Jay Menon

H041

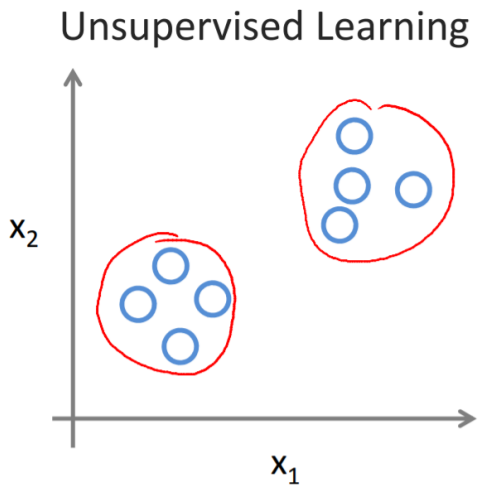
AI - LAB 5

AIM: To study and implement Unsupervised Learning: Clustering.

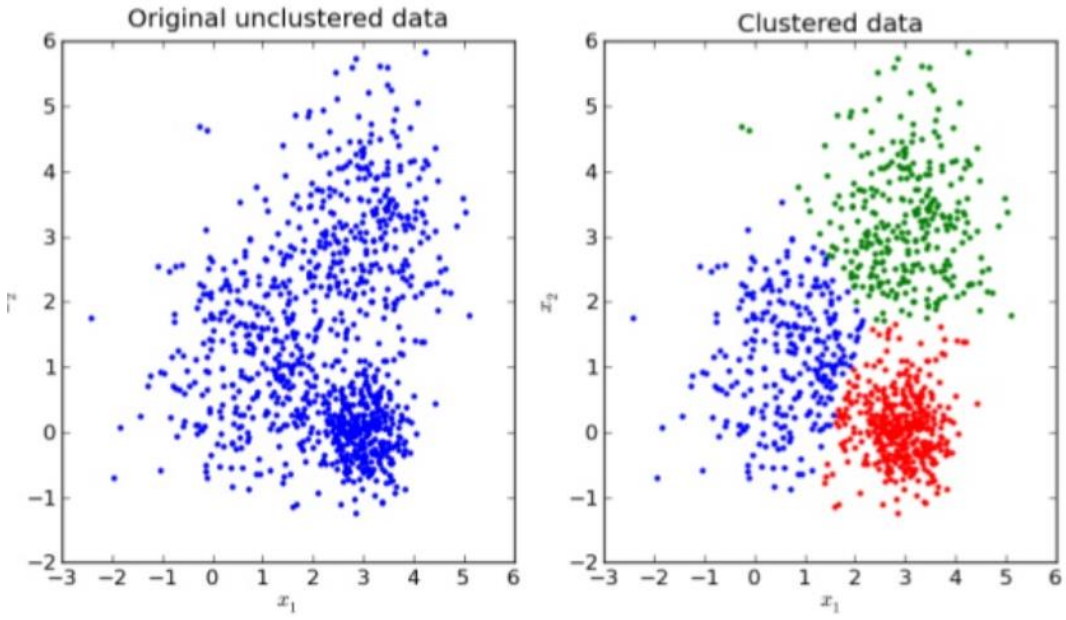
Theory: In Unsupervised Learning, the learners are required to form and evaluate the concepts on their own. The algorithms need to figure out the pattern in the loaded data and learn on the basis of generalisation.

Common scenarios for using unsupervised learning algorithms include:

- Data Exploration
- Outlier Detection
- Pattern Recognition



Such a technique is the Clustering. It helps in dividing the set of observations into small clusters based on their similarities. Data dissimilar to one cluster might be a part of another new cluster.

