

Trabajo práctico 0: Repaso de Probabilidades, Álgebra Lineal y el Algoritmo de Maximización de la Esperanza

Juan José Cordero
Luis Diego Hidalgo
Ricardo Sánchez
Instituto Tecnológico de Costa Rica
MC-8840 Deep Learning
Maestría en Ciencias de la Computación

9 de septiembre de 2022

1. (10 puntos) Demuestre que si $\vec{k} = (\vec{u} - \vec{w})$ y los vectores \vec{v} y \vec{k} son perpendiculares, entonces $\vec{v} \cdot \vec{u} = \vec{v} \cdot \vec{w}$. Enliste todas las propiedades del producto punto usadas.

Si $\vec{v} \perp \vec{k}$ (son perpendiculares) por definición se tiene que:

$$\vec{v} \cdot \vec{k} = \vec{v}^T \cdot \vec{k} = 0.$$

Lo anterior también se puede probar aplicando la fórmula del producto punto:

$$\vec{v} \cdot \vec{k} = \|\vec{v}\| \cdot \|\vec{k}\| \cdot \cos(\theta)$$

Al ser vectores perpendiculares se tiene que $\Theta = 90$, por lo que:

$$\|\vec{v}\| \cdot \|\vec{k}\| \cdot \cos(90) = \|\vec{v}\| \cdot \|\vec{k}\| \cdot 0 = 0$$

Para la demostración se asumen lo siguientes valores:

$$\vec{k} = (a, b)$$

$$\vec{u} = (c, d)$$

$$\vec{w} = (h, i)$$

$$\vec{v} = (j, k)$$

Si $\vec{k} = (\vec{u} - \vec{w})$ entonces $\vec{k} = ((c - h), (d - i))$ y como $\vec{v} \perp \vec{k} = \vec{v} \cdot \vec{k} = (ja + kb) = (aj + bk) = 0$

Ahora vamos a demostrar que $\vec{v} \cdot \vec{u} = \vec{v} \cdot \vec{w}$:

$$\vec{v} \cdot \vec{u} = \vec{v} \cdot \vec{w}$$

$$jc + kd = jh + ki$$

$$jc - jh = ki - kd$$

$$j(c - h) = -k(d - i)$$

A este punto vemos como es posible hacer el reemplazo términos ya que:

$$\vec{k} = (a, b) \text{ y } \vec{k} = ((c - h), (d - i))$$

Entonces se puede decir que $(c - h) = a$ y $(d - i) = b$

$$aj = -bk$$

$$aj + bk = 0$$

Lo cual, según lo demostrado anteriormente, es equivalente a:

$$\vec{v} \perp \vec{k} = \vec{v} \cdot \vec{k} = (ja + kb) = (aj + bk) = 0$$

Por lo que se puede concluir que efectivamente la expresión: $\vec{v} \cdot \vec{u} = \vec{v} \cdot \vec{w}$ se cumple.

2. (10 puntos) A continuación realice las siguientes demostraciones usando notación algebraica matricial (En algunos ejercicios se recomienda usar la notación de filas y columnas compacta (filas y columnas en vectores), como la utilizada en el material del curso):

a) Si $\vec{x} \in \mathbb{R}^n$ y $A = \frac{\vec{x}\vec{x}^T}{\vec{x}^T\vec{x}}$, entonces $A^2 = A$

b) Sean las matrices $A, B \in \mathbb{R}^{n \times n}$, con $A + B = I$ y $AB = 0_n$, demuestre que $A^2 = A$ y $B^2 = B$.

a)

Para la demostración se asumen los siguientes valores:

$$\vec{x} = (a, b)$$

$$\vec{x}^T = \begin{pmatrix} a \\ b \end{pmatrix}$$

Se tiene que:

$$\frac{\vec{x} \vec{x}^T}{\vec{x}^T \vec{x}} = \frac{a^2 + b^2}{\begin{pmatrix} a^2 & ab \\ ab & b^2 \end{pmatrix}} = A$$

Hay que demostrar que:

$$\frac{(a^2 + b^2)^2}{\begin{pmatrix} a^2 & ab \\ ab & b^2 \end{pmatrix}} = A$$

Se despeja:

$$\begin{aligned} \frac{(a^2 + b^2)^2}{\begin{pmatrix} a^2 & ab \\ ab & b^2 \end{pmatrix}} &= \frac{(a^2 + b^2) \cdot (a^2 + b^2)}{\begin{pmatrix} a^2 + a^2 b^2 & a^2 \cdot ab + ab \cdot b^2 \\ a^2 \cdot ab + b^2 \cdot ab & a^2 b^2 + b^4 \end{pmatrix}} = \frac{(a^2 + b^2) \cdot (a^2 + b^2)}{(a^2 + b^2) \cdot \begin{pmatrix} a^2 & ab \\ ab & b^2 \end{pmatrix}} \\ &= \frac{a^2 + b^2}{\begin{pmatrix} a^2 & ab \\ ab & b^2 \end{pmatrix}} = A \end{aligned}$$

Por lo tanto: $A^2 = A$

b)

Demostración para $A^2 = A$:

$$A + B = I$$

$$A = I - B$$

$$A \cdot A = A(I - B) \text{ por propiedad distributiva.}$$

$A^2 = AI - AB$ como $AB = 0_n$ y una matriz multiplicada por su matriz de identidad (I) es ella misma, es decir, $AI = A$, entonces:

$$A^2 = A - 0_n$$

$$A^2 = A$$

Demostración para $B^2 = B$:

$$A + B = I$$

$$B = I - A$$

$B \cdot B = B(I - A)$ por propiedad distributiva.

$$B^2 = BI - BA$$

Como $AB = 0_n$, y ambas son matrices cuadradas de $n \times n$, se puede decir que $BA = 0_n$ y una matriz multiplicada por su matriz de identidad (I) es ella misma, es decir, $BI = B$, entonces:

$$B^2 = B - 0_n$$

$$B^2 = B$$

Dado lo anterior se puede demostrar que tanto $A^2 = A$ como $B^2 = B$ se cumplen.

2. (30 puntos) Funciones multivariable

1. El vector Gradiente: Para cada una de las siguientes funciones multivariable: grafique su superficie, calcule el vector gradiente manualmente, evalúelo y grafique el vector unitario o de largo conveniente para su visualización, en la dirección del gradiente para los dos puntos especificados (en la misma figura de la superficie). Finalmente calcule la magnitud de tal vector gradiente en cada punto.

a) (5 puntos) $f(x, y) = \sqrt{x^2 + y^2}$, evaluación del gradiente en los puntos $P_0 = (5.2, 6.4)$ y $P_1 = (5.2, 2.3)$.

El vector gradiente corresponde a:

$$\frac{\partial f}{\partial x} \sqrt{x^2 + y^2} \rightarrow (x^2 + y^2)^{\frac{1}{2}} = \frac{(x^2 + y^2)^{\frac{1}{2}-1}}{2} \cdot 2x = \frac{x}{(x^2 + y^2)^{\frac{1}{2}}} = \frac{x}{\sqrt{x^2 + y^2}}$$

$$\frac{\partial f}{\partial y} \sqrt{x^2 + y^2} \rightarrow (x^2 + y^2)^{\frac{1}{2}} = \frac{(x^2 + y^2)^{\frac{1}{2}-1}}{2} \cdot 2y = \frac{y}{(x^2 + y^2)^{\frac{1}{2}}} = \frac{y}{\sqrt{x^2 + y^2}}$$

$$\nabla f(x, y) = \frac{x}{\sqrt{x^2 + y^2}} \hat{i} + \frac{y}{\sqrt{x^2 + y^2}} \hat{j}$$

Gráfica de la superficie de la función:

```
In [34]: import torch
import math
import numpy as np
import random
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
from matplotlib import cm
from matplotlib import style
from scipy.stats import norm
# Requerido para algoritmo de maximización de la esperanza, de lo contrario se mezclan
%matplotlib inline
style.use('default')
```

```
In [35]: #Función multivariable original
def funcion_z(X,Y):
    return torch.sqrt(X**2 + Y**2)

#Calculo del Vector Gradiente (Derivadas parciales)
def dx(X,Y):
    return X / torch.sqrt(X**2 + Y**2)

def dy(X,Y):
    return Y / torch.sqrt(X**2 + Y**2)

def vector_gradiente(X,Y):
    return [dx(X,Y), dy(X,Y), 0]
```

```

P0=torch.tensor([5.2,6.4], dtype = torch.float)
P1=torch.tensor([5.2,2.3], dtype = torch.float)

vector_p0=torch.tensor([vector_gradiente(P0[0],P0[1])], dtype = torch.float)
vector_p1=torch.tensor([vector_gradiente(P1[0],P1[1])], dtype = torch.float)

print("El valor del vector gradiente para el punto P0=(5.2, 6.4) es: ", vector_p0)
print("El valor del vector gradiente para el punto P1=(5.2, 2.3) es: ", vector_p1)

El valor del vector gradiente para el punto P0=(5.2, 6.4) es:  tensor([[0.6306, 0.776
1, 0.0000]])
El valor del vector gradiente para el punto P1=(5.2, 2.3) es:  tensor([[0.9145, 0.404
5, 0.0000]])

```

```

In [36]: #Inicia el plot
N=256
arrange=[-1,1]

x_values=torch.linspace(arrange[0],arrange[1],steps=N)
y_values=torch.linspace(arrange[0],arrange[1],steps=N)

X, Y = torch.meshgrid(x_values,y_values)
Z = funcion_z(X,Y)

fig=plt.figure(figsize=(7,7))

ax = plt.axes(projection='3d')

ax.set_xlabel('x')
ax.set_ylabel('y')
ax.set_zlabel('z')

origin=[0,0,0]
ax.quiver(origin[0], origin[1], origin[2], vector_p0[:,0],vector_p0[:,1],vector_p0[:,2])
ax.quiver(origin[0], origin[1], origin[2], vector_p1[:,0],vector_p1[:,1],vector_p1[:,2])

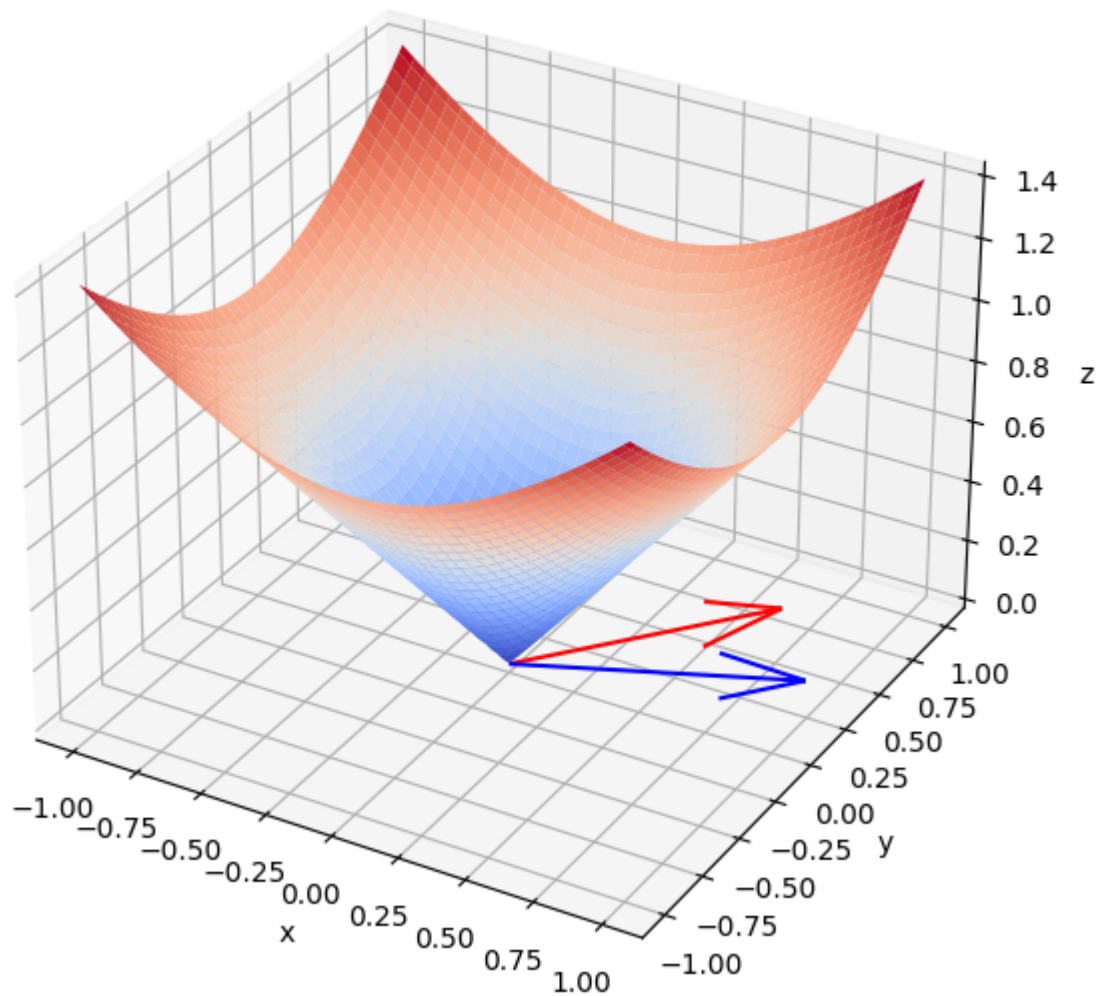
ax.plot_surface(X,Y,Z,cmap=cm.coolwarm,linewidth=10,antialiased=True)

```

```

Out[36]: <mpl_toolkits.mplot3d.art3d.Poly3DCollection at 0x29f9908f8e0>

```



```
In [4]: #Magnitud de Los vectores:
print("La magnitud del vector P0 es: ", torch.linalg.vector_norm(vector_p0).item())
print("La magnitud del vector P1 es: ", torch.linalg.vector_norm(vector_p1).item())
```

La magnitud del vector P0 es: 1.0

La magnitud del vector P1 es: 1.0

b) (5 puntos) $z = f(x, y) = 3x^2 + 2y^4$, evaluación del gradiente en los puntos $P_0 = (0, 0)$ y $P_1 = (7.4, -6.3)$.

