**Quick Check-Point #4**

**#1**

The vast majority of available data in many applications is usually **labeled**.

Yes/No (The vast majority of available data in many applications is usually **labeled**. We know the input features X, but we don’t have the labels y to train our model.)

**#2**

**Unsupervised learning** is learning without a “teacher” **supervising** the learning process. The goal is to identify automatically meaningful patterns in unlabeled data. We don’t need to provide the algorithm a labeled dataset, which makes it a very useful option for some use cases.

Yes/No

**#3**

**Clustering** is the task of identifying similar instances with **shared attributes** in a dataset and groupthem into **clusters**.

Yes/No

**#4**

When using clustering, each cluster is assigned a number called a **cluster-ID,** and each data point or instance will be assigned to one cluster-ID. This is used to label the data with a small group of clusters automatically.

Yes/No

**#5**

**Semi-supervised learning** is a method sitting between supervised and unsupervised learning. The idea here is to run a clustering algorithm on an unlabeled dataset that will create few clusters as labels, clusters #1, #2, etc. Then propagate the labels to all the instances in the same cluster. And now, we have labeled dataset that can be used for training a model in supervised learning.

Yes/No

**#6**

One of the main challenges in supervised learning is the number of input features that the algorithm needs to analyze. **Dimension reduction** algorithms are used for reducing the number of features under consideration, helping to simplify the data without losing too much information.

Yes/No