1. $d(A,B) = |1-1| + |4-3| = 1$

$d(A,C) = |1-3| + |4-4| = 2$

$d(A,D) = |1-5| + |4-2| = 6$

$d(A,E) = |1-3| + |4-2| = 4$

$d(A,F) = |1-3| + |4-0| = 6$

$d(B,C) = |1-3| + |3-4| = 3$

$d(B,D) = |1-5| + |3-2| = 5$

$d(B,E) = |1-3| + |3-2| = 3$

$d(B,F) = |1-3| + |3-0| = 5$

$d(C,D) = |3-5| + |4-2| = 4$

$d(C,E) = |3-3| + |4-2| = 2$

$d(C,F) = |3-3| + |4-0| = 4$

$d(D,E) = |5-3| + |2-2| = 2$

$d(D,F) = |5-3| + |2-0| = 4$

$d(E,F) = |3-3| + |2-0| = 2$

So the matrix D, of pairwise distances will be

	A	B	C	D	E	F
A	0	1	2	6	4	6
B	1	0	3	5	3	5
C	2	3	0	4	2	4
D	6	5	4	0	2	4
E	4	3	2	2	0	2
F	6	5	4	4	2	0

In this table, $D_1(A,B) = 1$ is the lowest value of D_1, so I cluster elements, A and B.

First Branch

Let u denote the node to which A and B are connected.

setting $\delta(A,u) = \delta(B,u) = D_1(A,B)/2 = 1/2 = 0.5$

First distance matrix update

$D_2((A,B),C) = \min(D_1(A,C), D_1(B,C)) = \min(2,3) = 2$

$D_2((A,B),D) = \min(D_1(A,D), D_1(B,D)) = \min(6,5) = 5$

$D_2((A,B),E) = \min(D_1(A,E), D_1(B,E)) = \min(4,3) = 3$

$D_2((A,B),F) = \min(D_1(A,F), D_1(B,F)) = \min(6,5) = 5$

Second clustering

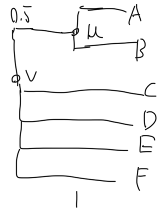
	(A,B)	C	D	E	F
(A,B)	0	2	5	3	5
C	2	0	4	2	4
D	5	4	0	2	4
E	3	2	2	0	2
F	5	4	4	2	0

Here $D_2((A,B),C) = 2 = D_2(C,E) = D_2(D,E) = D_2(E,F)$

Let v denote the node (A,B), C, D, E, F links

Here $\delta(A,v) = \delta(B,v) = \delta(C,v) = 2/2 = 1$

$\delta(u,v) = \delta(A,v) - \delta(A,u) = 0.5$



2. (a) $Z = (1-\varepsilon)e^{-a} + \varepsilon e^{a}$

$$\frac{dZ}{da} = -(1-\varepsilon)e^{-a} + \varepsilon e^{a}$$

let the derivative form $= 0$

$$-(1-\varepsilon)e^{-a} + \varepsilon e^{a} = 0$$

$$\varepsilon e^{a} = (1-\varepsilon)e^{-a}$$

$$a = \frac{\ln(1-\varepsilon) - \ln(\varepsilon)}{2}$$

(b) First consider the case $y = -1$

then I want to prove: $\mathbb{1}(H(x) \neq -1) \le \exp(f(x))$

$H(x) = \text{sign}(f(x))$

① If $f(x) \ge 0$ then $\mathbb{1}(H(x) \neq -1) = 1$

$\exp(f(x)) \ge 1$

then $1 \le \exp(f(x))$

$\mathbb{1}(H(x) \neq -1) \le \exp(f(x))$

② If $f(x) < 0$, Then $\mathbb{1}(H(x) \neq -1) = 0$

$0 < \exp(f(x)) < 1$

then $0 < \exp(f(x))$

$\mathbb{1}(H(x) \neq -1) \le \exp(f(x))$

Second consider the case $y = 1$

then I want to prove $\mathbb{1}(H(x) \neq 1) \le \exp(-f(x))$

① If $f(x) > 0$, then $\mathbb{1}(H(x) \neq 1) = 0$

$0 < \exp(-f(x)) < 1$

then $\mathbb{1}(H(x) \neq 1) < \exp(-f(x))$

② If $f(x) \le 0$, then $\mathbb{1}(H(x) \neq 1) = 1$,

$\exp(-f(x)) \ge 1$

$\mathbb{1}(H(x) \neq 1) \le \exp(-f(x))$

Thus, I proved that, for all values of x and either $y = -1$ or $y = 1$,

$$\mathbb{1}(H(x) \neq y) \le \exp(-yf(x)) = \exp\left(-y\sum_{t=1}^{T} a_t h_t(x)\right)$$

