

worksheet__07-Copy2

February 23, 2024

1 Worksheet 07

Name: Haya Naveed UID: U16758779

1.0.1 Topics

- Density-Based Clustering

1.0.2 Density-Based Clustering

Follow along with the live coding of the DBScan algorithm.

```
[1]: import numpy as np
import matplotlib.pyplot as plt
import sklearn.datasets as datasets

centers = [[1, 1], [-1, -1], [1, -1]]
X, _ = datasets.make_blobs(n_samples=750, centers=centers, cluster_std=0.4,
random_state=0)
plt.scatter(X[:, 0], X[:, 1], s=10, alpha=0.8)
plt.show()

class DBC():

    def __init__(self, dataset, min_pts, epsilon):
        self.dataset = dataset
        self.min_pts = min_pts
        self.epsilon = epsilon
        self.assignments = [-1 for _ in range(len(self.dataset))]

    def distance(self, i, j):
        return np.linalg.norm(self.dataset[i] - self.dataset[j])

    def is_core(self, i):
        return len(self.get_nbhd(i)) >= self.min_pts

    def get_nbhd(self, i):
        nbhd = []
        for j in range(len(self.dataset)):
```

```

        if i != j and self.distance(i, j) <= self.epsilon:
            nbhd.append(j)
        return nbhd

def is_unassigned(self, i):