El vector gradiente corresponde a:

$$\frac{\partial f}{\partial x} 3x^2 + 2y^4 \rightarrow 3x^2 = 6x$$

$$\frac{\partial f}{\partial y} 3x^2 + 2y^4 \rightarrow 2y^4 = 8y^3$$

$$\nabla f(x, y) = 6x\hat{i} + 8y^3\hat{j}$$

```
In [5]: #Función multivariable original
def funcion_z(X, Y):
    return (3*X**2) + (2*Y**4)
```

```

#Calculo del Vector Gradiente (Derivadas parciales)
def dx(X,Y):
    return 6*X

def dy(X,Y):
    return 8*Y**3

def vector_gradiente(X,Y):
    return [dx(X,Y),dy(X,Y),0]

P0=torch.tensor([0,0], dtype = torch.float)
P1=torch.tensor([7.4,-6.3], dtype = torch.float)

vector_p0=torch.tensor([vector_gradiente(P0[0],P0[1])], dtype = torch.float)
vector_p1=torch.tensor([vector_gradiente(P1[0],P1[1])], dtype = torch.float)

print("El valor del vector gradiente para el punto P0=(0, 0) es: ", vector_p0)
print("El valor del vector gradiente para el punto P1=(7.4, -6.3) es: ", vector_p1)

El valor del vector gradiente para el punto P0=(0, 0) es: tensor([[0., 0., 0.]])
El valor del vector gradiente para el punto P1=(7.4, -6.3) es: tensor([[ 44.4000,
-2000.3762,      0.0000]])

```

```

In [6]: #Inicia el plot
N=256
arrange=[-1,1]

x_values=torch.linspace(arrange[0],arrange[1],steps=N)
y_values=torch.linspace(arrange[0],arrange[1],steps=N)

X, Y = torch.meshgrid(x_values,y_values)
Z = funcion_z(X,Y)

fig=plt.figure(figsize=(7,7))

ax = plt.axes(projection='3d')

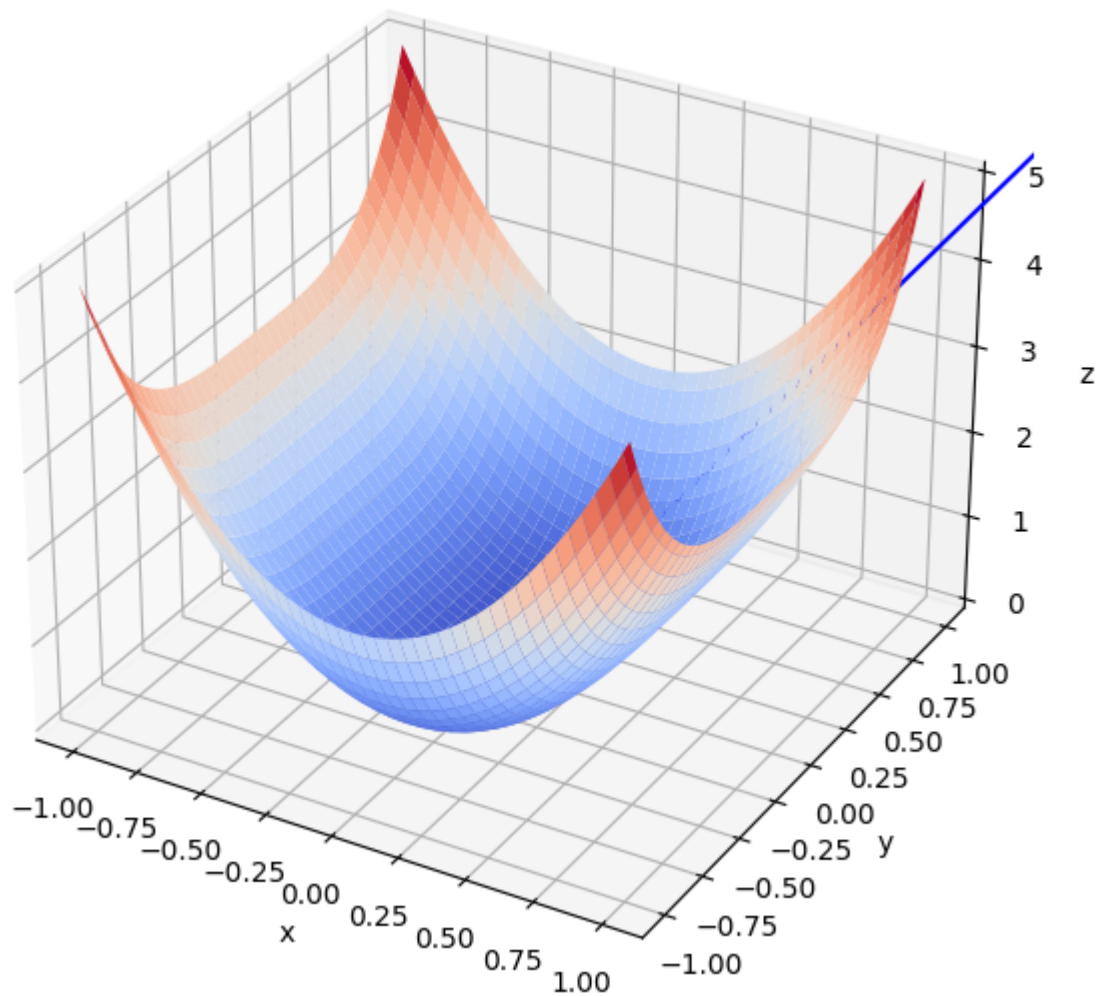
ax.set_xlabel('x')
ax.set_ylabel('y')
ax.set_zlabel('z')

origin=[0,0,0]
ax.quiver(origin[0], origin[1], origin[2], vector_p0[:,0],vector_p0[:,1],vector_p0[:,2])
ax.quiver(origin[0], origin[1], origin[2], vector_p1[:,0],vector_p1[:,1],vector_p1[:,2])

ax.plot_surface(X,Y,Z,cmap=cm.coolwarm,linewidth=10,antialiased=True)

```

Out[6]: <mpl_toolkits.mplot3d.art3d.Poly3DCollection at 0x29f97522e80>



```
In [7]: #Magnitud de Los vectores:
print("La magnitud del vector P0 es: ", torch.linalg.vector_norm(vector_p0))
print("La magnitud del vector P1 es: ", torch.linalg.vector_norm(vector_p1))
```

La magnitud del vector P0 es: tensor(0.)
 La magnitud del vector P1 es: tensor(2000.8689)

c) (5 puntos) $z = f(x, y) = 4x^2 + 2x + e^{2x} + 5y^2 + e^{3y} + 1$, evaluación del gradiente en los puntos $P_0 = (2, 1)$ y $P_1 = (5, 7)$.

El vector gradiente corresponde a:

$$\frac{\partial f}{\partial x} 4x^2 + 2x + e^{2x} + 5y^2 + e^{3y} + 1 \rightarrow 8x + 2 + 2e^{2x} + 0 + 0 + 0 = 8x + 2 + 2e^{2x} = \mathbf{2(4x + 1 + e^{2x})}$$

$$\frac{\partial f}{\partial y} 4x^2 + 2x + e^{2x} + 5y^2 + e^{3y} + 1 \rightarrow 0 + 0 + 0 + 10y + 3e^{3y} + 0 = \mathbf{10y + 3e^{3y}}$$

$$\nabla f(x, y) = 2(4x + e^{2x} + 1)\hat{i} + 10y + 3e^{3y}\hat{j}$$

Gráfica de la superficie de la función:

```
In [15]: #Función multivariable original
def funcion_z(X, Y):
    return (4*X**2) + (2*X) + (torch.e**(2*X)) + (5*Y**2) + (torch.e**(3*Y)) + 1

#Calculo del Vector Gradiente (Derivadas parciales)
def dx(X,Y):
    return 2*((4*X)+torch.e**(2*X)+1)

def dy(X,Y):
    return (10*Y)+(3*torch.e**(3*Y))

def vector_gradiente(X,Y):
    return [dx(X,Y),dy(X,Y),0]

P0=torch.tensor([2,1], dtype = torch.float)
P1=torch.tensor([5,7], dtype = torch.float)

vector_p0=torch.tensor([vector_gradiente(P0[0],P0[1])], dtype = torch.float)
vector_p1=torch.tensor([vector_gradiente(P1[0],P1[1])], dtype = torch.float)

print("El valor del vector gradiente para el punto P0=(2, 1) es: ", vector_p0)
print("El valor del vector gradiente para el punto P1=(5, 7) es: ", vector_p1)
```

El valor del vector gradiente para el punto P0=(2, 1) es: tensor([[127.1963, 70.2566, 0.0000]])

El valor del vector gradiente para el punto P1=(5, 7) es: tensor([[4.4095e+04, 3.9564e+09, 0.0000e+00]])

```
In [17]: #Inicia el plot
N=256
arrange=[-1,1]

x_values=torch.linspace(arrange[0],arrange[1],steps=N)
y_values=torch.linspace(arrange[0],arrange[1],steps=N)

X, Y = torch.meshgrid(x_values,y_values)
Z = funcion_z(X,Y)

fig=plt.figure(figsize=(7,7))

ax = plt.axes(projection='3d')

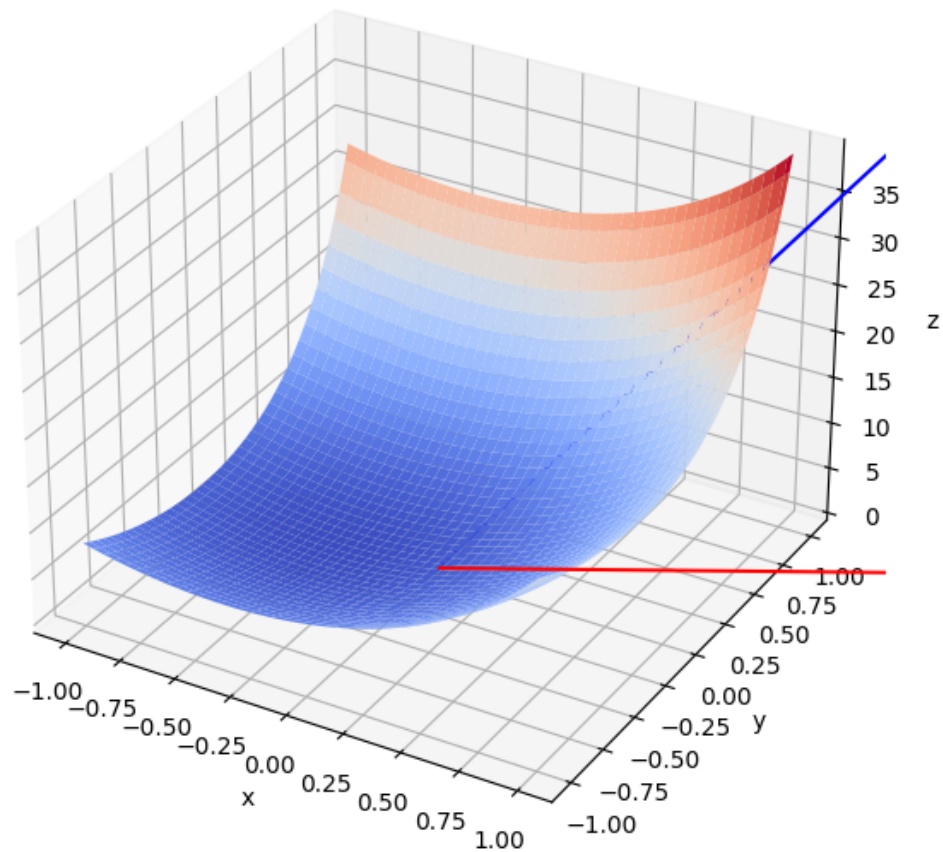
ax.set_xlabel('x')
ax.set_ylabel('y')
ax.set_zlabel('z')

origin=[0,0,0]
ax.quiver(origin[0], origin[1], origin[2], vector_p0[:,0],vector_p0[:,1],vector_p0[:,2])
ax.quiver(origin[0], origin[1], origin[2], vector_p1[:,0],vector_p1[:,1],vector_p1[:,2])

ax.plot_surface(X,Y,Z,cmap=cm.coolwarm,linewidth=10,antialiased=True)
```

Out[17]: <mpl_toolkits.mplot3d.art3d.Poly3DCollection at 0x20e39c20970>

Figure



```
In [18]: #Magnitud de Los vectores:
print("La magnitud del vector P0 es: ", torch.linalg.vector_norm(vector_p0))
print("La magnitud del vector P1 es: ", torch.linalg.vector_norm(vector_p1))
```

La magnitud del vector P0 es: tensor(145.3096)

La magnitud del vector P1 es: tensor(3.9564e+09)

d) (5 puntos) $z = f(x, y) = \sin(x^2) + x \cos(y^3)$, evaluación del gradiente en los puntos $P_0 = (-2, 6)$ y $P_1 = (0, 4)$.