(c) $\prod_{t=1}^{t} Z_t = Z_1 \cdot Z_2 \cdot Z_3 \cdots Z_t$

$Z_1 = \sum_{i=1}^{m} D_1(x_i)\exp(-a_1 y_i h_1(x_i))$ $\quad D_1(x_i) = \frac{1}{m}$

$Z_2 = \sum_{i=1}^{m} D_2(x_i)\exp(-a_2 y_i h_2(x_i))$ $\quad D_2(x_i) = \frac{D_1(x_i)\cdot\exp(-a_1 y_i h_1(x_i))}{Z_1}$

$Z_3 = \sum_{i=1}^{m} D_3(x_i)\exp(-a_3 y_i h_3(x_i))$ $\quad D_3(x_i) = \frac{D_2(x_i)\cdot\exp(-a_2 y_i h_2(x_i))}{Z_2}$

taking D_t out of the summation.

$$Z_1 = D_1 \sum_{i=1}^{m} \exp(-y_i a_1 h_1(x_i))$$

$$Z_2 = \sum_{i=1}^{m} \frac{D_1(i)\exp(-a_1 y_i h_1(x_i))}{Z_1}\cdot\exp(-a_2 y_i h_2(x_i))$$

$$= \frac{D_1}{Z_1}\sum_{i=1}^{m} \exp(-y_i a_1 h_1(x_i) - y_i a_2 h_2(x_i))$$

$$Z_3 = \sum_{i=1}^{m} \frac{D_2(i)\exp(-a_2 y_i h_2(x_i))}{Z_2}\cdot\exp(-a_3 y_i h_3(x_i))$$

$$= \sum^{m} D_1(i)\exp(-a_1 y_i h_1(x_i)) \cdot \exp(-a_2 y_i h_2(x_i))$$

$$\sum_{i=1} \quad \underbrace{Z_1 \qquad\qquad 1}_{Z_2} \cdot \exp(-a_3 y_i h_3(x_i)$$

$$= \sum_{i=1}^{m} \frac{D_1(i)}{Z_1 \cdot Z_2} \cdot \exp(-y_i a_1 h_1(x_i) - y_i a_2 h_2(x_i) - y_i a_3 h_3(x_i))$$

$$= \sum_{i=1}^{m} \frac{D_1}{Z_1 \cdot Z_2} \cdot \exp(-y_i \sum_{t=1}^{3} a_t h_t(x_i))$$

From here, I can see that

$$Z_T = \sum_{i=1}^{m} \frac{D_1}{Z_1 \cdot Z_2 \cdots Z_{T-1}} \cdot \exp(-y_i \sum_{t=1}^{T} a_t h_t(x_i))$$

$$\prod_{t=1}^{T} Z_t = Z_1 \cdot Z_2 \cdots Z_T$$
$$= \sum_{i=1}^{m} D_1 \cdot \exp(-y_i \sum_{t=1}^{T} a_t h_t(x_i))$$

$$D_1 = \frac{1}{m}$$
$$= \sum_{i=1}^{m} \frac{1}{m} \cdot \exp(-y_i \sum_{t=1}^{T} a_t h_t(x_i))$$

$\frac{1}{m}$ is constant, I can move it out side of $\sum_{i=1}^{m}$.

thus $\prod_{t=1}^{T} Z_t = \frac{1}{m} \sum_{i=1}^{m} \exp(-y_i \sum_{t=1}^{T} a_t h_t(x_i))$

$$\varepsilon_T = \sum_{i=1}^{m} D_T(x_i) \mathbb{1}(h_T(x_i) \neq y_i)$$

$$\mathbb{1}(h_T(x_i) \neq y_i) = 1 \Rightarrow h_T(x_i) y_i < 0 \Rightarrow \exp(-\alpha_T y_i h_T(x_i)) > 1$$

$$\mathbb{1}(h_T(x_i) \neq y_i) = 0 \Rightarrow h_T(x_i) y_i \geq 0 \Rightarrow \exp(-\alpha_T y_i h_T(x_i)) > 0$$

$\Rightarrow \mathbb{1}(h_T(x_i) \neq y_i)$ is upperbounded by $\exp(-\alpha_T y_i h_T(x_i))$