    return self.assignments[i]==-1

def get_unassigned_nbhd(self, i):
    nbhd = self.get_nbhd(i)
    return [pt for pt in nbhd if self.is_unassigned(pt)]

def make_cluster(self, i, cluster_num):
    self.assignments[i] = cluster_num
    nbhd_q = self.get_unassigned_nbhd(i)

    while nbhd_q:
        next_pt = nbhd_q.pop()
        if not self.is_unassigned(next_pt):
            continue
        self.assignments[next_pt] = cluster_num
        #self.snap()
        if self.is_core(next_pt):
            nbhd_q += self.get_unassigned_nbhd(next_pt)

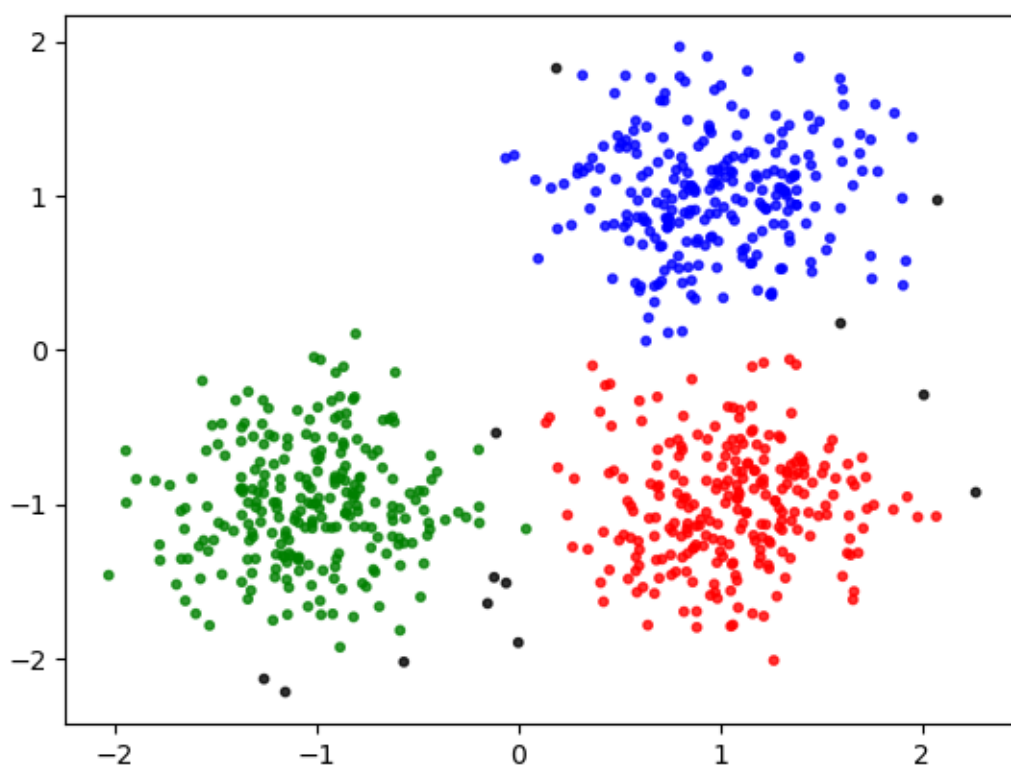
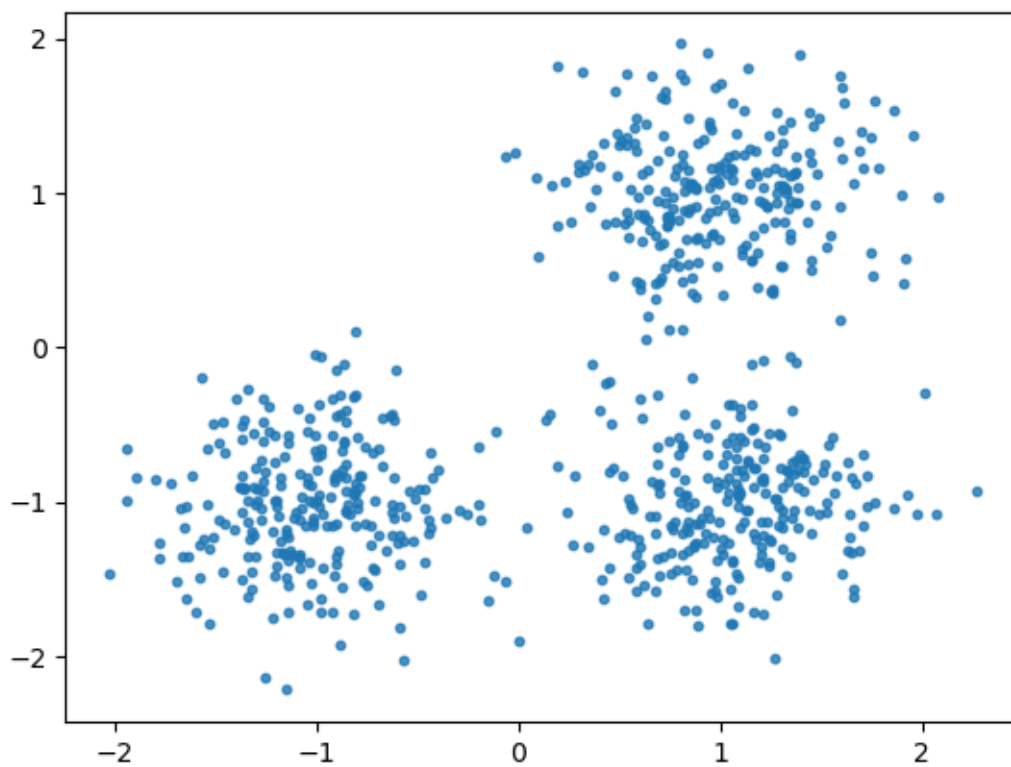
def dbscan(self):
    cluster_num = 0

    for i in range(len(self.dataset)):
        if self.assignments[i] != -1:
            continue
        if self.is_core(i):
            self.make_cluster(i, cluster_num)
            cluster_num += 1

    return self.assignments

clustering = DBC(X, 5, 0.3).dbscan()
colors = np.array([x for x in 'bgrcmykbgrcmykbgrcmykbgrcmyk'])
colors = np.hstack([colors] * 100)
plt.scatter(X[:, 0], X[:, 1], color=colors[clustering].tolist(), s=10, alpha=0.
↪8)
plt.show()