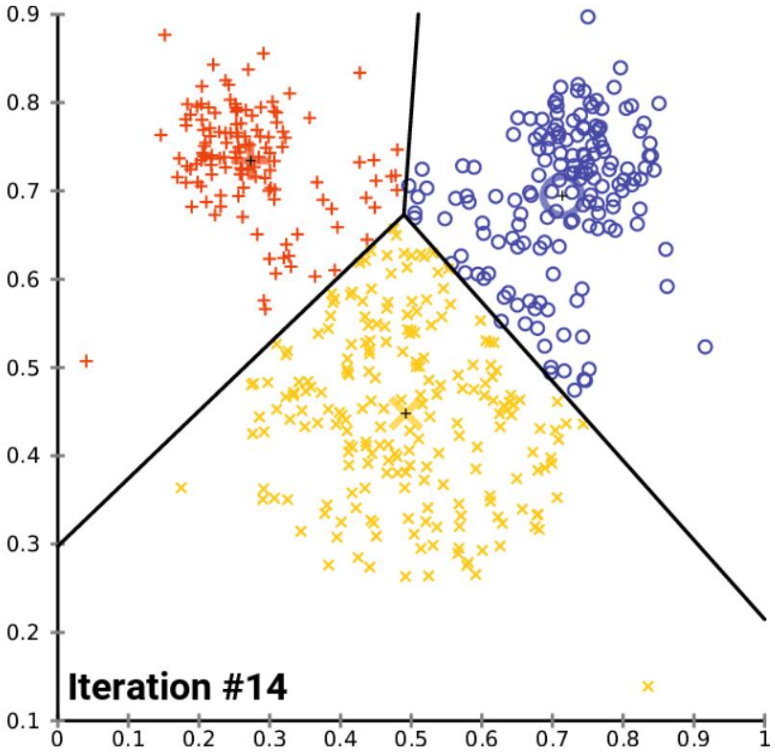


Algorithms for Data clustering are:

- K-Means Algorithm

  The algorithm then iteratively moves the k-centres and selects the datapoints that are closest to that centroid in the cluster.

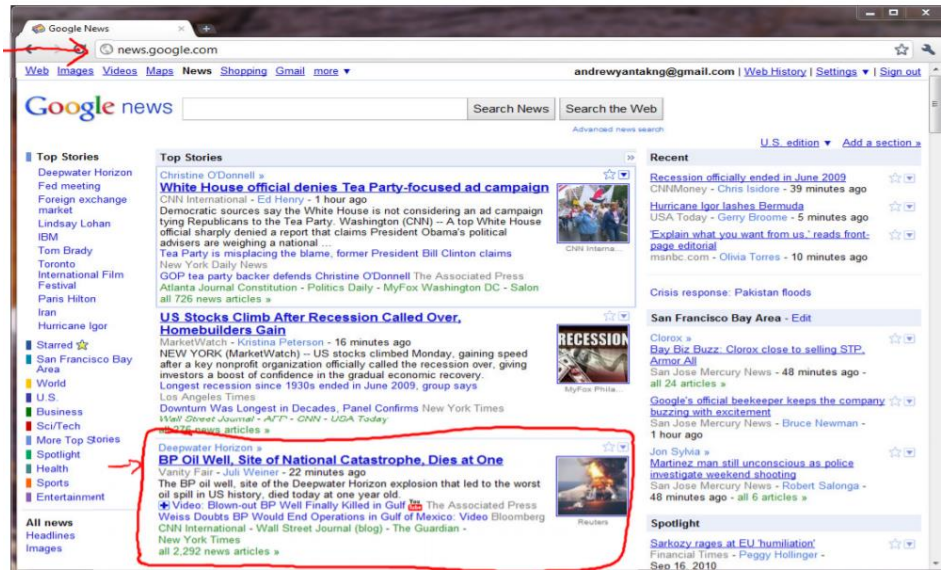
Taking K=3 as an example, the iterative process is given below:



- Mean-Shift Algorithm

It is a non-parametric algorithm. The central mean of the data keeps on shifting itself until it reaches the optimum point for the cluster.