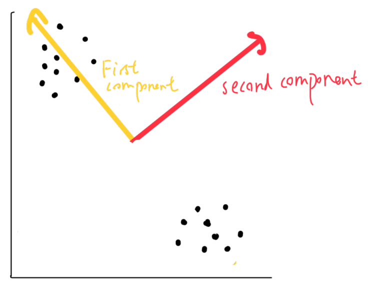
So $\varepsilon_T \leq \sum_{i=1}^{m} D_T(x_i) \exp(-\alpha_T y_i h_T(x_i)) = Z_T$

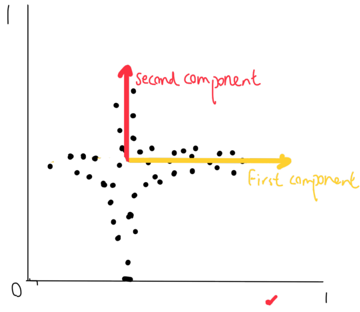
from part (a) I know that

$a = \frac{1}{2} \ln(\frac{1-\varepsilon}{\varepsilon})$ which minimize Z.

therefore it also minimize the upper bound Lost

3.(a)





(b) eigenvalue (Σ)

$$= \text{eigenvalue}\left(\begin{bmatrix} 6 & -3 & 0 \\ -3 & 6 & 0 \\ 0 & 0 & 10 \end{bmatrix} \right) = [3, 9, 10]$$

So $\sigma_1 = 10$

$\sigma_2 = 9$

$\sigma_3 = 3$

the ordered basis vector

$$[u_1, u_2, u_3] \in \mathbb{R}^3 = \begin{bmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \end{bmatrix}$$

4. (a) $L(\mu, \sigma^2; x, z) = \prod_{i=1}^{n} p(x_i \mid z_i, \mu_k) = \sum_{i=1}^{n} \sum_{k=1}^{3} \mathbb{I}[z_i = k] \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{1}{2\sigma^2}(x_i - \mu_k)^2\right)$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \mathbb{I}[z_i = k] (x_i - \mu_k)^2\right) = ($$

$$\mathbb{I}[z_i = k] = 1, \ \exp\left(\frac{1}{2\sigma^2}(x_i - \mu_k)^2\right)$$

(b) $p(\mu_1, \mu_2, \mu_3 \mid X, Z, \gamma^2, \sigma^2)$

$$= \prod_{k=1}^{3} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\gamma^2}\mu_k^2\right) \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \mathbb{I}[z_i = k](x_i - \mu_k)^2\right)$$

$$\propto \prod_{k=1}^{3} \prod_{i=1}^{n} \exp\left(-\frac{\mu_k^2}{2\gamma^2} - \frac{1}{2\sigma^2} \mathbb{I}[z_i = k](x_i - \mu_k)^2\right)$$

$$= \prod_{k=1}^{3} \exp\left(-\frac{\mu_k}{2\gamma^2} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \mathbb{I}[z_i = k](x_i - \mu_k)^2\right)$$

$$\propto \prod_{k=1}^{3} \prod_{i=1}^{n} \exp\left(-\frac{\mu_k^2}{2\gamma^2} - \frac{\mathbb{I}[z_i = k] x_i^2}{2\sigma^2} + \frac{\mathbb{I}[z_i = k] x_i \mu_k}{\sigma^2} - \frac{\mathbb{I}[z_i = k]\mu_k^2}{2\sigma^2}\right)$$

$$\propto \prod_{k=1}^{3} \prod_{i=1}^{n} \exp\left(-\frac{1}{2}\underbrace{\left(\frac{1}{\gamma^2} + \frac{\mathbb{I}[z_i = k]}{\sigma^2}\right)}_{a_1}\mu_k^2 + \underbrace{\frac{\mathbb{I}[z_i = k] x_i}{\sigma^2}}_{a_2}\mu_k\right)$$

$$\propto \prod_{k=1}^{3} \prod_{i=1}^{n} \exp\left(-\frac{1}{2}a_1\left(\mu_k^2 - \frac{2a_2}{a_1}\mu_k + \frac{a_2^2}{a_1^2}\right)\right)$$

$$= \prod_{k=1}^{3} \prod_{i=1}^{n} \exp\left(-\frac{1}{2}a_1\left(\mu_k - \frac{a_2}{a_1}\right)^2\right)$$

$$= \prod_{k=1}^{3} \exp\left(-\frac{1}{2\sigma_k^2}(\mu_k - \bar{\mu}_k)^2\right), \quad \sigma_k^2 = \left(\frac{1}{\gamma^2} + \frac{\sum \mathbb{I}(z_i = k)}{\sigma^2}\right)^{-1}$$

$$\propto \prod_{i=1}^{n} \exp\left(-\frac{\mu_1^2}{2\gamma^2} - \frac{1}{2\sigma^2}\mathbb{I}[z_i = 1](x_i - \mu_1)^2\right) \times \cdots \times \prod_{i=1}^{n} \exp\left(\frac{-\mu_3^2}{2\gamma^2} \cdots (x_i - \mu_3)^2\right)$$

$$= p(\mu_1 \mid X, Z, \gamma^2, \sigma^2) \times \cdots \times p(\mu_3 \mid X, Z, \gamma^2, \sigma^2)$$

