

Estados similares

Guillermo Ruiz

En esta tarea se proporcionan vectores para cada uno de los estados de la república. Los vectores tienen que ver con el vocabulario que se usa en cada uno. La idea es ver qué estados escriben de manera similar. Lo que se debe hacer es:

1. Leer el conjunto de datos en el archivo `data.npy`. Cada vector corresponde a un estado, el orden de los vectores corresponde con el orden alfabético de los estados.
2. Usar umap para reducir la dimensión a 3. Usar la configuración `umap.UMAP(n_components=3, n_neighbors=3, min_dist=.2, metric='cosine')` La reducción se interpretará como un color en la escala RGB. La idea es que colores similares indican que los estados escriben de manera similar.
3. Reescalar los vectores reducidos con umap para que estén entre 0 y 1 mediante el método `MinMaxScaler` de sklearn.
4. Colorear cada estado con su color correspondiente. La salida debe ser algo parecido al mostrado.
5. Interpretar el resultado.