El vector gradiente corresponde a:

$$\frac{\delta f}{\delta x} \sin(x^2) + x \cos(y^3) \rightarrow \cos(x^2) \cdot 2x + \cos(y^3) = 2\mathbf{x} \cdot \cos(\mathbf{x}^2) + \cos(\mathbf{y}^3)$$

$$\frac{\delta f}{\delta y} \sin(x^2) + x \cos(y^3) \rightarrow -x \cdot \text{sen}(y^3) \cdot 3y^2 = -3\mathbf{xy}^2 \cdot \text{sen}(\mathbf{y}^3)$$

$$\nabla f(\mathbf{x}, \mathbf{y}) = 2\mathbf{x} \cdot \cos(\mathbf{x}^2) + \cos(\mathbf{y}^3) \hat{\mathbf{i}} + -3\mathbf{xy}^2 \cdot \text{sen}(\mathbf{y}^3) \hat{\mathbf{j}}$$

```

In [19]: #Función multivariable original
def funcion_z(X, Y):
    return torch.sin(X**2) + (X*torch.cos(Y**3))

#Calculo del Vector Gradiente (Derivadas parciales)
def dx(X,Y):
    return 2*X*torch.cos(X**2)+torch.cos(Y**3)

def dy(X,Y):
    return -3*X*(Y**2)*torch.sin(Y**3)

def vector_gradiente(X,Y):
    return [dx(X,Y),dy(X,Y),0]

P0=torch.tensor([-2,6], dtype = torch.float)
P1=torch.tensor([0,4], dtype = torch.float)

vector_p0=torch.tensor([vector_gradiente(P0[0],P0[1])], dtype = torch.float)
vector_p1=torch.tensor([vector_gradiente(P1[0],P1[1])], dtype = torch.float)

print("El valor del vector gradiente para el punto P0=(-2, 6) es: ", vector_p0)
print("El valor del vector gradiente para el punto P1=(0, 4) es: ", vector_p1)

El valor del vector gradiente para el punto P0=(-2, 6) es: tensor([[ 1.8966, 150.34
86,  0.0000]])
El valor del vector gradiente para el punto P1=(0, 4) es: tensor([[0.3919, -0.0000,
0.0000]])

In [20]: #Inicia el plot
N=256
arrange=[-2,2]

x_values=torch.linspace(arrange[0],arrange[1],steps=N)
y_values=torch.linspace(arrange[0],arrange[1],steps=N)

X, Y = torch.meshgrid(x_values,y_values)
Z = funcion_z(X,Y)

fig=plt.figure(figsize=(7,7))

ax = plt.axes(projection='3d')

ax.set_xlabel('x')
ax.set_ylabel('y')
ax.set_zlabel('z')

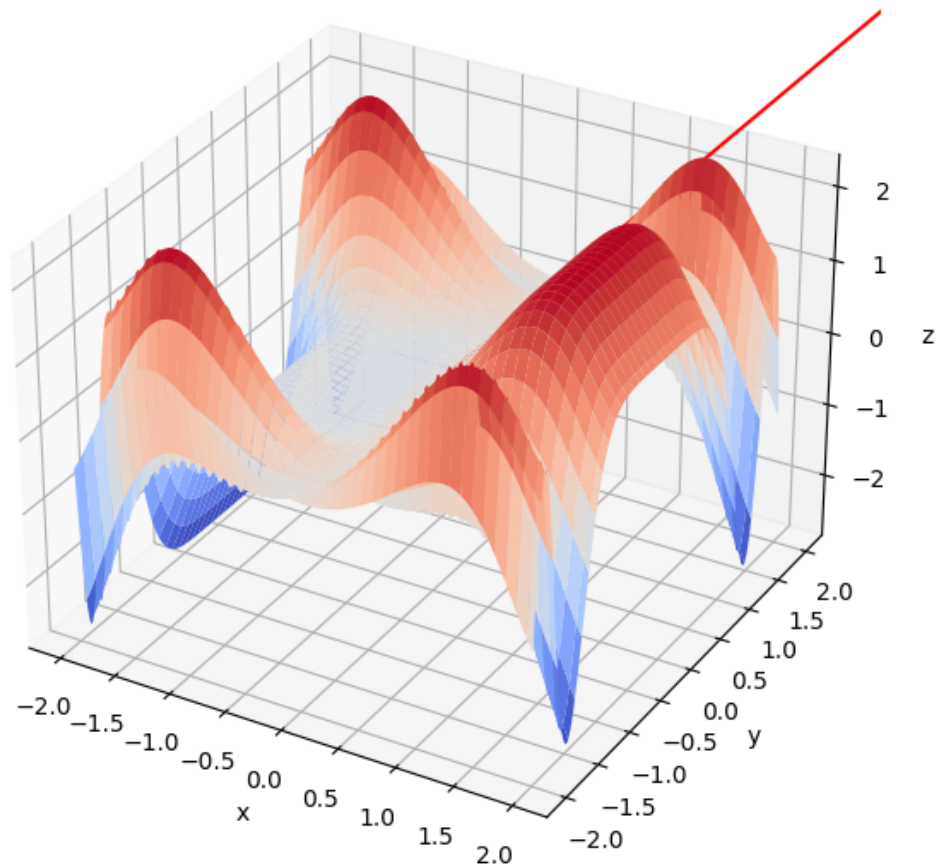
origin=[0,0,0]
ax.quiver(origin[0], origin[1], origin[2], vector_p0[:,0],vector_p0[:,1],vector_p0[:,2])
ax.quiver(origin[0], origin[1], origin[2], vector_p1[:,0],vector_p1[:,1],vector_p1[:,2])

ax.plot_surface(X,Y,Z,cmap=cm.coolwarm,linewidth=10,antialiased=True)

Out[20]: <mpl_toolkits.mplot3d.art3d.Poly3DCollection at 0x20e33ea6fa0>

```

Figure



```
In [21]: #Magnitud de Los vectores:
print("La magnitud del vector P0 es: ", torch.linalg.vector_norm(vector_p0))
print("La magnitud del vector P1 es: ", torch.linalg.vector_norm(vector_p1))

La magnitud del vector P0 es:  tensor(150.3606)
La magnitud del vector P1 es:  tensor(0.3919)
```

2. (10 puntos) En general, investigue ¿qué es y que indica la matriz Hessiana? Sobre las aplicaciones de esta matriz, se pueden citar los siguientes puntos:

- Se define como una matriz cuadrada de $n \times n$ que se compone de las segundas derivada parciales de la función multivariable, por ejemplo para funciones de dos variables:

$$\mathbf{H}_f(x, y) = \begin{pmatrix} \frac{\delta^2 f}{\delta x^2} & \frac{\delta^2 f}{\delta y \delta x} \\ \frac{\delta^2 f}{\delta y \delta x} & \frac{\delta^2 f}{\delta y^2} \end{pmatrix}$$

- Por Teorema de Schwarz se puede decir que $\frac{\delta^2 f}{\delta y \delta x} = \frac{\delta^2 f}{\delta x \delta y}$
- Permite encontrar ya sea máximos o mínimos de funciones multivariable. Para conseguir esto, el procedimiento a seguir consiste de obtener los puntos críticos de la función (igualando a cero el vector gradiente y obteniendo los puntos respectivos del despeje) y operarlos con la matriz Hessiana.
- Una vez obtenido el resultado, la matriz que se consiguió debe ser evaluada bajo el criterio de: definida positiva, definida negativa, indefinida, etc (lo cual puede ser establecido con el Criterio de los valores propios o con el Criterio de Sylvester).
- Dependiendo del tipo de matriz obtenida, así podrá definirse el punto que fue utilizado para el cálculo (máximo, mínimo o un punto neutro).
- También permite saber si una función es cóncava o convexa con respecto a un conjunto de puntos pertenecientes a la función, aplicando nuevamente el concepto de definida positiva, definida negativa, indefinida, etc.
- Fuentes utilizadas para cálculo y usos:
 - [Matriz Hessiana \(o Hessiano\)](#)
 - [La matriz hessiana](#)

a) Para cada una de los puntos 1.a, 1.b, 1.c, y 1.d calcule la matriz Hessiana:

- 1.a Para $f(x, y) = \sqrt{x^2 + y^2}$
 - $\frac{\delta f}{\delta x} = \frac{x}{\sqrt{x^2 + y^2}}$
 - $\frac{\delta f}{\delta y} = \frac{y}{\sqrt{x^2 + y^2}}$
 - $\frac{\delta^2 f}{\delta x^2} = \frac{y^2}{\sqrt{(x^2 + y^2)^3}}$
 - $\frac{\delta^2 f}{\delta y^2} = \frac{x^2}{\sqrt{(x^2 + y^2)^3}}$
 - $\frac{\delta^2 f}{\delta y \delta x} = \frac{\delta^2 f}{\delta x \delta y} = \frac{2xy^2 - x^3}{(x^2 + y^2)\sqrt{(x^2 + y^2)^3}}$
 - $\mathbf{H}_f(x, y) = \begin{pmatrix} \frac{\delta^2 f}{\delta x^2} & \frac{\delta^2 f}{\delta x \delta y} \\ \frac{\delta^2 f}{\delta y \delta x} & \frac{\delta^2 f}{\delta y^2} \end{pmatrix} = \begin{pmatrix} \frac{y^2}{\sqrt{(x^2 + y^2)^3}} & \frac{2xy^2 - x^3}{(x^2 + y^2)\sqrt{(x^2 + y^2)^3}} \\ \frac{2xy^2 - x^3}{(x^2 + y^2)\sqrt{(x^2 + y^2)^3}} & \frac{x^2}{\sqrt{(x^2 + y^2)^3}} \end{pmatrix}$
- 1.b Para $z = f(x, y) = 3x^2 + 2y^4$
 - $\frac{\delta f}{\delta x} = 6x$
 - $\frac{\delta f}{\delta y} = 8y^3$

- $\frac{\delta^2 f}{\delta x^2} = 6$
 - $\frac{\delta^2 f}{\delta y^2} = 24y^2$
 - $\frac{\delta^2 f}{\delta y \delta x} = \frac{\delta^2 f}{\delta x \delta y} = 0$
 - $\mathbf{H}_f(x, y) = \begin{pmatrix} \frac{\delta^2 f}{\delta x^2} & \frac{\delta^2 f}{\delta x \delta y} \\ \frac{\delta^2 f}{\delta y \delta x} & \frac{\delta^2 f}{\delta y^2} \end{pmatrix} = \begin{pmatrix} 6 & 0 \\ 0 & 24y^2 \end{pmatrix}$
- 1.c Para $z = f(x, y) = 4x^2 + 2x + e^{2x} + 5y^2 + e^{3y} + 1$
- $\frac{\delta f}{\delta x} = 2(4x + e^{2x} + 1)$
 - $\frac{\delta f}{\delta y} = 10y + 3e^{3y}$
 - $\frac{\delta^2 f}{\delta x^2} = 4(e^{2x} + 2)$
 - $\frac{\delta^2 f}{\delta y^2} = 9e^{3y} + 10$
 - $\frac{\delta^2 f}{\delta y \delta x} = \frac{\delta^2 f}{\delta x \delta y} = 0$
 - $\mathbf{H}_f(x, y) = \begin{pmatrix} \frac{\delta^2 f}{\delta x^2} & \frac{\delta^2 f}{\delta x \delta y} \\ \frac{\delta^2 f}{\delta y \delta x} & \frac{\delta^2 f}{\delta y^2} \end{pmatrix} = \begin{pmatrix} 4(e^{2x} + 2) & 0 \\ 0 & 9e^{3y} + 10 \end{pmatrix}$
- 1.b Para $z = f(x, y) = \sin(x^2) + x \cos(y^3)$
- $\frac{\delta f}{\delta x} = 2x \cdot \cos(x^2) + \cos(y^3)$
 - $\frac{\delta f}{\delta y} = -3xy^2 \cdot \sin(y^3)$
 - $\frac{\delta^2 f}{\delta x^2} = 2\cos(x^2) - 2x^2 \sin(x^2)$
 - $\frac{\delta^2 f}{\delta y^2} = -3xy(2\sin(y^3) + 3y^3 \cos(y^3))$
 - $\frac{\delta^2 f}{\delta y \delta x} = \frac{\delta^2 f}{\delta x \delta y} = -3y^2 \sin(y^3)$
 - $\mathbf{H}_f(x, y) = \begin{pmatrix} \frac{\delta^2 f}{\delta x^2} & \frac{\delta^2 f}{\delta x \delta y} \\ \frac{\delta^2 f}{\delta y \delta x} & \frac{\delta^2 f}{\delta y^2} \end{pmatrix} = \begin{pmatrix} 2\cos(x^2) - 2x^2 \sin(x^2) & -3y^2 \sin(y^3) \\ -3y^2 \sin(y^3) & -3xy(2\sin(y^3) + 3y^3 \cos(y^3)) \end{pmatrix}$