$\Downarrow$

μ_1, μ_2, μ_3 are conditionally independent

(c) $p(Z \mid \mu, X, \gamma^2, \sigma^2) \propto \prod_{k=1}^{3} \prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2} \mathbb{I}[z_i = k](x_i - \mu_k)^2\right)$

$$\propto \prod_{k=1}^{3} \prod_{i=1}^{n} \left[\exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma^2}\right)\right]^{\mathbb{I}[z_i = k]}$$

$$= \prod_{i=1}^{n} \exp\left(-\frac{(x_i - \mu_1)^2}{2\sigma^2}\right)^{\mathbb{I}[z_i = 1]} \cdot \exp\left(-\frac{x_i - \mu_2}{2\sigma^2}\right)^{\mathbb{I}[z_i = 2]} \cdots$$

Thus Z is not independent given μ_1, μ_2, μ_3

In [1]:
```
import numpy as np;
import pandas as pd;
import matplotlib.pyplot as plt;
import seaborn as sns;
```

In [2]:
```
from sklearn.mixture import GaussianMixture
from sklearn.cluster import KMeans
```

In [3]:
```
from scipy.stats import norm
```

In [4]:
```
import statistics
```

In [5]:
```
df = pd.read_csv("hw3_Q5.txt", sep = " ", header = None, names = ["D1","D2","D3","D4","
data = pd.read_csv("hw3_Q5.txt", sep = " ", header = None, names = ["D1","D2","D3","D4"
```

a

In [6]:
```
gm = GaussianMixture(n_components = 3, covariance_type='spherical', init_params = 'rand
```

In [7]:
```
gm.fit(df)
```

Out[7]:
```
GaussianMixture(covariance_type='spherical', init_params='random',
                n_components=3)
```

In [8]:
```
gm.means_
```

Out[8]:
```
array([[ 1.05250251e-02,  4.16618640e-03,  1.35683863e-02,
         1.00207924e-02, -2.47443547e-02],
       [-3.06176428e+00, -3.96696240e+00, -5.02048769e+00,
        -5.02021252e+00, -6.00420335e+00],
       [ 2.99520441e+00,  3.97929290e+00,  4.93552594e+00,
         4.95000613e+00,  6.02926365e+00]])
```

In [9]:
```
u1 = gm.means_[0]
print('The means u1 is', u1)
```

```
The means u1 is [ 0.01052503  0.00416619  0.01356839  0.01002079 -0.02474435]
```

In [10]:
```
u2 = gm.means_[1]
print('The means u2 is', u2)
```

```
The means u2 is [-3.06176428 -3.9669624  -5.02048769 -5.02021252 -6.00420335]
```

In [11]:
```
u3 = gm.means_[2]
print('The means u3 is', u3)
```

```
The means u3 is [2.99520441 3.9792929  4.93552594 4.95000613 6.02926365]
```

In [12]:
```python
v = np.sqrt(gm.covariances_)
print('The variances gamma1 is', v[0])
```

```
The variances gamma1 is 0.4958449228619954
```

In [13]:
```python
print('The variances gamma2 is', v[1])
```

```
The variances gamma2 is 0.9848634682444027
```

In [14]:
```python
print('The variances gamma3 is', v[2])
```

```
The variances gamma3 is 0.9892007698529822
```

b

In [15]:
```python
km_per = KMeans(n_clusters = 3)
```

In [16]:
```python
km_per.fit(df)
```

Out[16]:
```
KMeans(n_clusters=3)
```

In [17]:
```python
km_per.n_iter_
```

Out[17]:
```
2
```

In [18]:
```python
km_rand = KMeans(n_clusters = 3, init = "random")
```

In [19]:
```python
km_rand.fit(df)
```

Out[19]:
```
KMeans(init='random', n_clusters=3)
```

In [20]:
```python
km_rand.n_iter_
```

Out[20]:
```
3
```

Apparently, random initialization and pre initialization from K-Means take different times of iterations to converge

c

In [21]:
```python
label = gm.fit_predict(df)
```

In [22]:
```python
df["predicted_cluster"]=label
```

In [23]:

```
df
```

Out[23]:

