2.

* 1. Overfitting occurs when a model learns the intricacies in the training data and learns the noise to the point where the model starts to interpret it as a pattern in the dataset.
  2. Underfitting occurs when a model does not correctly learn the dataset and cannot generalise the testing data.
  3. “An extremely simple hypothesis”, i.e., a weak data-target relation, would mean that the data cannot be generalised properly, leading to underfitting.
  4. “An overly complex hypothesis”, i.e., a strong data-target relation, would mean that the model would learn the data, and the noise that exists in the data would also be learnt by the model, leading to wrong results computed due to the noise, leading to overfitting.

3. Disadvantages of K-Means:

1. The number of clusters (k) has to be mentioned in advance.
2. Outliers in the dataset would be interpreted and considered by the model, resulting in inconsistencies.
3. Even if the training dataset contains clusters of different sizes, the output produced would contain clusters of uniform size.
4. Scaling of the training data may result in drastic change in output.

4. Let us consider the student dataset. In the student data, taking Name, Address, Gender, Marks, and Birthday exist. In this, we can take and plot the frequency of the student’s state to get an insight into how many students live in a particular state.