```



[5]: *# Challenge part 1 - animation of DBScan algorithm, result is in 'dbscan2.gif'*

```
import numpy as np
import matplotlib.pyplot as plt
import sklearn.datasets as datasets
from matplotlib.patches import Circle
from PIL import Image as im

centers = [[1, 1], [-1, -1], [1, -1]]
X, _ = datasets.make_blobs(n_samples=750, centers=centers, cluster_std=0.4,
    ↪random_state=0)
plt.scatter(X[:, 0], X[:, 1], s=10, alpha=0.8)
plt.show()
TEMPFILE = 'temp.png'

class DBC():

    def __init__(self, dataset, min_pts, epsilon):
        self.dataset = dataset
        self.min_pts = min_pts
        self.epsilon = epsilon
        self.assignments = [-1 for _ in range(len(self.dataset))]
        self.snaps = []

    def distance(self, i, j):
        return np.linalg.norm(self.dataset[i] - self.dataset[j])

    def is_core(self, i):
        return len(self.get_nbhd(i)) >= self.min_pts

    def get_nbhd(self, i):
        nbhd = []
        for j in range(len(self.dataset)):
            if i != j and self.distance(i, j) <= self.epsilon:
                nbhd.append(j)
        return nbhd

    def is_unassigned(self, i):
        return self.assignments[i] == -1

    def get_unassigned_nbhd(self, i):
        nbhd = self.get_nbhd(i)
        return [pt for pt in nbhd if self.is_unassigned(pt)]
```

```

def make_cluster(self, i, cluster_num):
    self.assignments[i] = cluster_num
    nbhd_q = self.get_unassigned_nbhd(i)

    while nbhd_q:
        next_pt = nbhd_q.pop()
        if not self.is_unassigned(next_pt):
            continue
        self.assignments[next_pt] = cluster_num
        print("Taking a snapshot...")
        self.snapshot(next_pt) # Take a snapshot when a point is assigned
        if self.is_core(next_pt):
            nbhd_q += self.get_unassigned_nbhd(next_pt)

def snapshot(self, assigned_point):
    fig, ax = plt.subplots()
    colors = np.array([x for x in 'bgrcmkykbgrcmkykbgrcmkykbgrcmky'])
    colors = np.hstack([colors] * 100)

    ax.scatter(self.dataset[:, 0], self.dataset[:, 1], color=colors[self.
↪ assignments].tolist(), s=10, alpha=0.8)

    for i, assignment in enumerate(self.assignments):
        if assignment != -1 and i == assigned_point:
            cir = Circle((self.dataset[i, 0], self.dataset[i, 1]), self.
↪ epsilon, fill=False, color='black')
            ax.add_patch(cir)

    ax.set_xlim(np.min(self.dataset[:, 0]) - 1, np.max(self.dataset[:, 0]))
↪ + 1)
    ax.set_ylim(np.min(self.dataset[:, 1]) - 1, np.max(self.dataset[:, 1]))
↪ + 1)
    ax.set_aspect('equal') # necessary or else the circles appear to be
↪ oval shaped

    temp_file = 'temp1.png'
    fig.savefig(temp_file)
    plt.close()

    with im.open(temp_file) as img:
        img_copy = img.copy()
        self.snaps.append(img_copy)

def dbscan(self):
    cluster_num = 0

```

```

        for i in range(len(self.dataset)):
            if self.assignments[i] != -1:
                continue
            if self.is_core(i):
                self.make_cluster(i, cluster_num)
                cluster_num += 1

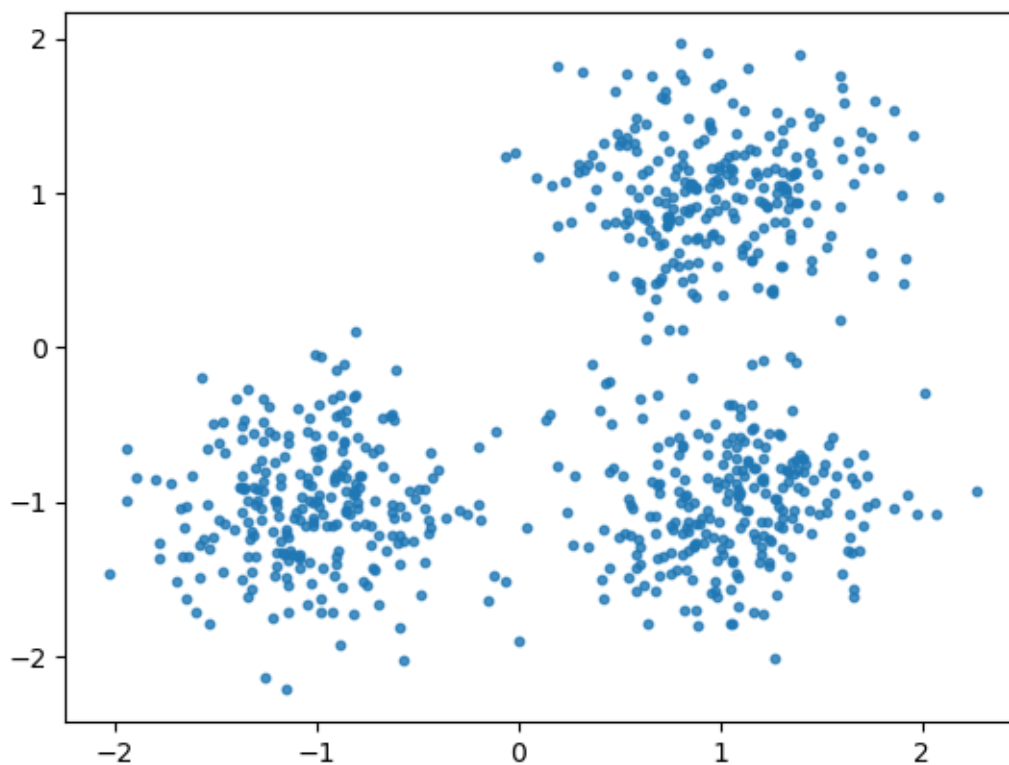
        return self.assignments

dbc = DBC(X, 5, 0.3)
clustering = dbc.dbscan()

dbc.snaps[0].save(
    'dbscan2.gif',
    optimize=False,
    save_all=True,
    append_images=dbc.snaps[1:],
    loop=0,
    duration=25
)

#####