	D1	D2	D3	D4	D5	predicted_cluster
0	3.805564	5.384041	4.564335	4.266158	7.348113	2
1	-2.545971	-2.601637	-2.910390	-3.785392	-6.392758	1
2	-4.667780	-3.411962	-3.442990	-4.115973	-8.019540	1
3	0.766852	-0.386541	-0.200627	-0.057678	0.505778	0
4	-0.207983	-0.104326	0.016167	0.368387	-0.096917	0
...	...	...	...	...	...	...
2995	3.138888	3.903238	5.403115	5.628149	6.567997	2
2996	-2.640751	-5.582494	-2.756398	-6.422795	-4.077675	1
2997	2.313257	5.220735	4.589131	4.117231	5.454372	2
2998	-2.515320	-4.219420	-5.240284	-5.228420	-5.342212	1
2999	0.348203	-0.826678	0.727258	-0.015884	-0.118320	0

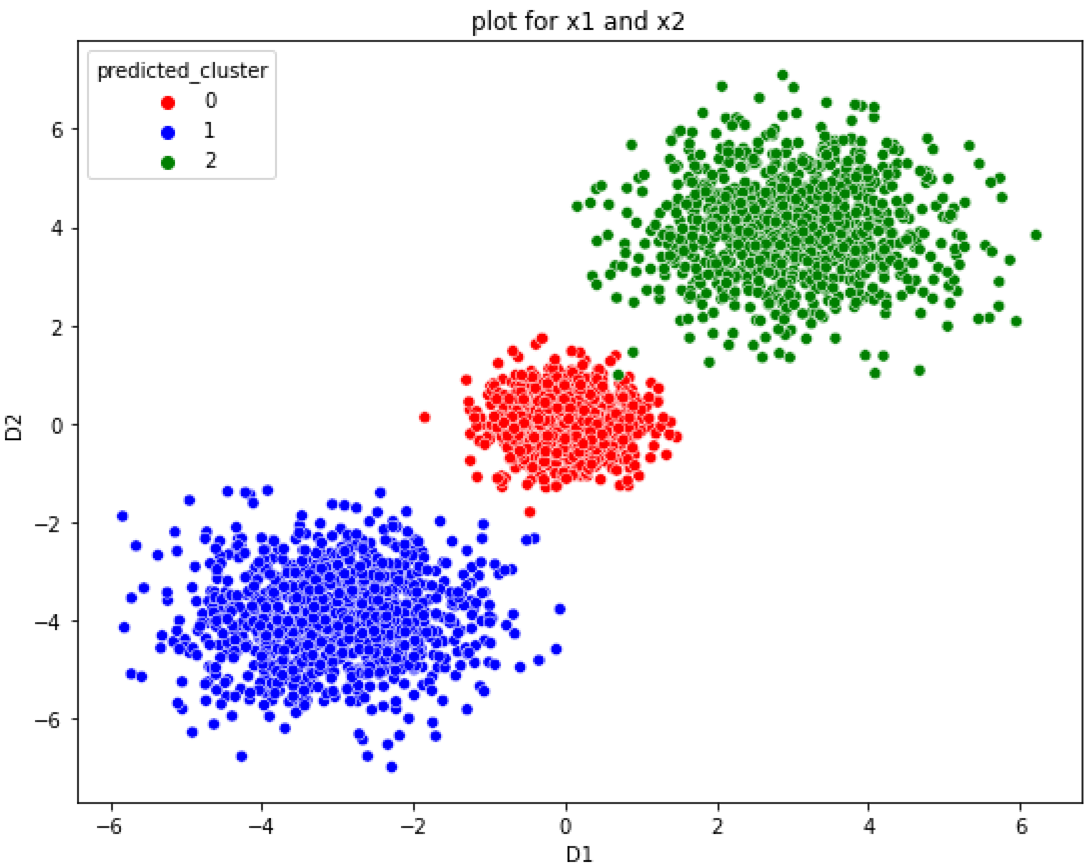
3000 rows × 6 columns

In [24]:

```
plt.figure(figsize=(9,7))
sns.scatterplot(data = df,
                x = df['D1'],
                y = df['D2'],
                hue = "predicted_cluster",
                palette=["red","blue","green"]).set(title = "plot for x1 and x2")
```