3. (30 puntos) Probabilidades: Algoritmo de Maximización de la Esperanza

A continuación, implemente el algoritmo de maximización de la esperanza (descrito en el material del curso), usando la definición y descripción de las siguientes funciones como base:

```
In [8]: # Constantes
# Rangos para inicializar mu
MU_START = 10
MU_END = 50
# Rangos para inicializar sigma
SIGMA_START = 3.1
SIGMA_END = 6.2
# Dato para solucion heuristica
HEURISTIC_STEP = 5

In [9]: # Graficar observacion con su funcion de densidad de probabilidad
def plot_observation(observation, show=False):
    mu = torch.mean(observation)
    sigma = torch.std(observation, unbiased=True)
    x_axis = torch.arange(min(observation) - 5, max(observation) + 5, 0.01)
    plt.scatter(observation.numpy(), torch.zeros(len(observation)), s=5, alpha=0.5)
    plt.plot(x_axis.numpy(), norm.pdf(x_axis.numpy(), mu.numpy(), sigma.numpy()),
             label=r'$\mu=' + str(round(mu.item(), 2)) + r',\ \sigma=' + str(round(sigma.item(), 2)))
    if show:
        plt.legend()
        plt.show()

# Graficar las distribuciones aleatorias junto con las observaciones
def plot_gaussian_distribution_and_observations(distribution_parameters, observations):
    for observation in observations:
        plot_observation(observation)
    param_number = 1
    for parameters in distribution_parameters:
        mu = parameters[0]
        sigma = parameters[1]
        x_axis = torch.arange(mu / 2, mu * 2, 0.01)
        plt.plot(x_axis.numpy(), norm.pdf(x_axis.numpy(), mu.numpy(), sigma.numpy()),
                 label=r'$\mu_' + str(param_number) + r'=' + str(round(mu.item(), 2))
                 + r',\ \sigma_' + str(param_number) + r'=' + str(round(sigma.item(), 2)))
        param_number += 1
    if show:
        plt.legend()
        plt.show()
```

1. Función generate_data:

```
In [10]: # Genera datos siguiendo una distribucion gaussiana
def generate_data(n_observations: int, k_parameters=2, show=False, heuristic=False):
    gaussian_distributions = []
    heuristic_mu = random.uniform(MU_START, MU_END) if heuristic else 0
    for k in range(k_parameters):
        mu = torch.tensor(random.uniform(MU_START, MU_END)) if not heuristic else torch.tensor(heuristic_mu)
```



```

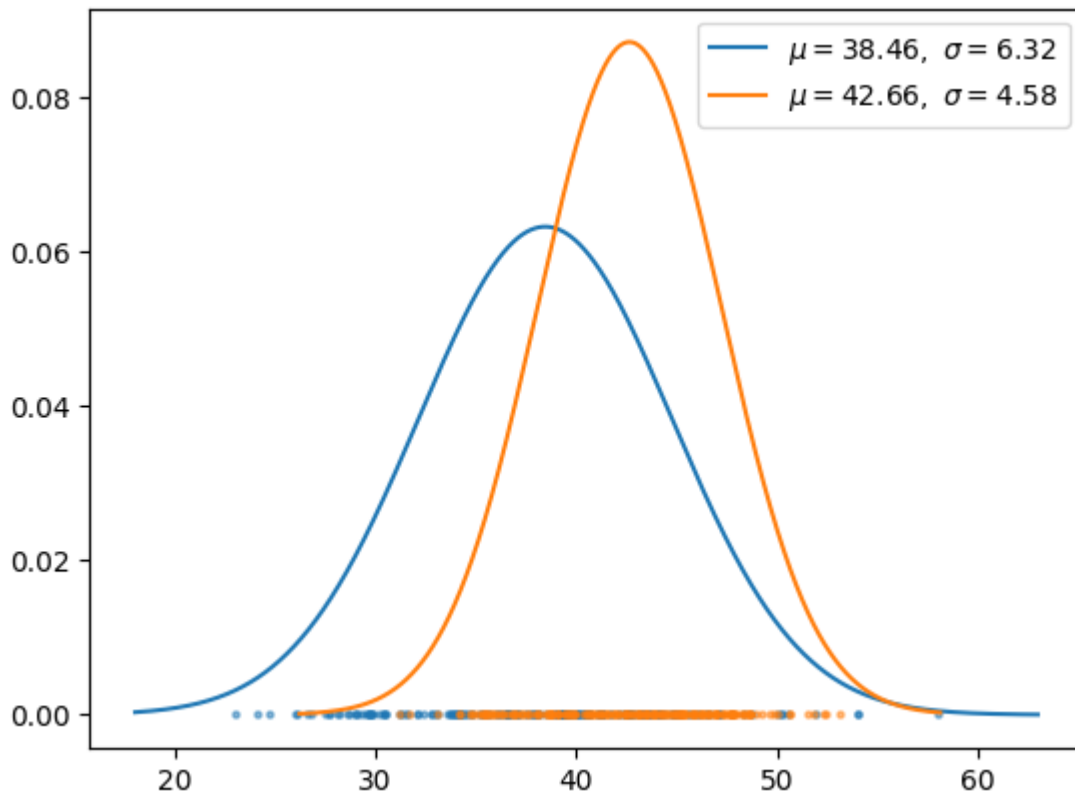
    heuristic_mu += HEURISTIC_STEP if heuristic else 0
    sigma = torch.tensor(random.uniform(SIGMA_START, SIGMA_END))
    normal_dist = torch.distributions.Normal(mu, sigma)
    sample = normal_dist.sample((n_observations, 1)).squeeze()
    gaussian_distributions.append(sample)
for distribution in gaussian_distributions:
    plot_observation(distribution)
if show:
    plt.legend()
    plt.show()
return gaussian_distributions

```

```

In [13]: plt.figure()
dataset = generate_data(200, show=False, heuristic=True)
plt.legend()
plt.show()
print(dataset)

```



```
[tensor([40.0973, 39.4247, 40.1122, 42.4325, 36.7756, 32.1311, 48.3529, 36.7557,
42.5862, 39.7117, 29.0037, 44.8239, 43.1669, 29.0770, 41.7910, 53.9897,
33.5972, 29.3193, 39.1221, 44.5079, 42.0217, 40.6877, 29.6682, 32.7780,
41.2338, 39.7538, 37.0846, 32.1090, 43.7242, 45.0002, 38.9886, 31.2561,
34.7561, 40.0426, 38.7902, 24.0729, 40.1048, 42.5734, 42.1284, 37.4647,
24.7083, 34.3932, 42.5168, 42.7289, 23.0259, 44.3069, 29.8323, 39.4567,
43.5241, 37.5675, 40.3557, 47.0306, 36.3094, 50.2619, 39.2355, 39.6107,
30.5091, 44.0632, 34.1547, 28.9910, 33.7371, 26.5671, 38.1134, 44.7620,
35.2382, 27.7134, 33.9015, 50.0032, 26.0685, 43.0462, 39.8573, 36.9930,
28.1020, 36.4979, 34.0300, 43.5469, 41.0486, 40.1708, 58.0116, 44.1577,
39.6291, 46.6742, 41.4126, 42.7254, 37.8765, 34.3034, 43.7800, 42.3314,
29.9138, 51.9334, 38.1004, 39.3672, 29.6318, 38.9954, 40.4020, 42.6939,
44.2603, 30.5418, 36.5470, 42.6023, 28.2442, 40.5630, 36.5578, 35.6233,
30.1589, 46.0059, 38.4346, 39.5107, 37.9538, 37.8053, 43.1825, 42.4567,
37.8204, 43.3430, 34.4583, 45.5109, 35.3889, 44.0467, 39.2058, 34.6078,
27.7445, 40.5606, 39.4518, 33.6630, 33.0729, 39.2260, 34.7617, 43.3417,
36.5716, 29.6909, 30.3923, 40.3820, 34.1036, 39.5285, 47.1182, 31.4598,
40.6396, 35.6993, 34.0271, 39.4070, 37.0345, 27.4985, 43.4175, 41.1972,
39.2301, 43.6017, 33.7442, 40.4110, 50.2283, 39.3264, 45.8371, 41.2516,
42.5444, 45.1952, 32.8030, 41.6965, 39.2833, 35.9864, 28.7190, 45.1755,
32.3955, 34.1720, 54.0496, 46.5623, 48.7495, 36.1789, 34.2262, 29.5253,
28.5769, 29.7454, 36.0517, 46.4243, 41.0759, 44.7345, 39.0480, 37.2109,
46.1662, 25.9960, 29.6916, 45.6427, 43.2312, 44.1070, 36.5449, 30.3471,
35.1751, 26.8354, 41.2719, 31.5593, 47.7225, 38.1847, 40.6263, 42.8558,
31.2003, 47.1741, 36.4518, 44.9960, 42.4885, 50.5912, 36.5439, 33.0235]), ten
sor([42.5309, 44.6570, 43.1648, 40.6364, 39.9287, 34.1384, 46.9272, 44.2545,
36.8895, 45.2306, 43.7496, 46.8387, 46.6801, 43.4845, 35.4643, 41.1480,
38.8032, 45.4214, 43.4406, 45.1660, 39.4152, 38.0620, 40.6106, 46.6351,
45.3243, 41.8828, 44.5648, 43.1299, 38.0099, 46.9901, 48.6889, 43.5257,
34.8329, 49.6799, 40.7350, 40.5485, 45.7261, 37.4297, 35.4270, 44.7262,
48.6300, 43.4898, 36.5769, 42.0661, 35.5667, 47.0608, 41.3746, 45.5682,
38.5499, 48.5745, 50.6884, 47.7701, 47.8091, 42.0470, 47.2957, 36.4114,
35.4885, 41.8534, 43.9367, 47.6500, 46.0641, 48.3252, 42.4945, 43.8856,
38.8087, 45.8360, 46.2975, 46.2357, 38.1499, 42.4434, 46.3470, 46.5563,
45.3127, 34.1740, 42.1558, 36.2185, 42.7876, 35.8188, 41.5557, 47.6877,
44.8136, 44.2412, 35.3183, 48.1250, 34.6748, 37.2538, 50.6491, 35.5139,
45.0524, 45.9217, 43.0864, 38.6944, 37.0352, 43.2052, 48.3739, 51.8810,
46.5464, 35.6323, 47.9001, 43.1153, 39.8453, 42.4983, 51.5894, 41.1045,
49.2382, 43.9279, 45.2167, 35.9617, 46.1626, 41.3152, 42.6747, 48.6500,
40.7504, 41.1241, 45.2102, 41.3176, 44.3977, 47.7186, 49.8582, 42.9421,
38.5650, 41.6114, 45.1760, 36.9484, 47.7360, 53.1588, 50.0016, 41.8978,
45.8149, 47.3151, 39.0228, 44.1616, 35.9856, 43.8370, 44.9403, 44.4296,
37.6684, 33.0617, 52.4062, 41.5298, 40.8451, 39.0995, 45.7193, 46.9748,
47.6466, 37.6154, 44.8770, 36.7684, 52.2993, 46.4246, 42.3816, 43.3380,
38.1187, 39.5620, 44.9421, 36.9111, 31.7233, 39.5017, 36.2817, 45.8690,
31.2188, 37.3256, 48.6723, 38.6506, 45.3565, 39.0429, 43.6957, 35.3109,
45.1643, 42.4209, 39.6769, 39.4325, 38.7883, 38.4628, 39.8661, 39.6408,
43.4364, 46.9331, 39.6362, 34.7542, 40.5132, 47.9625, 35.7739, 40.9259,
48.9576, 44.1535, 44.4575, 39.0871, 40.3216, 43.8710, 47.4880, 39.9133,
47.3098, 42.7788, 48.3280, 35.0338, 46.2769, 38.7616, 42.9644, 47.8799)])]
```

2. Función init_random_parameters:

```
In [14]: # Genera una matriz k x 2 con mu y sigma aleatorios
def init_random_parameters(k_parameters=2):
    p_matrix = []
```

```

for k in range(k_parameters):
    mu = torch.tensor(random.uniform(MU_START, MU_END))
    sigma = torch.tensor(random.uniform(SIGMA_START, SIGMA_END))
    p_matrix.append([mu, sigma])
p_matrix = torch.tensor(p_matrix)
return p_matrix

```

```

In [15]: random_parameters = init_random_parameters()
print(random_parameters)

```

```

tensor([[48.3876,  3.9910],
        [25.6663,  3.1677]])

```

3. Función calculate_likelihood_gaussian_observation:

```

In [16]: # Calcula la verosimilitud de una observación con respecto a un mu y sigma
# Se utiliza la función de densidad de probabilidad gaussiana
def calculate_likelihood_gaussian_observation(x_n, mu_k, sigma_k):
    def probability_density_function(x, mu, sigma):
        return (1/math.sqrt(2 * math.pi * sigma**2)) * math.e**(-(1/2) * ((x-mu) / sigma)**2)
    return probability_density_function(x_n, mu_k, sigma_k)

```

```

In [17]: calculate_likelihood_gaussian_observation(42.3022, 28.0083, 1.5705)

```

```

Out[17]: 2.612131073896484e-19

```

4. Función calculate_membership_dataset:

```

In [18]: # Calcula la pertenencia de cada observación a los parámetros
# Se utiliza one hot vector
def calculate_membership_dataset(x_dataset, parameters_matrix):
    likelihood_matrix = []
    for dataset in x_dataset:
        for data in dataset:
            data_likelihood = []
            for matrix in parameters_matrix:
                mu = matrix[0]
                sigma = matrix[1]
                likelihood = calculate_likelihood_gaussian_observation(data.item(), mu, sigma)
                data_likelihood.append(likelihood)
            for index in range(len(data_likelihood)):
                data_likelihood[index] = 0 if data_likelihood[index] != max(data_likelihood) else 1
            likelihood_matrix.append(data_likelihood)
    likelihood_matrix = torch.tensor(likelihood_matrix)
    return likelihood_matrix