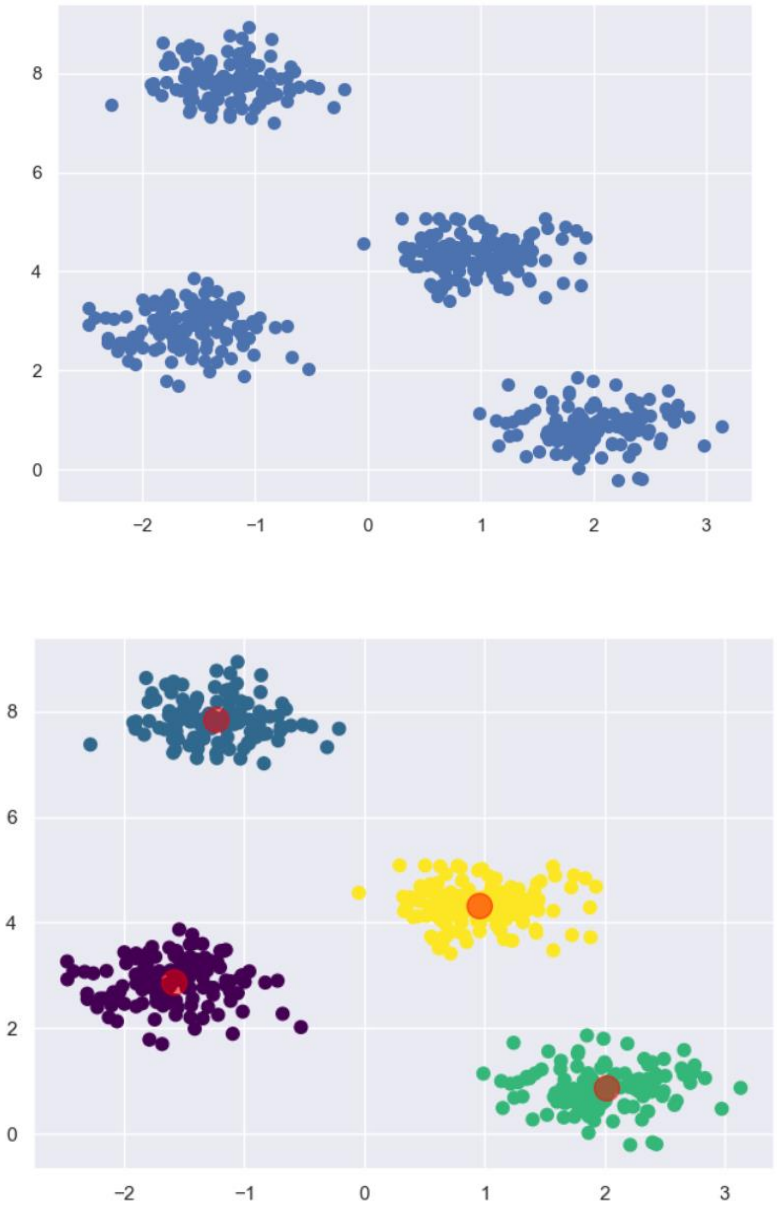
In the given Google webpage, the news articles are clustered on various parameters such as the user's location, recent search history, news priority and more.



Code:

```
import matplotlib.pyplot as plt
import seaborn as sns; sns.set()
import numpy as np
from sklearn.cluster import KMeans

from sklearn.datasets import make_blobs

X, y_true = make_blobs(n_samples = 500, centers = 4, cluster_std = 0.40,
random_state = 0)

plt.scatter(X[:, 0], X[:, 1], s = 50);
plt.show()

km = KMeans(n_clusters = 4)

km.fit(X)
y_kmeans = km.predict(X)
plt.scatter(X[:, 0], X[:, 1], c = y_kmeans, s = 50, cmap = 'viridis')

centers = km.cluster_centers_
plt.scatter(centers[:, 0], centers[:, 1], c = 'red', s = 200, alpha = 0.5);
plt.show()
```

Output:



Conclusion: In this experiment we have successfully implemented the K Means algorithm of Clustering from Unsupervised Learning. We saw how the clusters and their respective centroid correctly positioned themselves.

References:

- AI with Python - Unsupervised Learning: Clustering - Tutorialspoint
- Clustering Based Unsupervised Learning | by Syed Sadat Nazrul | Towards Data Science
- https://www.springboard.com/blog/lp-machine-learning-unsupervised-learning-supervised-learning/
- Machine Learning by Andrew Ng
- https://www.slideshare.net/annafensel/kmeans-clustering-122651195