```



Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

1.1 Challenge Problem

Using the code above and the template provided below, create the animation below of the DBScan algorithm.

```
[7]: from IPython.display import Image
Image(filename="dbscan_2.gif", width=500, height=500)
```

```
[7]: <IPython.core.display.Image object>
```

Hints:

- First animate the dbscan algorithm for the dataset used in class (before trying to create the above dataset)
- Take a snapshot of the assignments when the point gets assigned to a cluster
- Confirm that the snapshot works by saving it to a file
- Don't forget to close the matplotlib plot after saving the figure
- Gather the snapshots in a list of images that you can then save as a gif using the code below
- Use `ax.set_aspect('equal')` so that the circles don't appear to be oval shaped
- To create the above dataset you need two blobs for the eyes. For the mouth you can use the following process to generate (x, y) pairs:
 - Pick an x at random in an interval that makes sense given where the eyes are positioned
 - For that x generate y that is $0.2 * x^2$ plus a small amount of randomness
 - zip the x's and y's together and append them to the dataset containing the blobs

```
[20]: # Challenge part 2 - animation of Smiley face, result is in 'dbscan_smile4.gif'
```

```
import numpy as np
import matplotlib.pyplot as plt
import sklearn.datasets as datasets
from matplotlib.patches import Circle
from PIL import Image as im

# centers = [[1, 1], [-1, -1], [1, -1]]
# X, _ = datasets.make_blobs(n_samples=750, centers=centers, cluster_std=0.4,
# random_state=0)
# plt.scatter(X[:, 0], X[:, 1], s=10, alpha=0.8)
# plt.show()
# TEMPFILE = 'temp.png'

class DBC():

    def __init__(self, dataset, min_pts, epsilon):
        self.dataset = dataset
        self.min_pts = min_pts
        self.epsilon = epsilon
        self.assignments = [-1 for _ in range(len(self.dataset))]
```

```

self.snaps = []

def distance(self, i, j):
    return np.linalg.norm(self.dataset[i] - self.dataset[j])

def is_core(self, i):
    return len(self.get_nbhd(i)) >= self.min_pts

def get_nbhd(self, i):
    nbhd = []
    for j in range(len(self.dataset)):
        if i != j and self.distance(i, j) <= self.epsilon:
            nbhd.append(j)
    return nbhd

def is_unassigned(self, i):
    return self.assignments[i] == -1

def get_unassigned_nbhd(self, i):
    nbhd = self.get_nbhd(i)
    return [pt for pt in nbhd if self.is_unassigned(pt)]

def make_cluster(self, i, cluster_num):
    self.assignments[i] = cluster_num
    nbhd_q = self.get_unassigned_nbhd(i)

    while nbhd_q:
        next_pt = nbhd_q.pop()
        if not self.is_unassigned(next_pt):
            continue
        self.assignments[next_pt] = cluster_num
        print("Taking a snapshot...")
        self.snapshot(next_pt) # Take a snapshot when a point is assigned
        if self.is_core(next_pt):
            nbhd_q += self.get_unassigned_nbhd(next_pt)

def snapshot(self, assigned_point):
    fig, ax = plt.subplots()
    colors = np.array([x for x in 'bgrcmykbgrcmykbgrcmykbgrcmyk'])
    colors = np.hstack([colors] * 100)

    ax.scatter(self.dataset[:, 0], self.dataset[:, 1], color=colors[self.
↪assignments].tolist(), s=10, alpha=0.8)

    for i, assignment in enumerate(self.assignments):
        if assignment != -1 and i == assigned_point:

```



```

        cir = Circle((self.dataset[i, 0], self.dataset[i, 1]), self.
↪epsilon, fill=False, color='black')
        ax.add_patch(cir)

        ax.set_xlim(np.min(self.dataset[:, 0]) - 1, np.max(self.dataset[:, 0]))
↪+ 1)
        ax.set_ylim(np.min(self.dataset[:, 1]) - 1, np.max(self.dataset[:, 1]))
↪+ 1)
        ax.set_aspect('equal') # necessary or else the circles appear to be
↪oval shaped

        temp_file = 'temp_smile3.png'
        fig.savefig(temp_file)
        plt.close()

        with im.open(temp_file) as img:
            img_copy = img.copy()
            self.snaps.append(img_copy)

    def dbscan(self):
        cluster_num = 0

        for i in range(len(self.dataset)):
            if self.assignments[i] != -1:
                continue
            if self.is_core(i):
                self.make_cluster(i, cluster_num)
                cluster_num += 1

        return self.assignments

#centers = [[-1.5, 3], [1.5, 3]]
# centers = [[-1, 2], [1, 2]]
# eyes, _ = datasets.make_blobs(n_samples=500, centers=centers, cluster_std=0.
↪4, random_state=0)
centers = [[-1, 2], [1, 2]]
eyes, _ = datasets.make_blobs(n_samples=500, centers=centers, cluster_std=0.2,
↪random_state=0)
# eyes, _ = datasets.make_blobs(n_samples=750, centers=centers, cluster_std=0.
↪4, random_state=0)

# mouth_x = 6 * np.random.random(150) - 3 # Random x values in the range [-3,
↪3]

```

```

# mouth_y = 0.2 * mouth_x**2 + 0.1 * np.random.randn(150) # Generate y values
↳for the smiley face mouth
# mouth_x = 5 * np.random.random(150) - 2.5 # Random x values in the range [-2.
↳5, 2.5]
# mouth_y = 0.07 * mouth_x**2 + 0.1 * np.random.randn(150) # Generate y values
↳for the smiley face mouth
mouth_x = 4 * np.random.random(250) - 2 # Random x values in the range [-2, 2]
mouth_y = 0.2 * mouth_x**2 + 0.1 * np.random.randn(250) # Generate y values
↳for the smiley face mouth

face = np.append(eyes, np.column_stack((mouth_x, mouth_y)), axis=0)

# Run DBSCAN on smiley face dataset
dbc = DBC(face, 5, 0.3)
clustering = dbc.dbscan()

dbc.snaps[0].save(
    'dbscan_smile4.gif',
    optimize=False,
    save_all=True,
    append_images=dbc.snaps[1:],
    loop=0,
    duration=25
)

```

Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...
 Taking a snapshot...

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...
Taking a snapshot...

```
[ ]: # import numpy as np
# from PIL import Image as im
# import matplotlib.pyplot as plt
# import sklearn.datasets as datasets

# TEMPFILE = 'temp.png'

# class DBC():

#     def __init__(self, dataset, min_pts, epsilon):
#         self.dataset = dataset
#         self.min_pts = min_pts
#         self.epsilon = epsilon
#         self.snaps = []

#     def snapshot(self):
#         fig, ax = plt.subplots()
#         colors = ...

#         ax.scatter(...)
#         cir = plt.Circle(...) # create circle around the point assigned
#         ax.add_patch(cir)
#         ax.set_xlim(...)
#         ax.set_ylim(...)
#         ax.set_aspect('equal') # necessary or else the circles appear to be
#         ↪ oval shaped

#         fig.savefig(TEMPFILE)
#         plt.close()

#         return im.fromarray(np.asarray(im.open(TEMPFILE)))

#     def dbscan(self):
#         ...
#         return
```

```

# centers = [[-2, 2], [2, 2]]
# eyes, _ = datasets.make_blobs(n_samples=750, centers=centers, cluster_std=0.
    ↪4, random_state=0)

# mouth_x = 6 * np.random.random(150) - 3 # Random x values in the range [-3,
    ↪3]
# mouth_y = 0.2 * mouth_x**2 + 0.1 * np.random.randn(150) # Generate y values
    ↪for the smiley face mouth

# face = np.append(eyes, np.column_stack((mouth_x, mouth_y)), axis=0)

# # Run DBSCAN on smiley face dataset
# dbc = DBC(face, 5, 0.3)
# clustering = dbc.dbscan()

# dbc.snaps[0].save(
#     'dbscan_smile.gif',
#     optimize=False,
#     save_all=True,
#     append_images=dbc.snaps[1:],
#     loop=0,
#     duration=25
# )

```