Self-assessment questions 12

January 30, 2026

1. Suppose to perform anomaly detection via the K-means algorithm. In order to choose K, you apply the algorithm to the entire dataset (without dividing into training and test set), you then check the average silhouette score and choose the K that maximizes it. Is it correct to do so? Or must we apply the K-means algorithm only on the training set?
2. Write the formulas of the silhouette score and of the inertia (also called total within-cluster variation). What is the advantage of using one instead of the other?
3. Consider a dataset made of 4 samples: A=(1,1), B=(2,1), C=(2,2), D=(3,3). Represent the samples in an x-y plot and write the steps of the K-means algorithm. At each step, write and draw on the graph how the centroids change.
4. Write the formula of feature importance given by a random forest.
5. Suppose to use a random forest to perform regression. If you vary the number of trees, how do you expect the error on the test set to change? Draw a hypothetical curve that one could observe, with x = number of trees and y = error.
6. Write the formulas of entropy, Gini impurity index and information gain.
7. If you train a neural network on Keras, using different seeds, do you obtain always the same weights for the neurons? Why?
8. Why is it important to repeat the K-means algorithm with multiple initializations of the centroids? After all the repetitions, how do we get the final clusters?
9. What is the difference between feature selection techniques (refer to an example of technique we have seen) and dimensionality reduction?
10. Is dimensionality reduction a supervised or unsupervised technique?
11. To what can dimensionality reduction be useful?
12. How can we use dimensionality reduction for anomaly detection?

13. When we apply Principal Component Analysis (PCA), how do we choose the number of components to keep?

14. Write the formula of Singular Value Decomposition (SVD) of a data matrix $\mathbf{A}$. Write the properties of every element in the formula.

15. If you perform Singular Value Decomposition (SVD) on a data matrix and you draw a plot of the variance captured vs. the number of components, what is the shape that you obtain? Draw an example of such a plot.

16. Consider a data matrix $\mathbf{A}$ and apply Singular Value Decomposition (SVD). Denote with $\mathbf{A}^{(k)}$ the reduced matrix, obtained keeping only k components. Consider the Froebinius norm of the difference between $\mathbf{A}$ and $\mathbf{A}^{(k)}$, i.e., $||\mathbf{A} - \mathbf{A}^{(k)}||_F$. Is it equal to what?

17. Consider a dataset where you want to perform binary classification. Would you be able to generate such a dataset so that any classification model, even the most complex, has a very low precision (about 50%)? How?