash
  - c. **All the above**
  - d. Cross-correlation
- 9. In agreement to the class discussion, which description better describes the design activity?
  - a. Similarity in the dataset requires more space and processing time
  - b. Similarity in the dataset can improve generalization
  - c. **Both of the above**
  - d. None of the above
- 10. In agreement to the class discussion, in a dataset of 1100 labelled images, the search for duplications is typically achieved...
  - a. by manual exploration of the dataset for better results since the number of images is not critical
  - b. **by automatic iterations**
- 11. In agreement to the class discussion, what kind of labelling error is generally the worst case for the accuracy of the generalization of the model? ERR1 = Duplications with same labels ERR2 = Duplications with different labels
  - a. ERR1
  - b. **ERR2**
  - c. ERR1 = ERR2
- 12. According to the class discussion, about the relationship between the operation of cross-correlation and convolution it is possible to say that:
  - a. **They are very similar in meaning and mathematical expression**
  - b. Despite the mathematical expression is similar, the meaning and their use is completely different
  - c. There is no specific relationship since they are different in meaning and mathematical expressions
- 13. According to the class discussion, what is the characteristic of the self-correlation ( $O = xcor2(A, A)$ ) map produced by a generic image?
  - a. A flat and noisy central plateau
  - b. **An evident spike at the center with a very well-defined maximum**
  - c. It is not possible to create an autocorrelation map from one single images, two different images are needed
- 14. If your data set contains extreme outliers, it is better to use as preprocessing
  - a. **Feature clipping**
  - b. Min-max normalization
  - c. Z' norm
- 15. A logarithmic scaling to one feature values is typically applied in a case of
  - a. Outliers' presence
  - b. Negative values
  - c. **A very large range in the values (>0)**
- 16. According to the scientific visualization rules presented in class, if you are plotting many figures of merit obtained by your trained neural network on a new dataset,

which is the correct ranking of visual attributes to be used? Left: low accuracy Right: HIGH ACCURACY

- a. Color intensity > Hue > Length
  - b. Area > Length > Hue
  - c. Slope > Angle > Volume
  - d. Hue > Area > Length**
17. According to the scientific visualization rules presented in class, is it possible to plot a graphical representation of the confidence level of your figures of merit of your trained model?
- a. No, it is a statistical index with different units and meaning and hence cannot be represented in the same plot
  - b. Yes, the confidence interval data have the same units and meaning, and they can be represented in the same plot**
18. According to the discussion presented in class about the data visualization, and considering the following steps of the design workflow 1) Get Data, 2) Clean Manipulate Data, 3) Train models, 4) Test Data, 5) Improve the design, which are the main step/steps where data visualization should be involved?
- a. #1
  - b. #5
  - c. #1 and #5
  - d. #2, #3 and #5**
19. According to the discussion presented in class about the similarity, consider an image  $A(x, y)$  with internal similarity (repetitions of patterns). What happens to the output of the self-cross correlation ( $O = \text{xcorr2}(A, A)$ )
- a. It is not possible to apply the cross correlation to the same image
  - b. Output O tends to be a flat plateau with one clear central peak
  - c. Output O tends to have many peaks and one evident maximum**
  - d. Output O tends to have many equivalent peaks with the same maximum value

### Exam simulation 3

1. You have a dataset  $X$  of 1000 samples and number of features  $F = 4$  features. You want to reduce the number of features  $F$  to 2 for data visualization. According to the goal, consider the following options. OPTION A: Apply PCA to  $X$  and select only the first 2 Principal Components. OPTION B: Apply the Feedforward Feature Selection to  $X$  and select only the first 2 more relevant features.
  - a. **Option A is possible. Option B is possible.**
  - b. Option A is NOT possible. Option B is possible.
  - c. Option A is possible. Option B is NOT possible.
  - d. Option A is NOT possible. Option B is NOT possible.
2. You have a feature in your dataset with the following values  $F1 = [-5, 0, +5]$ , which normalization will give you the following  $F1\_norm = [0, 0.5, 1]$ 
  - a. **Min-MAX**
  - b. Z-score
  - c. Clipping
  - d. A different type of normalization
3. According to the class discussion, in general for a given small dataset  $X$ , if you train a feed-forward neural models (of the same type) with an increasing number of neurons, which case is more probable?
  - a. **None of the below**
  - b. The training error and the validation will decrease indefinitely
  - c. The training error will increase
  - d. The validation error will decrease indefinitely
4. According to the class discussion, in a cross-validation single test, which train/test partition of the samples will provide the lower training error but the lower confidence in the test results?
  - a. **Training set = 99%, Test Set = 01%**
  - b. Training set = 75%, Test Set = 25%
  - c. Training set = 50%, Test Set = 50%
  - d. Training set = 25%, Test Set = 75%
  - e. Training set = 01%, Test Set = 99%
5. According to the class discussion, what kind of activity can be performed on the test set?
  - a. **All the below**
  - b. Mean test error estimation
  - c. Mean test error estimation and standard deviation
  - d. Confusion matrix test
6. According to the class discussion, what kind of activity can be performed on the train set?
  - a. **All the other options**
  - b. Design of the #of neurons
  - c. Design of the #of layers
  - d. Normalization
  - e. PCA
7. According to the class discussion, where can be performed the feature engineering?
  - a. **Only on the train set**
  - b. Only on the test set