```

```

In [19]: membership_matrix = calculate_membership_dataset(dataset, random_parameters)
print(membership_matrix)

```

[illegible]

[illegible]

$[0, 1],$
 $[1, 0],$
 $[0, 1],$
 $[1, 0],$
 $[1, 0],$
 $[0, 1],$
 $[0, 1],$
 $[1, 0],$
 $[1, 0],$
 $[0, 1],$
 $[0, 1],$
 $[1, 0],$
 $[0, 1],$
 $[1, 0],$
 $[1, 0],$
 $[0, 1],$
 $[0, 1],$
 $[1, 0],$
 $[0, 1],$
 $[1, 0],$
 $[1, 0],$
 $[0, 1],$
 $[1, 0],$
 $[0, 1],$
 $[0, 1],$
 $[1, 0],$
 $[1, 0],$
 $[0, 1],$
 $[1, 0],$
 $[1, 0],$
 $[1, 0],$
 $[1, 0],$
 $[0, 1],$
 $[1, 0],$
 $[1, 0],$
 $[1, 0],$
 $[0, 1],$
 $[1, 0],$
 $[0, 1],$
 $[0, 1],$
 $[0, 1],$
 $[0, 1],$
 $[0, 1],$
 $[1, 0],$

[illegible]

[illegible]

[illegible]

[illegible]

```
[1, 0],
[1, 0]])
```

```
In [20]: # Se realiza el recálculo de Los parámetros
def recalculate_parameters(x_dataset, membership_data):
    membership_data = torch.transpose(membership_data, 0, 1)
    complete_dataset = torch.Tensor()
    new_parameters = []
    for dataset in x_dataset:
        complete_dataset = torch.cat((complete_dataset, dataset))
    for k in membership_data:
        data_set_one = []
        for one_hot_data in range(len(k)):
            if k[one_hot_data].item() == 1:
                data_set_one.append(complete_dataset[one_hot_data])
        data_set_one = torch.Tensor(data_set_one)
        mu = torch.mean(data_set_one)
        sigma = torch.std(data_set_one, unbiased=True)
        # Heurístico: En caso de que ninguna distribución tenga membresía con Los pará
        if mu.item() != mu.item() or sigma.item() != sigma.item(): # if nan
            params = init_random_parameters(1)
            mu = params[0][0]
            sigma = params[0][1]
            new_parameters.append([mu.item(), sigma.item()])
        else:
            new_parameters.append([mu.item(), sigma.item()])
    new_parameters = torch.Tensor(new_parameters)
    return new_parameters
```

```
In [21]: membership_matrix = calculate_membership_dataset(dataset, random_parameters)
recalculate_parameters(dataset, membership_matrix)
```

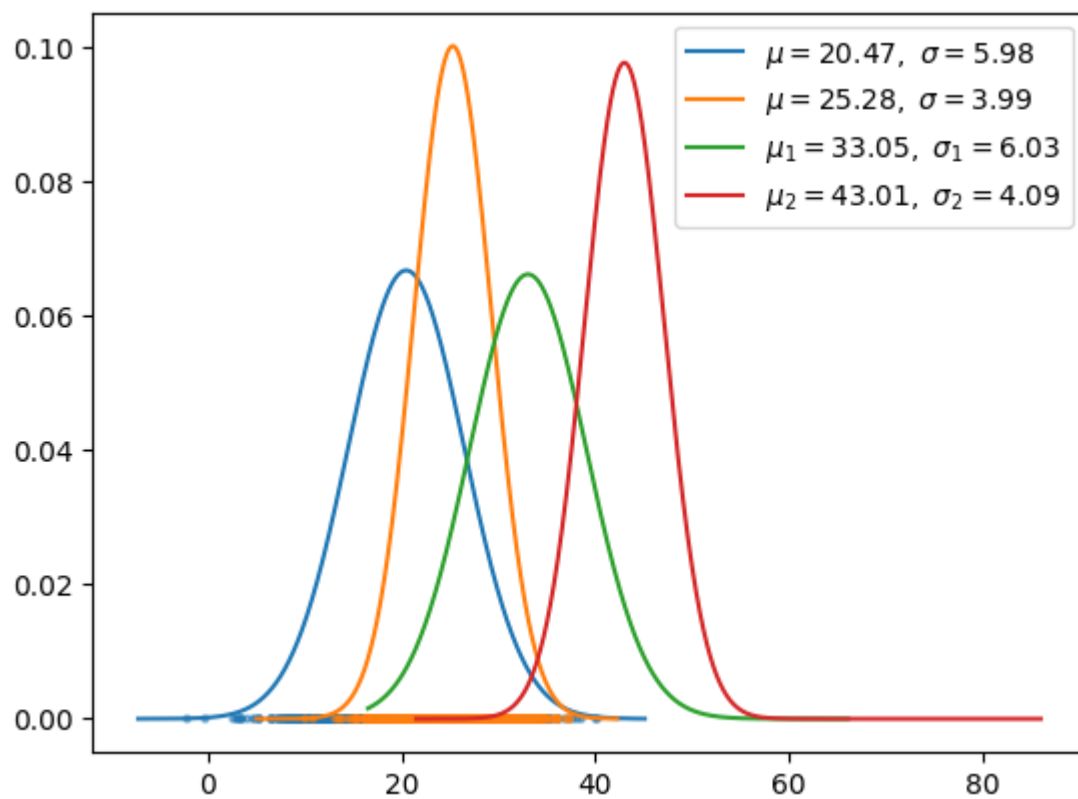
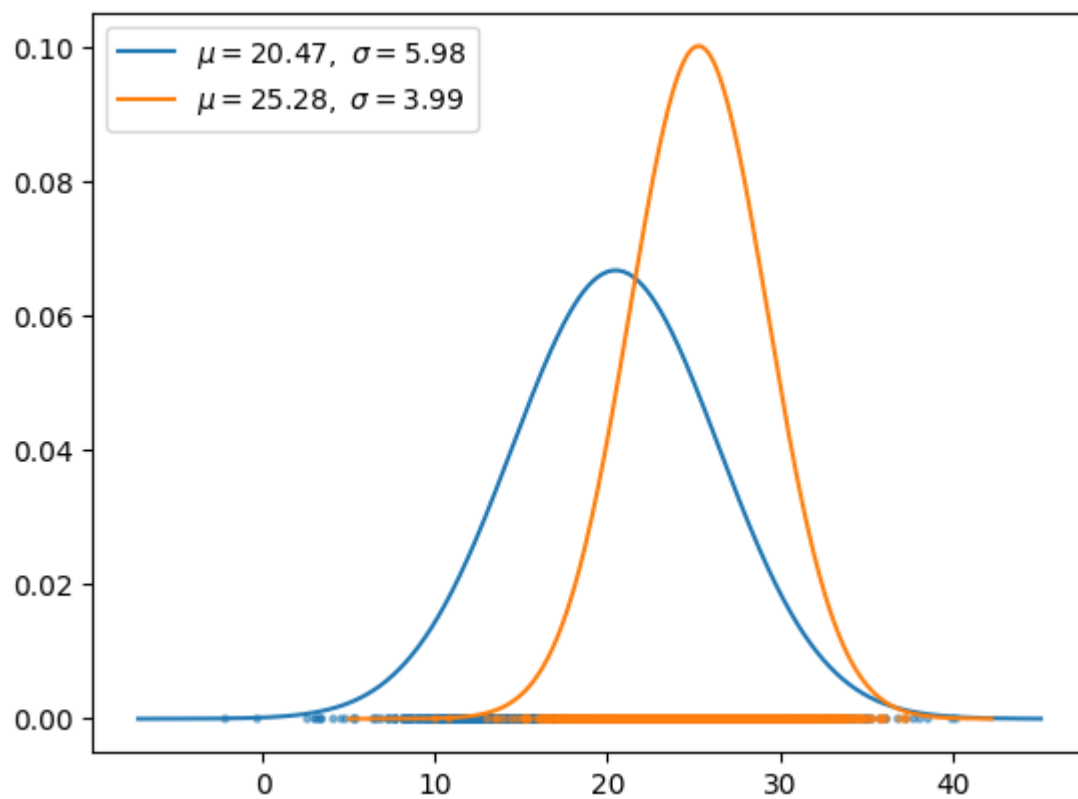
```
Out[21]: tensor([[42.8157,  4.0755],
                [31.9324,  3.1895]])
```

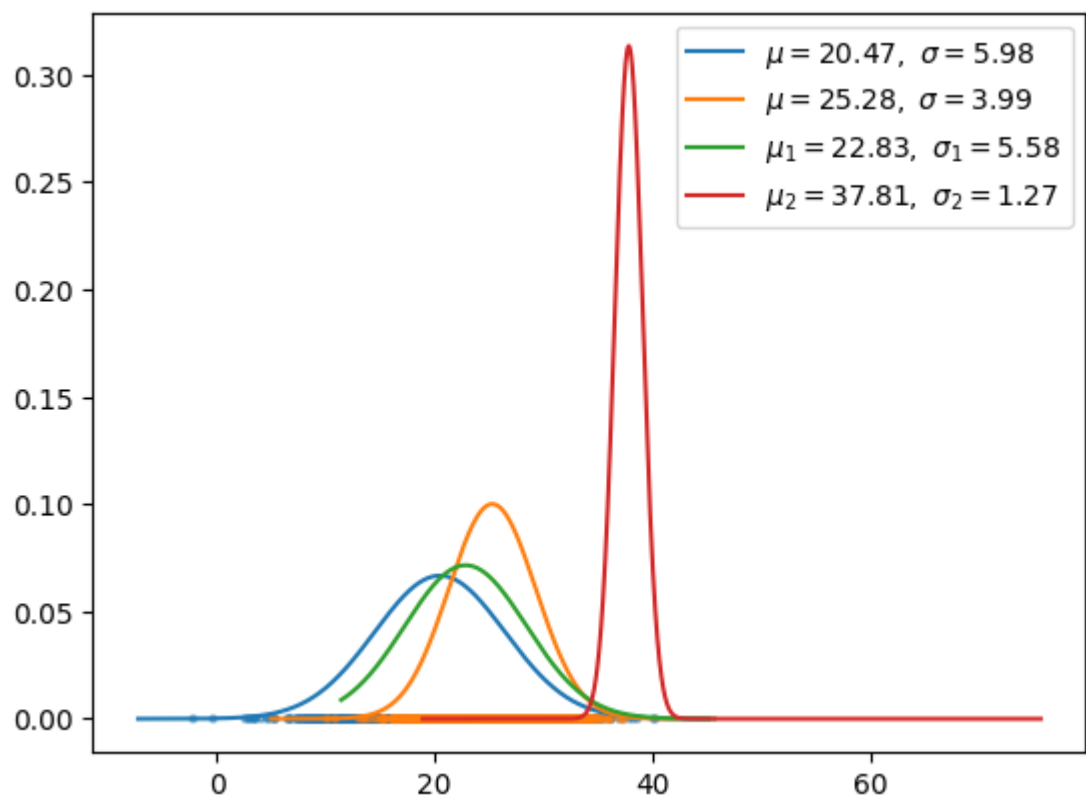
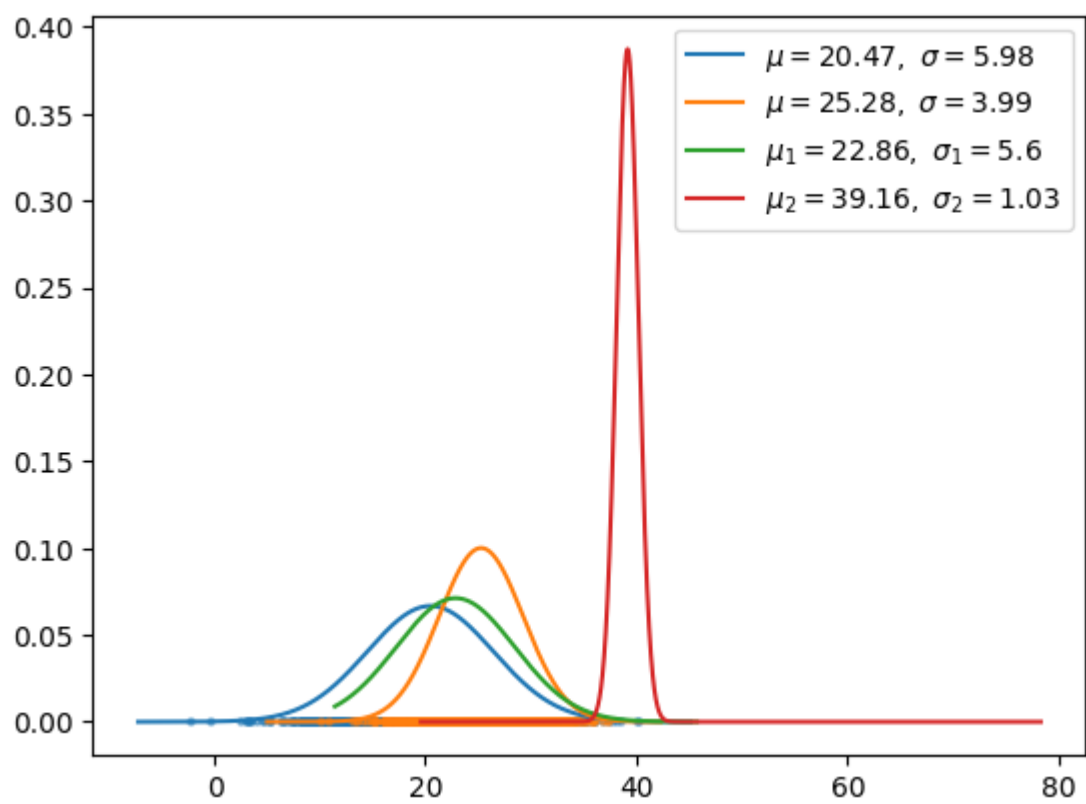
```
In [27]: # Algoritmo de maximización de la esperanza junto
def expectation_maximization(observations=200, k_parameters=2, iterations=5, heuristic
my_data = generate_data(observations, k_parameters, show=True, heuristic=heuristic)
parameters = init_random_parameters(k_parameters)
plot_gaussian_distribution_and_observations(parameters, my_data, show=True)
for iteration in range(iterations):
    membership_data = calculate_membership_dataset(my_data, parameters)
    parameters = recalculate_parameters(my_data, membership_data)
    plot_gaussian_distribution_and_observations(parameters, my_data, show=True)
```