Out[24]:

```
[Text(0.5, 1.0, 'plot for x1 and x2')]
```

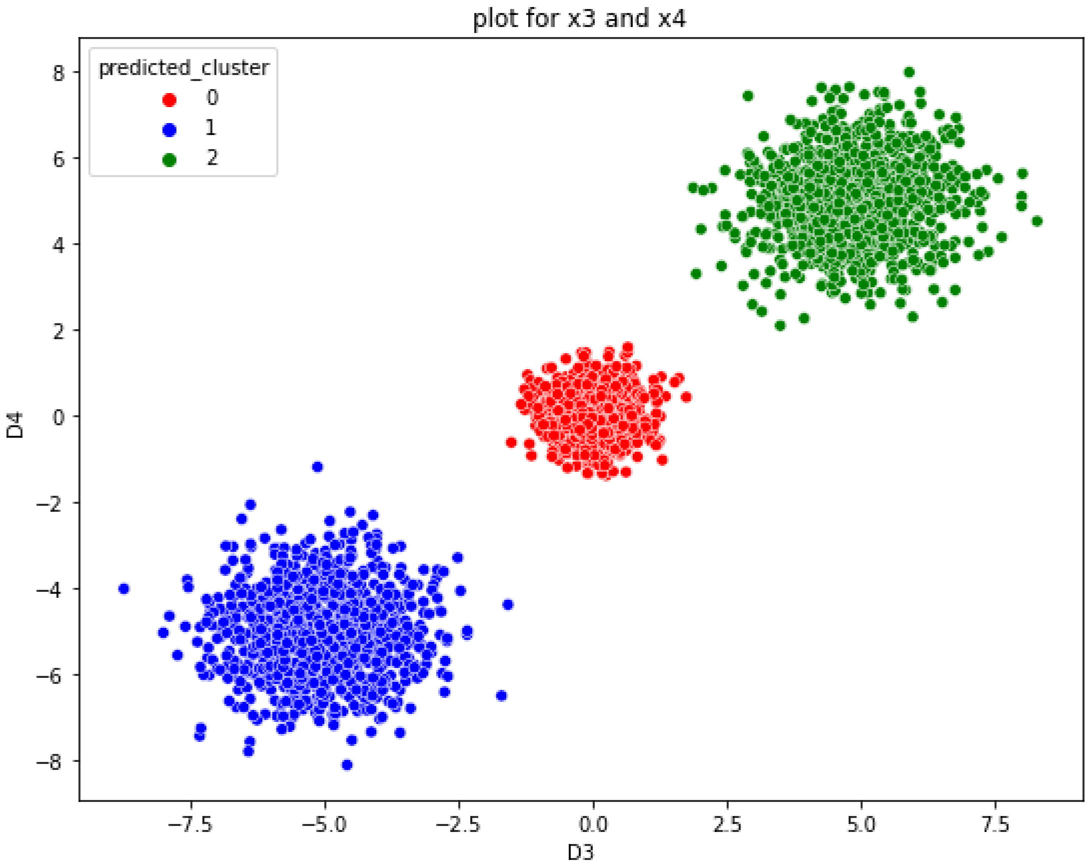


In [25]:

```
plt.figure(figsize=(9,7))
sns.scatterplot(data = df,
                x = df['D3'],
                y = df['D4'],
                hue = "predicted_cluster",
                palette=["red","blue","green"]).set(title = "plot for x3 and x4")
```