- c. On the train set and the test set
  - d. Not on the train, not on the test set, but only on a different dataset
8. A simple k-Fold Cross Validation procedure may
- a. **Lead to disarranging the proportion of examples from each class in the test partitions**
  - b. Making impossible to process the test error
  - c. Get stuck into one the local minima
  - d. Produce severe overfitting
  - e. None of the other answers
9. Which option is correct?
- a. **From the confusion matrix is possible to process the classification error**
  - b. From the confusion matrix is possible to process the classification error and vice versa
  - c. The confusion matrix is applicable only to binary classification systems
  - d. The classification error is equal to the sum of the diagonal elements of the confusion matrix
10. According to the notation used in class, which kind of a model is described by the equation  $f(x) = \text{sgn}(w \cdot x + b)$
- a. **Liner classifier**
  - b. Liner regressor
  - c. Soft-max neuron
  - d. Sigmoidal neuron
  - e. Gradient descent formula
  - f. Number of the model's parameters
11. According to the notation used in class, which kind of a classifier is better described by the following definition: "the output is the label produced by the most probable classifier"
- a. **Bayes Optimal Classifier**
  - b. Supervised Classifier
  - c. K-means
  - d. None of the other options
12. According to the class discussion the kNN classifier, what kind of learning is it?
- a. **Instance-based Learning**
  - b. Eager Learning
  - c. Hard-limited Learning
  - d. Unsupervised Clustering
  - e. None of the other options
13. According to the class discussion, what is the classifier with the following properties: not based on neural techniques; it's deterministic with no random initialization; perfect repeatability; a minimum number of parameters is needed; learning is very simple but effective; perfect explain ability
- a. **kNN**
  - b. Linear classifier
  - c. Decision Tree
  - d. K-means
  - e. None of the other options

14. According to the class discussion on kNN classifiers about the k parameter and its relationship to regularization of the decision boundaries and the computational complexity, what is the correct option about larger values of k?
  - a. **More regularization and more complexity**
  - b. Less regularization and more complexity
  - c. More regularization and less complexity
  - d. Less regularization and less complexity
  - e. The parameter k is not related to regularization and complexity
15. According to the class discussion on PCA what is the correct option?
  - a. PCA vectors are originating from the center of mass of the points
  - b. All subsequent principal component vectors are orthogonal
  - c. **All the other options**
16. According to the class discussion on PCA what is the correct option?
  - a. All subsequent principal component vectors are orthogonal
  - b. The variance of the data projection on the first PCA vectors is maximized
  - c. **All the other options**
17. According to the class discussion about unsupervised learning, what is the method with the following properties: you need to specify the number of clusters k in advance, is unable to handle noisy data and outliers, it is not suitable to discover clusters with non-convex shapes
  - a. **K-means**
  - b. kNN
  - c. Decision tree
  - d. None of the other options
18. According to the class discussion, considering the equation of the backpropagation in a feedforward neural network of weight  $w_{ij}$  connected to the following output neuron  $k$ , which is the missing term?  $DELTA W_{ij} = ? * y_j * delta_k$ 
  - a. **??? = alfa (the regularization term < 1)**
  - b. ??? = alfa (the regularization term > 1)
  - c. ??? =  $x_j$  (the input vector)
  - d. ??? =  $x_j$  (the input vector error)
19. According to the class discussion, considering a general CNN architecture, what is the sequence of modules which is more likely
  - a. **Input layer → Convolution → Relu → Max Pooling → Softmax → Output layer**
  - b. Input layer → Relu → Convolution → Max Pooling → Softmax → Output layer
  - c. Input layer → Relu → Max Pooling Convolution → Softmax → Output layer
  - d. Input layer → Relu → Max Pooling → Softmax → Convolution → Output layer
20. According to the class discussion, considering a standard intelligent vision system, which capability can be processed onboard on a recent smart industrial camera?
  - a. Segmentation
  - b. Segmentation, Measurement
  - c. Segmentation, Measurement, Classification with trained non-deep models
  - d. **Segmentation, Measurement, Classification with trained deep models**
  - e. Segmentation, Measurement, Classification with trained deep models and training of deep models



21. According to the class discussion, Traditional Segmentation methods are quite useful to produce blobs or object candidates to be further processed by deep models for classification or measurements. Traditional Segmentation methods can be partitioned in
- a. Global knowledge, Edge-based
  - b. Edge-based, Region-based
  - c. Global knowledge, Edge-based, Region-based**
  - d. None of the other options
22. According to the class discussion referred to edge computing, is it possible to process images with trained deep learning models on external small, dedicated devices connect via USB connection?
- a. True: the usage of dedicated processors and the USB bandwidth make this option possible**
  - b. False: the USB bandwidth make this option not possible
  - c. False: the needed computational complexity needed to run trained deep learning models make this option not possible
  - d. False: the bandwidth and the computational complexity need to process images with trained deep learning model is not adequate
23. According to the class discussion what is Greedy Layer-Wise Training?
- a. A supervised training step to improve auto-encoders
  - b. A supervised training step to classical feedforward networks
  - c. An unsupervised training step to classical feedforward networks
  - d. An unsupervised training step to improve auto-encoders**