# Ciencia de Datos
# Tarea 4

Para entregar el 21 de abril de 2021

1. En el archivo `data_world.csv` se encuentran datos en 2 dimensiones con 5 grupos que se muestran en la parte superior de la Figura 1.

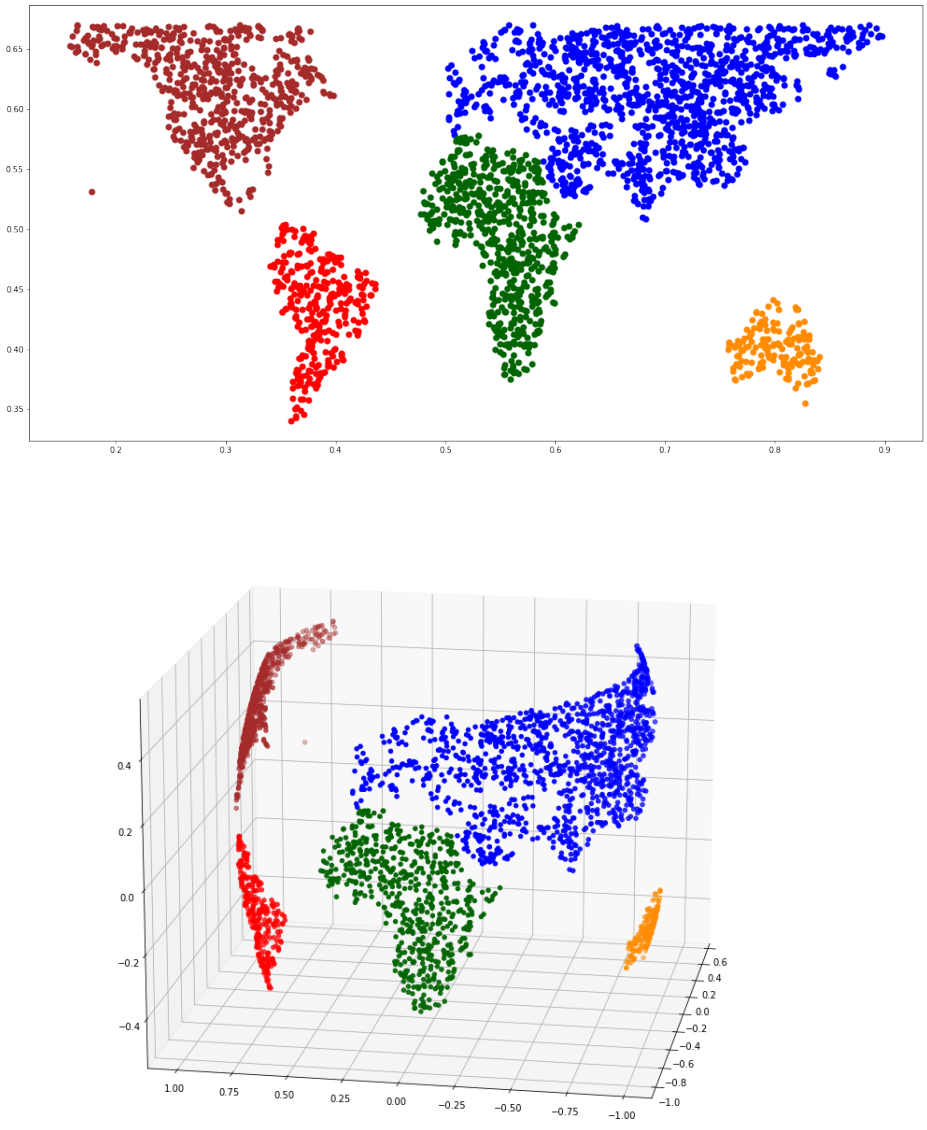


Figura 1: Visualización de los datos para el ejercicio 1. Superior: datos en 2 dimensiones. Inferior: embedding en un 3D manifold.

Ahora considera el embedding de los datos en una variedad no lineal en 3D como se muestra en la parte inferior de la misma Figura 1, y que puedes obtener con

el siguiente código (donde `X` contiene las coordenadas de los datos y `y` el color del clúster):

```
p = X[:, 0] * (2 * np.pi - 0.55)
t = X[:, 1] * np.pi

x_sphere = np.sin(t) * np.cos(p)
y_sphere = np.sin(t) * np.sin(p)
z_sphere = np.cos(t)

X_sphere = np.array([x_sphere, y_sphere, z_sphere]).T
X_sphere.shape

from mpl_toolkits import mplot3d
plt.figure(figsize=(20,15))
ax = plt.axes(projection = '3d')
ax.view_init(16, -170)
ax.scatter3D(X_sphere[:, 0],
             X_sphere[:, 1], -X_sphere[:, 2], c = y)
plt.show()
```

*a*) Aplica métodos de manifold learning basados en PCA, Kernel PCA, Spectral Embeddings y $t-$SNE para tratar de reconstruir los patrones encontrados en los datos 2D. ¿Cuál método prefieres y porqué? Documenta todos tus hallazgos incluyendo las parametrizaciones que usaste.

2. Considera nuevamente los datos LFW que usaste en el ejercicio 3 de la tarea 2, los cuales representan fotografías de rostros recolectados de internet. Usa los mismos criterios para la selección de rostros, pero sin crear conjuntos de entrenamiento y prueba, es decir, en este ejercicio usarás las 1288 imágenes que corresponden a 7 personas.

   Aplica los métodos de manifold learning basados en Kernel PCA, Spectral Embeddings y $t-$SNE para obtener representaciones en 2D de los rostros. Compáralos con los resultados que obtuviste con PCA en la tarea 2 ¿Qué método te parece más adecuado y porqué? Documenta todos tus hallazgos incluyendo las parametrizaciones que usaste.