Out[25]:

```
[Text(0.5, 1.0, 'plot for x3 and x4')]
```

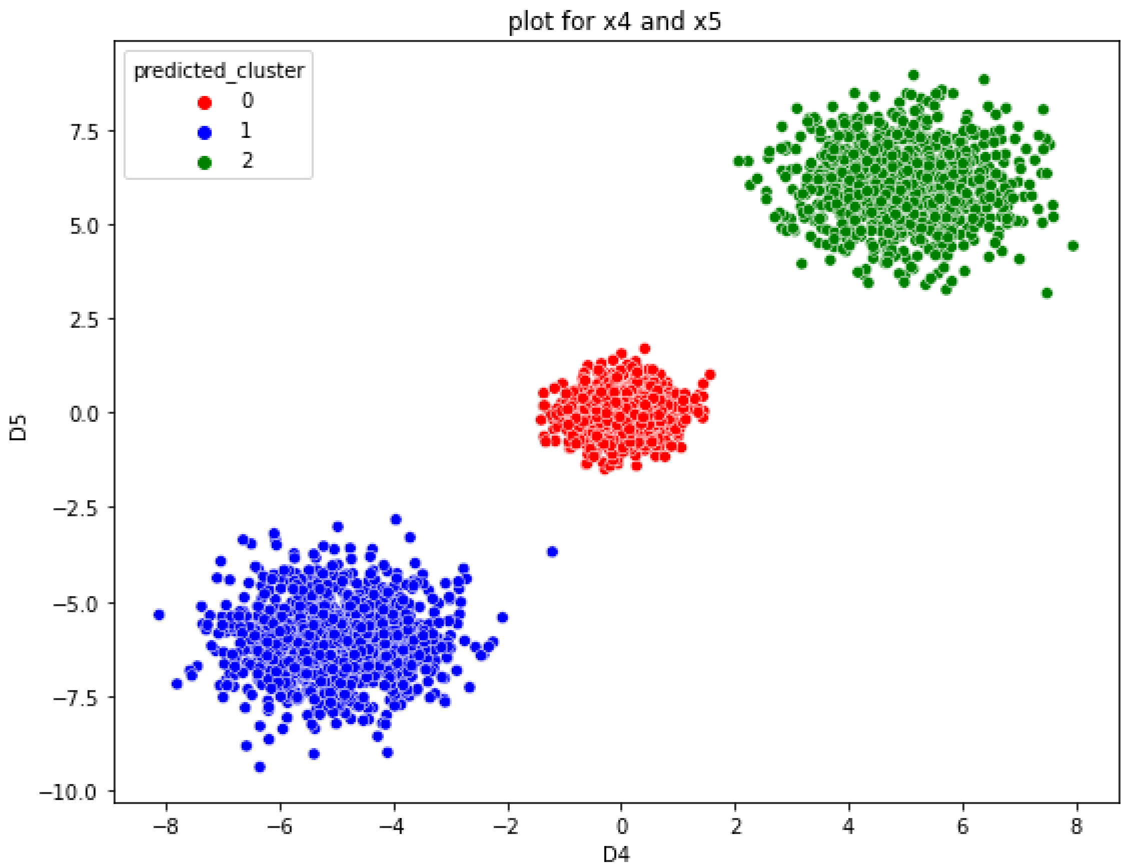


In [26]:

```
plt.figure(figsize=(9,7))
sns.scatterplot(data = df,
                x = df['D4'],
                y = df['D5'],
                hue = "predicted_cluster",
                palette=["red","blue","green"]).set(title = "plot for x4 and x5")
```

Out[26]:

```
[Text(0.5, 1.0, 'plot for x4 and x5')]
```



d

In [27]:

```
import matplotlib.pyplot as plt
from matplotlib.patches import Ellipse
from scipy.stats import multivariate_normal
plt.style.use('seaborn')
from sklearn import mixture
```

In [28]:

```
# update W
def update_W(data, Mu, Var, Pi):
    n_points, n_clusters = len(data), len(Pi)
    pdfs = np.zeros(((n_points, n_clusters)))
    for i in range(n_clusters):
        pdfs[:, i] = Pi[i] * multivariate_normal.pdf(data, Mu[i], np.diag(Var[i]))
    W = pdfs / pdfs.sum(axis=1).reshape(-1, 1)
    return W


# update pi
def update_Pi(W):
    Pi = W.sum(axis=0) / W.sum()
    return Pi


# calculate loglikelihoood function
def logLH(data, Pi, Mu, Var):
    n_points, n_clusters = len(data), len(Pi)
    pdfs = np.zeros(((n_points, n_clusters)))
    for i in range(n_clusters):
```

```
        pdfs[:, i] = Pi[i] * multivariate_normal.pdf(data, Mu[i], np.diag(Var[i]))
    return np.mean(np.log(pdfs.sum(axis=1)))

# plot the clusterings
def plot_clusters(X, Mu, Var, Mu_true=None, Var_true=None):
    colors = ['b', 'g', 'r']
    n_clusters = len(Mu)
    plt.figure(figsize=(10, 8))
    plt.axis([-10, 15, -5, 15])
    plt.scatter(X[:, 0], X[:, 1], s=5)
    ax = plt.gca()
    for i in range(n_clusters):
        plot_args = {'fc': 'None', 'lw': 5, 'edgecolor': colors[i], 'ls': ':'}
        ellipse = Ellipse(Mu[i], 3 * Var[i][0], 3 * Var[i][1], **plot_args)
        ax.add_patch(ellipse)
    if (Mu_true is not None) & (Var_true is not None):
        for i in range(n_clusters):
            plot_args = {'fc': 'None', 'lw': 5, 'edgecolor': colors[i], 'alpha': 0.5}
            ellipse = Ellipse(Mu_true[i], 3 * Var_true[i][0], 3 * Var_true[i][1], **plo
            ax.add_patch(ellipse)
    plt.show()

# update mu
def update_Mu(data, W):
    n_clusters = W.shape[1]
    Mu = np.zeros((n_clusters, 5))
    for i in range(n_clusters):
        Mu[i] = np.average(data, axis=0, weights=W[:, i])
    return Mu


# update Var
def update_Var(data, Mu, W):
    n_clusters = W.shape[1]
    Var = np.zeros((n_clusters, 5))
    for i in range(n_clusters):
        Var[i] = np.average((data - Mu[i]) ** 2, axis=0, weights=W[:, i])
    return Var
```

