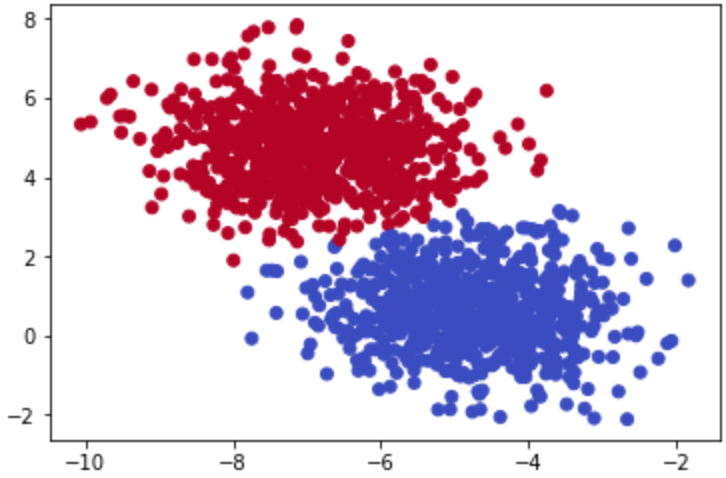
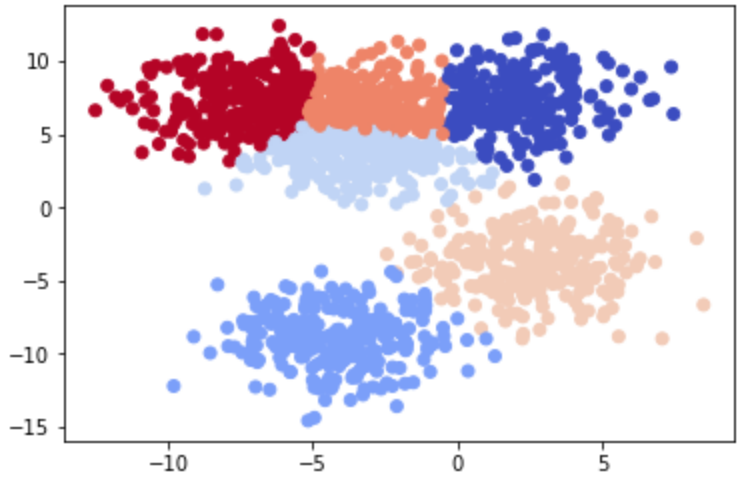
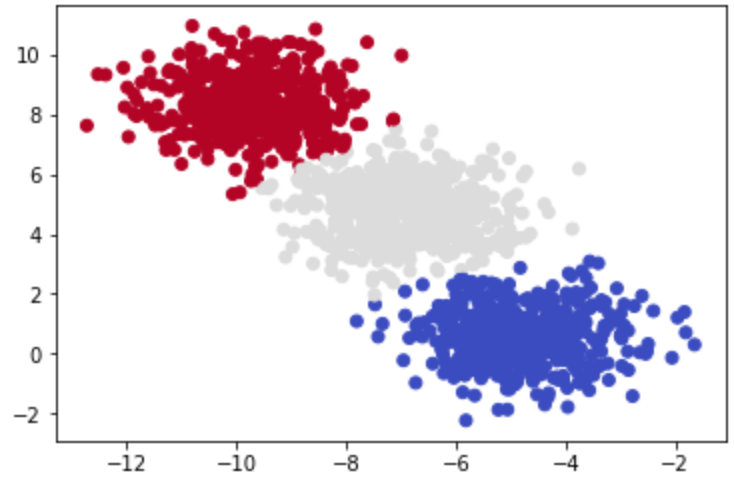
Chapter 1: Introduction to Unsupervised Learning Answer Key

Activity 1 Solutions:







Activity 2 Solutions:

Please refer to the provided Jupyter notebook for solution code.

Practice Question Solutions:

1) How are supervised and unsupervised learning different?

**Supervised has labels, Unsupervised does not**

Supervised means you have to monitor the process the whole time, Unsupervised does not

Supervised only works on smaller datasets, and you need to use Unsupervised with larger datasets

2) What is a cluster?

A group of models used to train on data

A collection of features that are important to your model

**A grouping of similar data**

3) Finding clusters in your data is always valuable

True

**False**

4) How are dimensions expressed in a dataset?

**As features, typically the number of columns in your data**

As features, typically the number of rows in your data

As features, typically the number of tables in your database

5) How many dimensions can be calculated by computers?

Less than 3

3

**Greater than 3**

6) What does the “K” represent in K-Means clustering?

How many models will be fit to your data, with performance results averaged

**How many clusters you expect to be in your data**

How many dimensions your data set has

7) The starting points for K-Means are determined by taking the mean of all the points in the space.

True

**False**

8) What is the formula that underpins K-Means clustering?

**Euclidean Distance**

Manhattan Distance

Cosine Distance

9) When does a K-Means clustering algorithm finish running?

After specified number of iterations

After K/2 iterations

**After convergence, when there is no longer a difference in calculated cluster centers**

10) K-Means will still find clusters in a dataset even if all of the data is fairly similar.

**True**

False