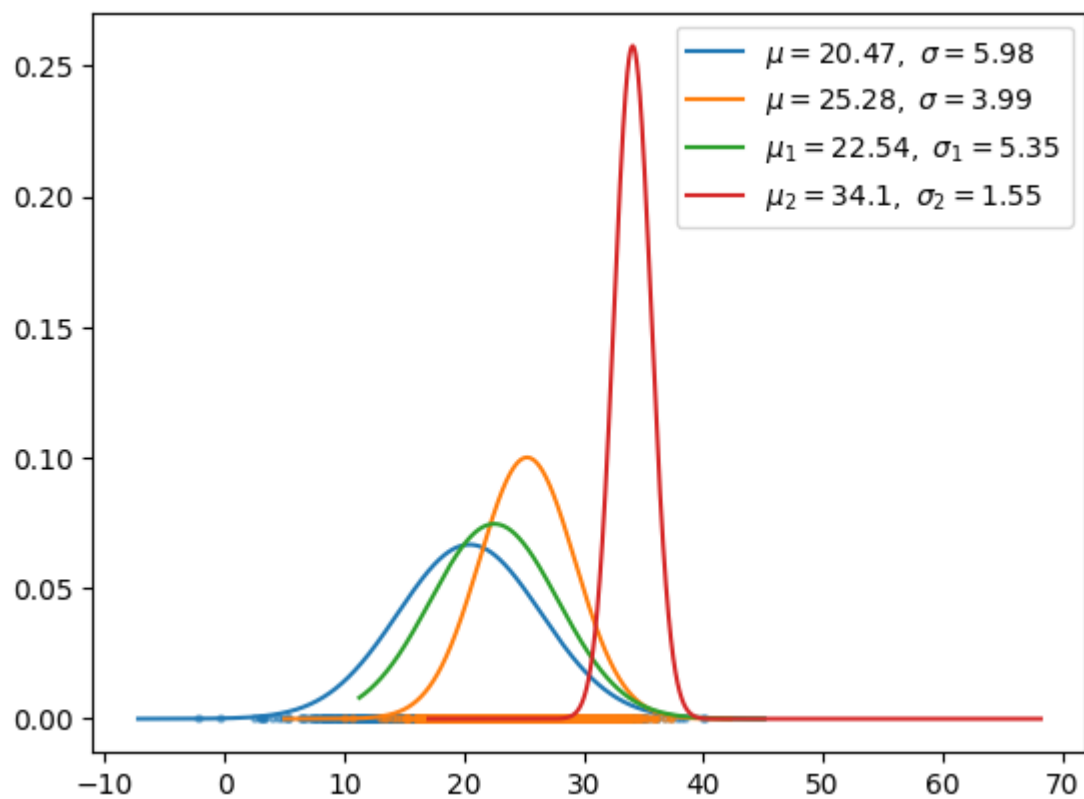
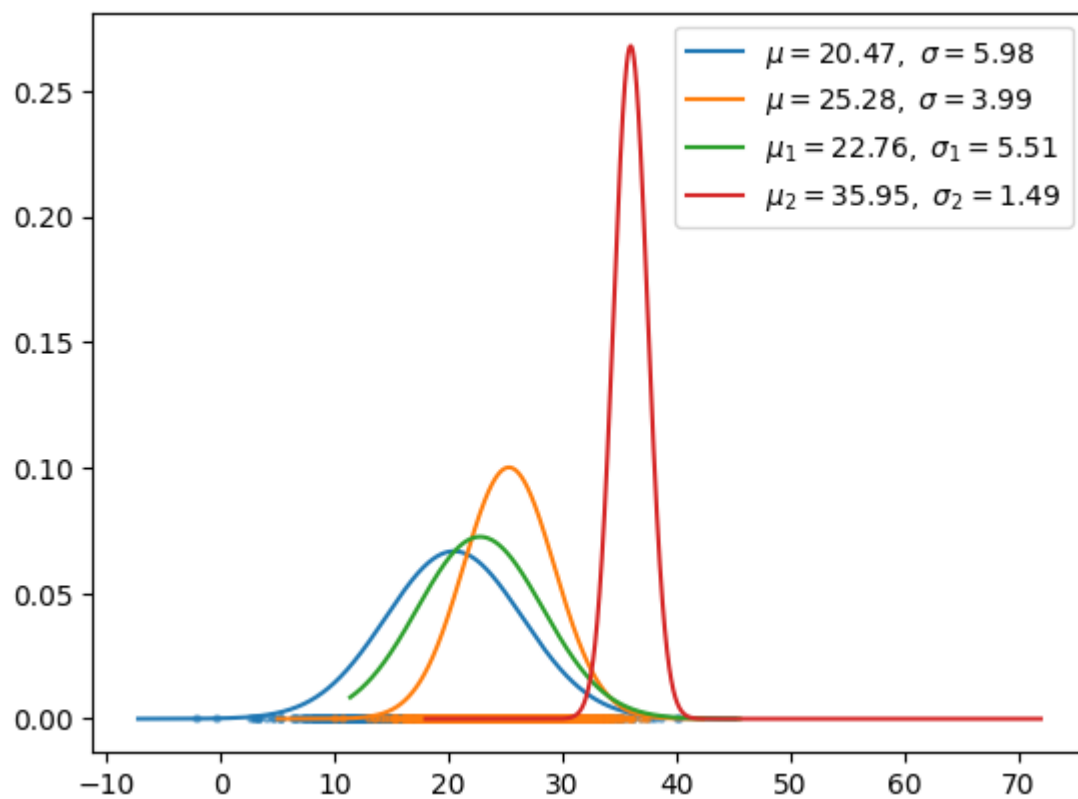
6. Ejecuta 5 corridas, donde por cada una documento los parámetros a los que se arribó.

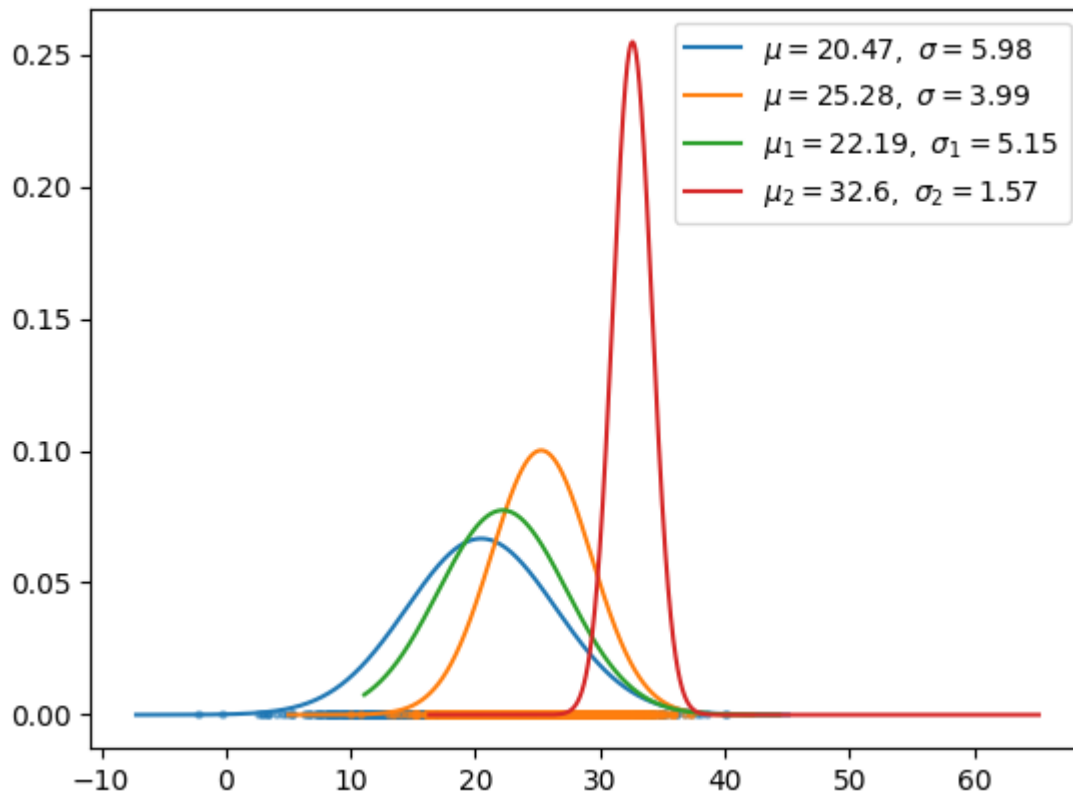
A continuación se muestran las corridas, los parámetros a los que se arribó se imprimen en los gráficos.

```
In [28]: expectation_maximization(observations=2000, heuristic=True, k_parameters=2, iterations
```









7. Proponga una mejor heurística para inicializar los parámetros del modelo aleatoriamente.

La mejora heurística consiste en inicializar las observaciones de manera que algunos de sus datos se mezclen. Esto se logra inicializando μ 's que tengan una diferencia pequeña definida por una constante que hemos denominado "HEURISTIC_STEP". Del mismo modo, se puede aumentar la cantidad de observaciones y el sigma se puede hacer más grande.

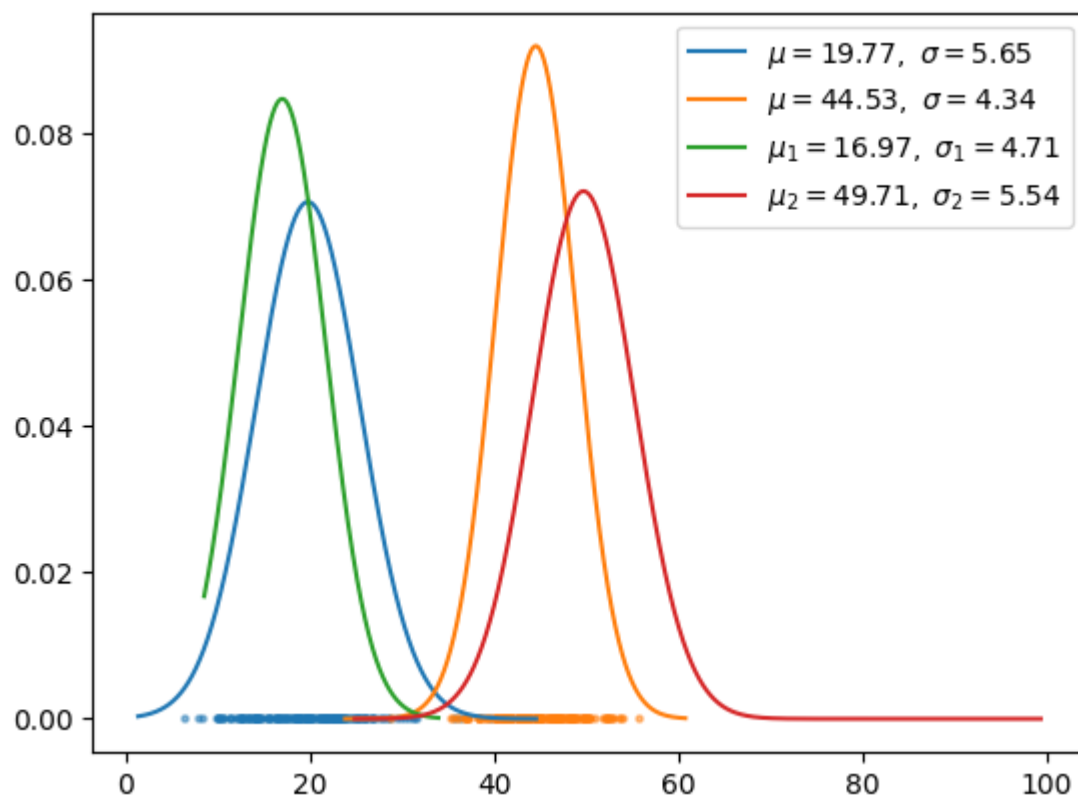
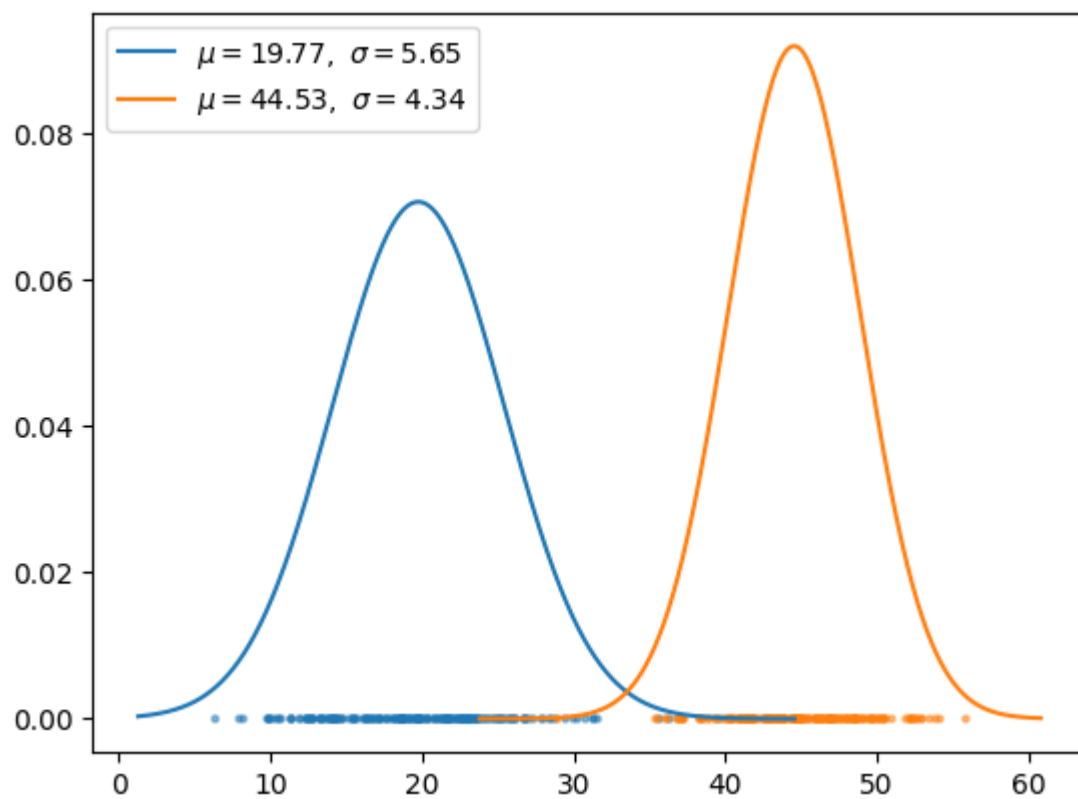
También, en caso de que la membresía de todas las observaciones se ajusten a un y solo un set de parámetros, se implementó la reinicialización aleatoria de los parámetros para aquel set que no tuvo ninguna membresía; esto fue colocado en la función `recalculate_parameters`.