In [29]:

```
#initialization
true_Mu=[[-3,-4,-5,-5,-6],[3,4,5,5,6],[0,0,0,0,0]]
true_Var = [np.ones(5),np.ones(5),np.ones(5)*0.25]
true_Pi = np.ones(3)/3
Mu = np.random.rand(3,5)
Var=true_Var
Pi = true_Pi
n_clusters=3
n_points = len(data)
W = np.ones((n_points, n_clusters)) / n_clusters

loglh = []
iter=1
diff=None
while len(loglh)==0 or diff>=0.001:
    newlog=logLH(data, Pi, Mu, Var)
    loglh.append(newlog)
    if len(loglh)!=1:
        diff=abs(loglh[-1]-loglh[-2])
```

```python
    else:
        diff=abs(loglh[-1])
    W = update_W(data, Mu, Var, Pi)
    Pi = update_Pi(W)
    Mu = update_Mu(data, W)
    print('step %1d:log-likehood:%.4f'%(iter,loglh[-1]))
    Var = update_Var(data, Mu, W)
    iter+=1

print(np.round(Mu,4))
print(true_Mu)
```

```
step 1:log-likehood:-44.3113
step 2:log-likehood:-10.2410
step 3:log-likehood:-6.9994
step 4:log-likehood:-6.9785
step 5:log-likehood:-6.9785
[[ 2.9952e+00  3.9793e+00  4.9355e+00  4.9500e+00  6.0293e+00]
 [-3.0618e+00 -3.9670e+00 -5.0205e+00 -5.0202e+00 -6.0042e+00]
 [ 1.0500e-02  4.2000e-03  1.3600e-02  1.0000e-02 -2.4700e-02]]
[[-3, -4, -5, -5, -6], [3, 4, 5, 5, 6], [0, 0, 0, 0, 0]]
```

e

In [30]:

```python
#initialization
true_Mu=[[-3,-4,-5,-5,-6],[3,4,5,5,6],[0,0,0,0,0]]
true_Var = [np.ones(5),np.ones(5),np.ones(5)*0.25]
true_Pi=np.ones(3)/3

Mu = np.random.rand(3,5)
Var=true_Var
Pi = np.random.rand(3)
Pi = Pi/sum(Pi)
n_points = len(data)
W = np.ones((n_points, n_clusters)) / n_clusters
#Pi = W.sum(axis=0) / W.sum()

loglh = []
iter=1
diff=None
while len(loglh)==0 or diff>=0.001:
    newlog=logLH(data, Pi, Mu, Var)
    loglh.append(newlog)
    if len(loglh)!=1:
        diff=abs(loglh[-1]-loglh[-2])
    else:
        diff=abs(loglh[-1])
    print('step %1x: log-likehood:%.4f'%(iter,loglh[-1]),'Pi:',Pi)
    W = update_W(data, Mu, Var, Pi)
    Pi = update_Pi(W)
    Mu = update_Mu(data, W)
    Var = update_Var(data, Mu, W)
    iter+=1

print(np.round(Mu,4))
print(true_Mu)
print(np.round(Pi,4))
print(true_Pi)
```

```
step 1: log-likehood:-41.6272 Pi: [0.34797088 0.42444785 0.22758127]
step 2: log-likehood:-9.2146 Pi: [0.49038822 0.41908448 0.0905273 ]
step 3: log-likehood:-7.0494 Pi: [0.3417465  0.33397229 0.32428121]
step 4: log-likehood:-6.9785 Pi: [0.33333333 0.33333333 0.33333333]
step 5: log-likehood:-6.9785 Pi: [0.33333333 0.33333333 0.33333333]
[[-3.0618e+00 -3.9670e+00 -5.0205e+00 -5.0202e+00 -6.0042e+00]
 [ 2.9952e+00  3.9793e+00  4.9355e+00  4.9500e+00  6.0293e+00]
 [ 1.0500e-02  4.2000e-03  1.3600e-02  1.0000e-02 -2.4700e-02]]
[[-3, -4, -5, -5, -6], [3, 4, 5, 5, 6], [0, 0, 0, 0, 0]]
[0.3333 0.3333 0.3333]
[0.33333333 0.33333333 0.33333333]
```