# Machine Learning Lab 5

## K-Means Clustering

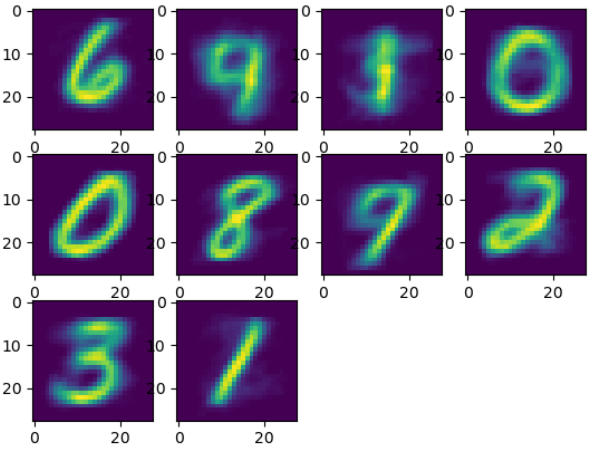
In this lab, you will use K-Means clustering techniques in classification tasks. As for the dataset, we still use the MNIST which we used in our previous labs. You should finish the following task.

1. Use `KMeans` from `sklearn.cluster` to do K-Means clustering.

   See the following as an example:

   ```
   from sklearn.cluster import KMeans
   kmeans = KMeans(n_clusters=10).fit(X_small)
   kmeans.cluster_centers_
   kmeans.labels_ = y_small
   kmeans.predit(X_test)
   ```

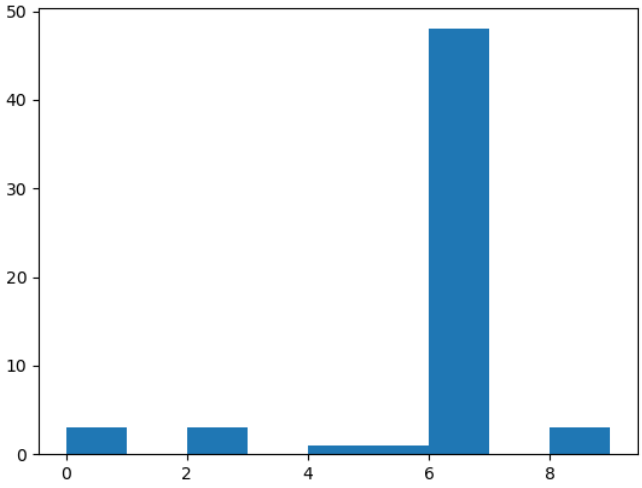
   You can use `cluster_centers_` from `KMeans` to check the center of those clusters by plotting each of the center as images, then calculate the distribution of the ten digits for each cluster to see how well your clustering performs.



   In the ideal case, each of your cluster should contain only images with the same digit. You may use `hist` from `matplotlib.pyplot` for plotting a histogram for distribution.

For example, if `y_cluster1` contains the labels for all the points in the first cluster, you can do as following:

```
from matplotlib import plot as plt
plt.hist(y_cluster1, [0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
```




In the given example in this lab script, the first cluster center seems like digit 6 where 6 appears most in its corresponding histogram as well.

Note that although there are ten digits in the dataset, you can try different Ks other than 10 to see whether you can cluster different digits with similar shape together.