# 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?
   1. Passive data IoT devices
   2. Active data IoT devices
   3. Dynamic data IoT devices

## All of the above

* 1. None of the above

1. Referring to the class discussion, the (correct) design practice for neural networks considers
   1. Start with deep learning models since they are the cutting edge and most advanced technology that we have now
   2. Start with deep learning models since they are the cutting edge and most advanced technology we have now, and then use classicals method as reference

## Start with simple neural networks before to consider deep learning models

1. The missing values can also be occupied by computing mean, mode or median of the observed given values.
   1. This is very unusual and not common in practice

## This is a very simple and effective solution in case the learning method is not capable to deal with missing data

* 1. This is not possible, since that is just descriptive statistics about the features, and cannot be used to fill missing data

1. Referring to the class discussion on data leakage what is the worst situation?

## The unwanted leakage of data from test dataset to training data set

* 1. The unwanted leakage of data from training dataset to test data set
  2. None of the above since transferring data from test and/or training dataset is normal when the accuracy of the model is tested

1. 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

## Data leakage

* 1. Data pre-processing
  2. Data harmonization
  3. Data wrangling

1. 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
   1. To design the number of vectors in the learning dataset.
   2. To avoid overfitting problem in the model

## All the above

* 1. None of the above

1. About the cosine metrics it is possible to say that
   1. Two vectors with the same orientation have a cosine similarity of 1
   2. Two vectors oriented at 90° relative to each other have a similarity of 0

## All of the above

* 1. None of the above

1. What similarity feature/features discussed in class offers/offer the property to allow a fast comparison based on a short 1D vector of elements or bits
   1. phash
   2. ahash

## All the above

* 1. Cross-correlation

1. In agreement to the class discussion, which description better describes the design activity?
   1. Similarity in the dataset requires more space and processing time
   2. Similarity in the dataset can improve generalization

## Both of the above

* 1. None of the above

1. In agreement to the class discussion, in a dataset of 1100 labelled images, the search for duplications is typically achieved...
   1. by manual exploration of the dataset for better results since the number of images is not critical

## by automatic iterations

1. 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 EER2 = Duplications with different labels
   1. ERR1

## ERR2

* 1. ERR1 = EE2

1. According to the class discussion, about the relationship between the operation of cross-correlation and convolution it is possible to say that:

## They are very similar in meaning and mathematical expression

* 1. Despite the mathematical expression is similar, the meaning and their use is completely different
  2. There is no specific relationship since they are different in meaning and mathematical expressions

1. According to the class discussion, what is the characteristic of the self-correlation

(𝑂 = 𝑥𝑐𝑜𝑟2(𝐴, 𝐴)) map produced by a generic image?

* 1. A flat and noisy central plateau

## An evident spike at the center with a very well-defined maximum

* 1. It is not possible to create an autocorrelation map from one single images, two different images are needed

1. If your data set contains extreme outliers, it better to use as preprocessing

## Feature clipping

* 1. Min-max normalization
  2. Z’ norm

1. A logarithmic scaling to one feature values is typically applied in a case of
   1. Outliers’ presence
   2. Negative values

## A very large range in the values (>0)

1. 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
   1. Color intensity > Hue > Length
   2. Area > Length > Hue
   3. Slope > Angle > Volume

## Hue > Area > Length

1. 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?
   1. No, it is a statistical index with different units and meaning and hence cannot be represented in the same plot

## Yes, the confidence interval data have the same units and meaning, and they can be represented in the same plot

1. 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?
   1. #1
   2. #5
   3. #1 and #5

## #2, #3 and #5

1. 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 = 𝑥𝑐𝑜𝑟𝑟2(𝐴, 𝐴))
   1. It is not possible to apply the cross correlation to the same image
   2. Output O tends to be a flat plateau with one clear central peak

## Output O tends to have many peaks and one evident maximum

* 1. Output O tends to have many equivalent peaks with the same maximum value