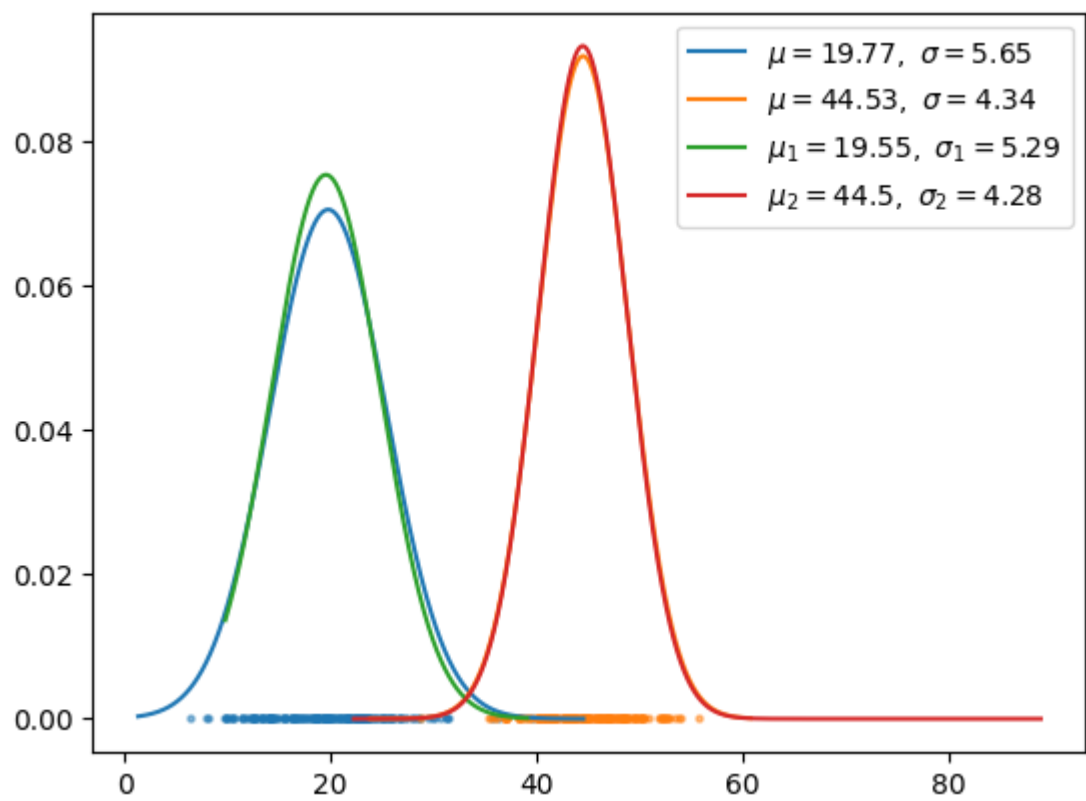
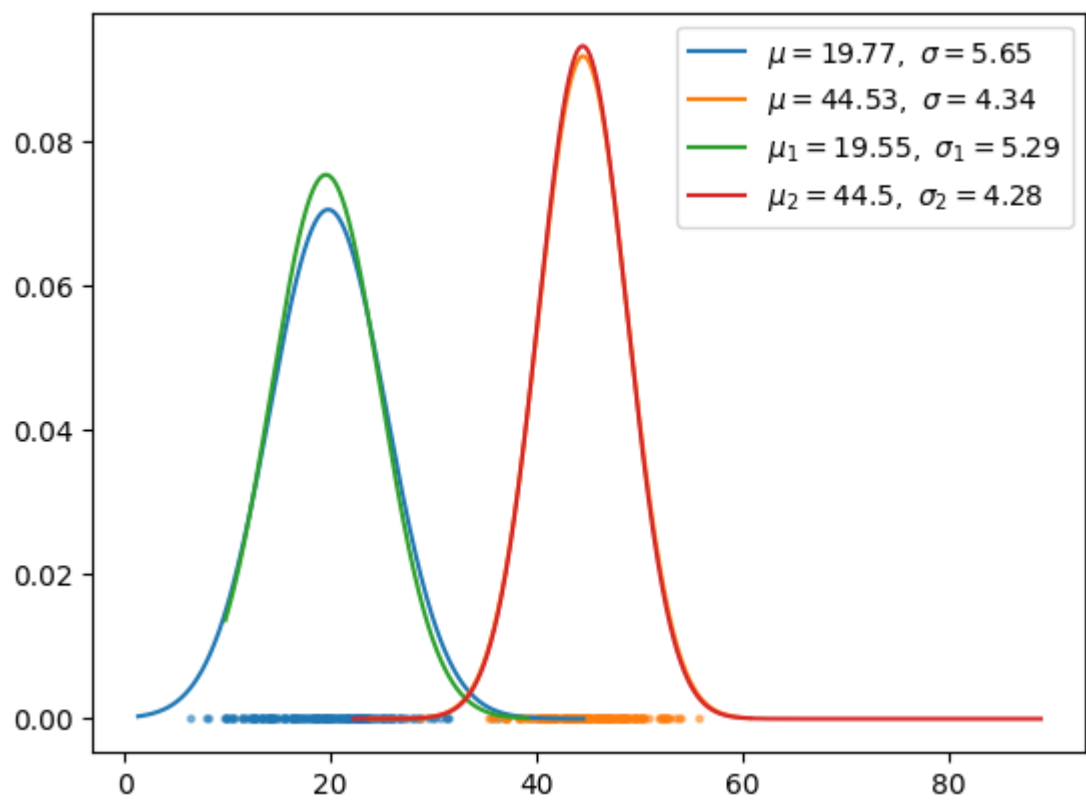
Sin esta mejora, el algoritmo tiende mucho a quedarse atascado después de la primera iteración, no sucede el 100% de las veces, pero tiene tendencia a suceder bastante.

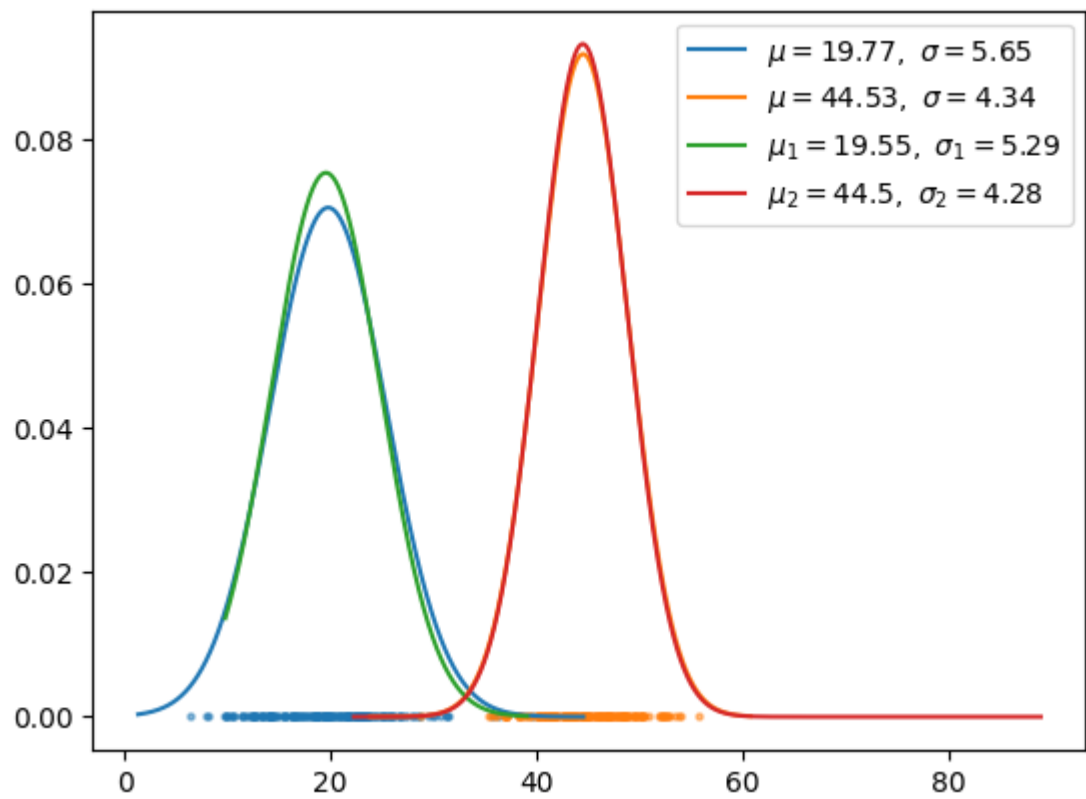
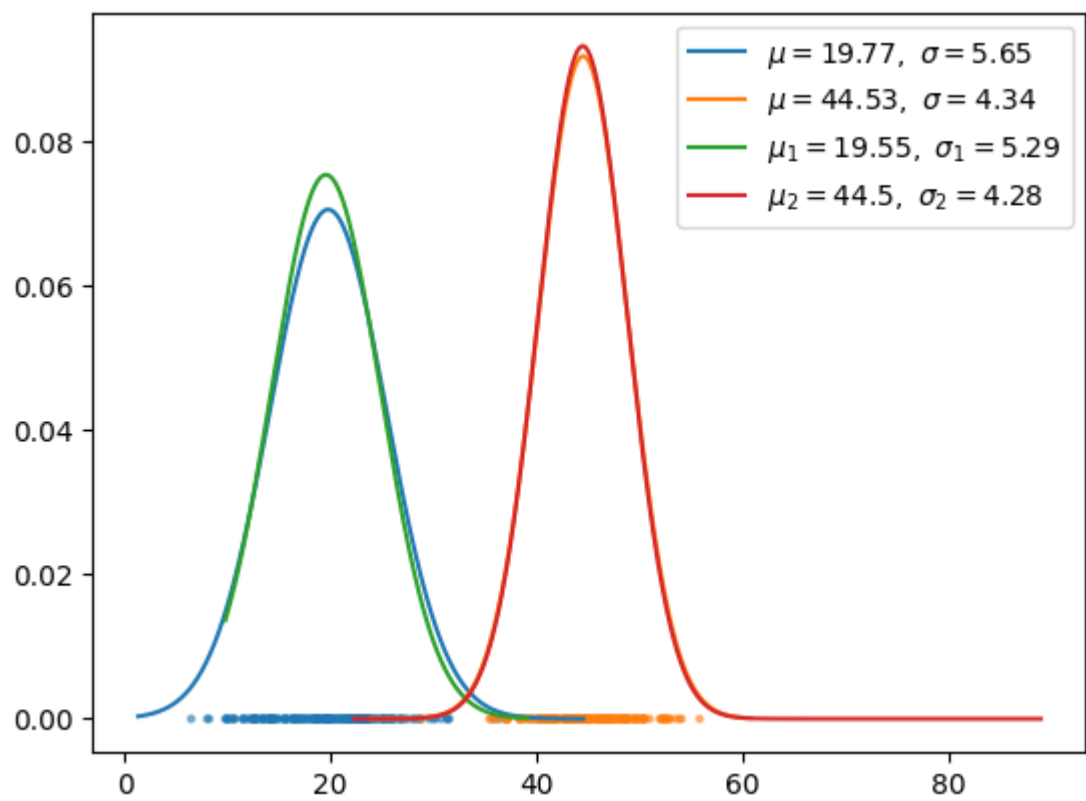
Se pone un ejemplo para ejecutar con el parámetro `heuristic=False`, en caso de que no suceda el comportamiento anteriormente descrito (lo cual es poco probable), se adjunta un ejemplo al final donde se vé que sí sucede:

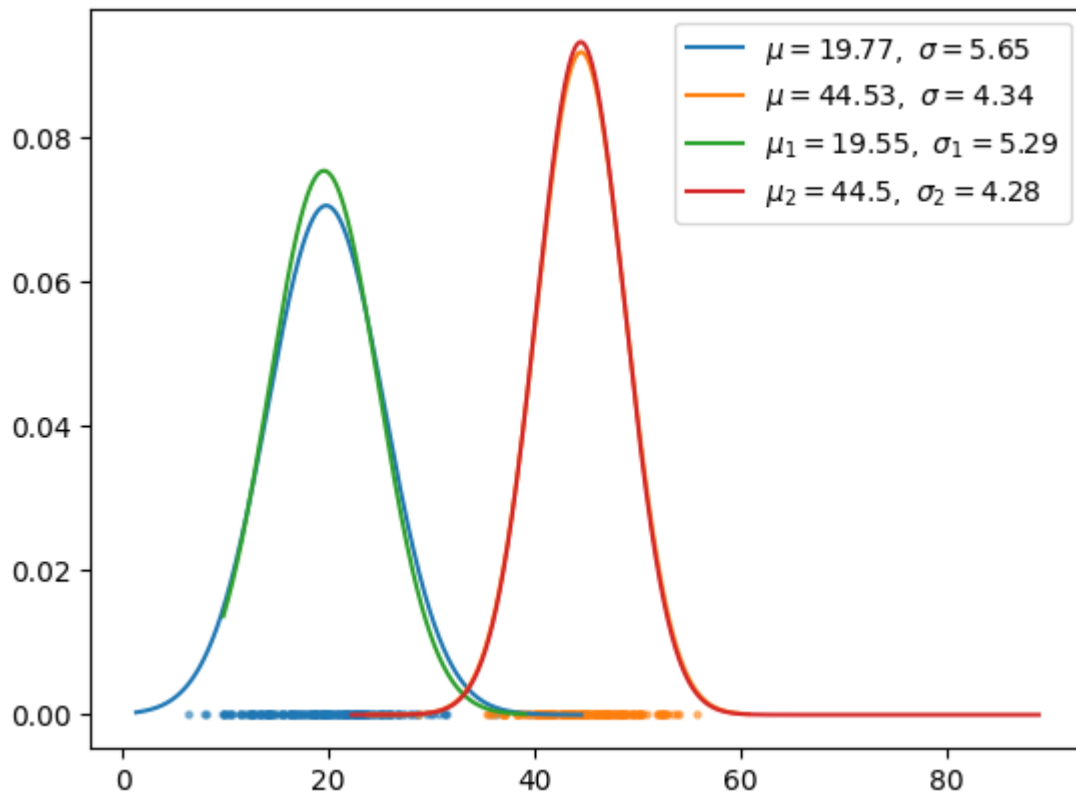
(Se sigue tomando en cuenta la reinicialización aleatoria, porque de lo contrario podría causar errores)

In [49]: `expectation_maximization(observations=200, heuristic=False, k_parameters=2, iterations`



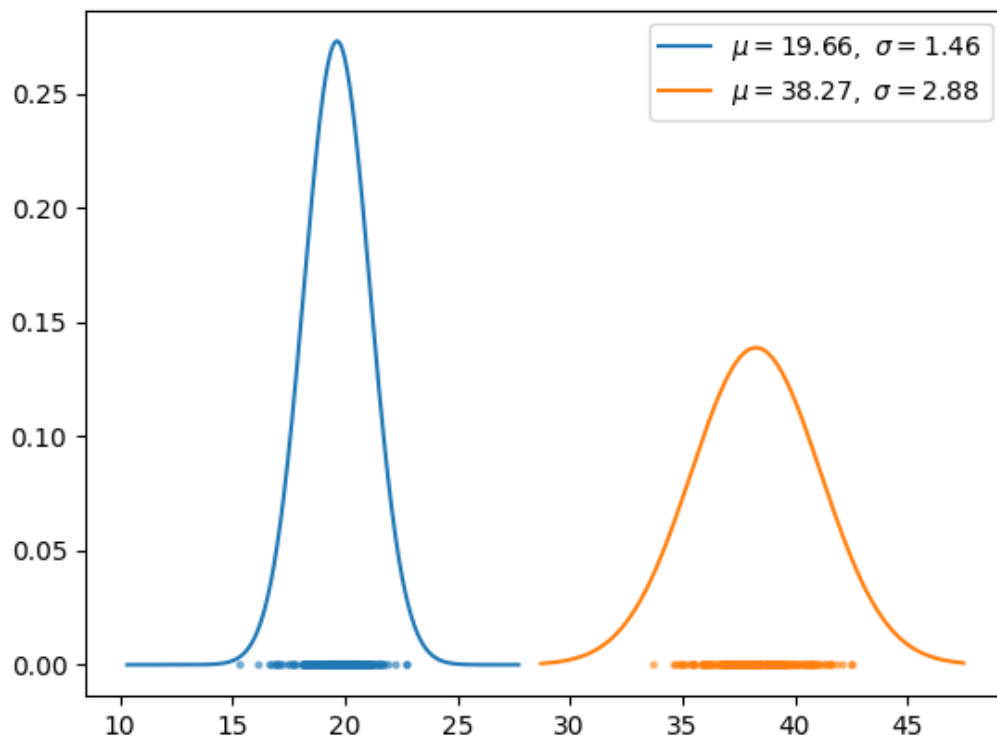


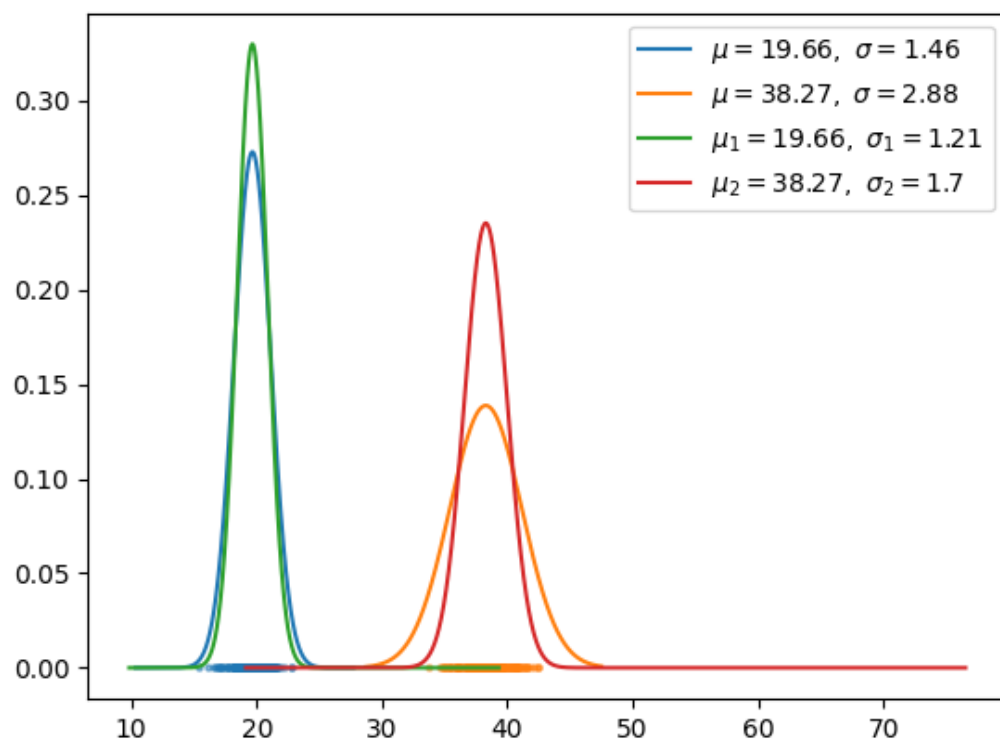
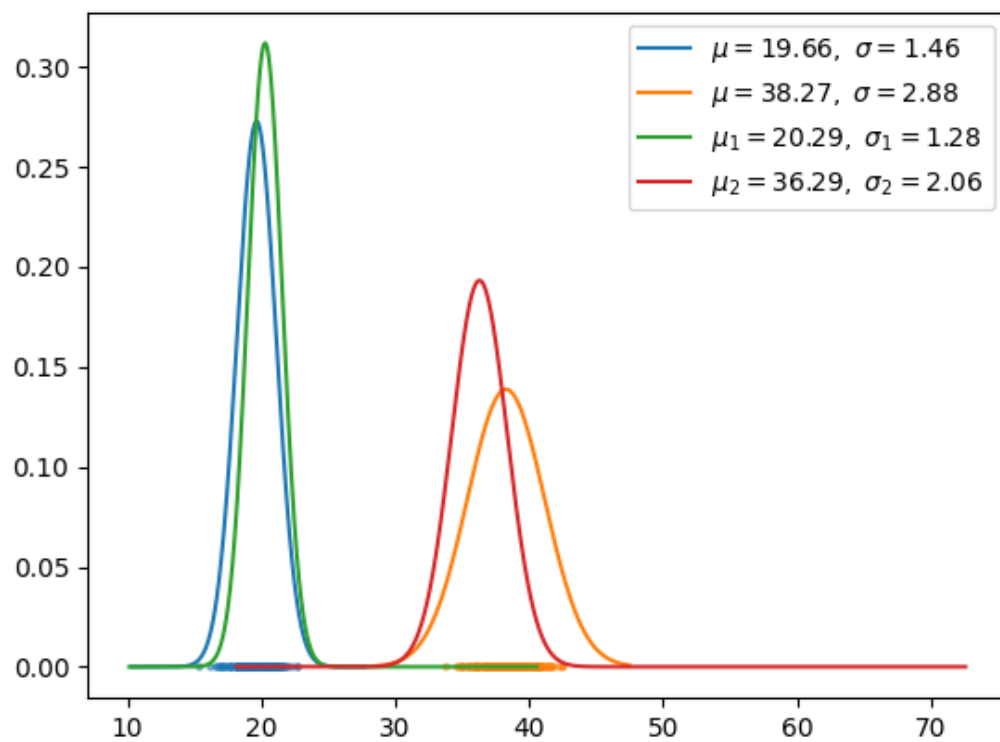


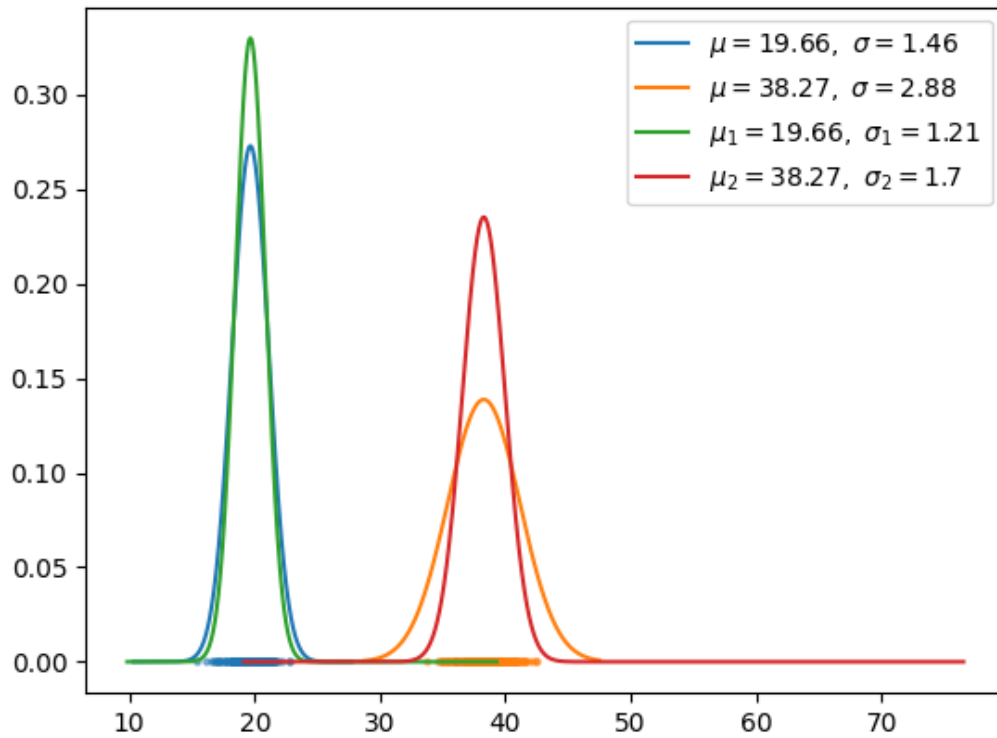


Ejemplo de imágenes:

IN [] : expectation_maximization(observations=200, heuristic=False, k_parameters=2, iterations=5)







Como se puede ver, después de la primera interacción, los parámetros no volvieron a cambiar y así se mantiene hasta el final.

Con la mejora, esto no ha sucedido, por lo que nuestra propuesta tiene potencial para mejorar la inicialización de los parámetros.