README



Código base para la tarea 5

En la tarea 5 se utilizan cuadernos de Jupyter para explorar varios algoritmos de reducción de dimensiones, de aglomeración, y de clasificación generativa.

Se utilizan dos casos de evaluación de medicina.

Integrantes

- Carlos Andrés Mata Calderón
- David Badilla Marin

Dependencias

Para micromamba, conda, o mamba

Se tiene que crear un entorno llamado t5.

```
micromamba create -n t5 -y

micromamba install -c conda-forge \
matplotlib numpy \
scikit-learn jupyter pillow \
joblib scikit-learn-intelex \
pandas \
umap-learn \
-y
```

Utilización del código

Datos de cancer.ipynb

- Abrir la terminal en el directorio del proyecto.
- Ejecutar micromamba activate t5.
- Ejecutar jupyter-notebook ./cancer.ipynb
- Seleccionar Run>Run All Cells

Datos de heart.ipynb

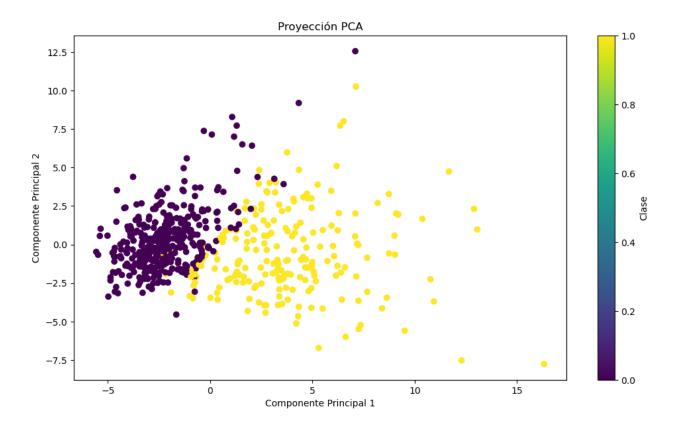
- Abrir la terminal en el directorio del proyecto.
- Ejecutar micromamba activate t5.
- Ejecutar jupyter-notebook ./heart.ipynb
- Seleccionar Run>Run All Cells

PDF con las gráficas

Se creo un documento pdf del README, ya que el readme también tiene el resultados y Análisis. Este pdf se encuentra en la carpeta docs/solution.pdf . O bien, puede presionar este enlace con enlace solution.pdf

Resultados y Análisis

Resultados de cancer.ipynb

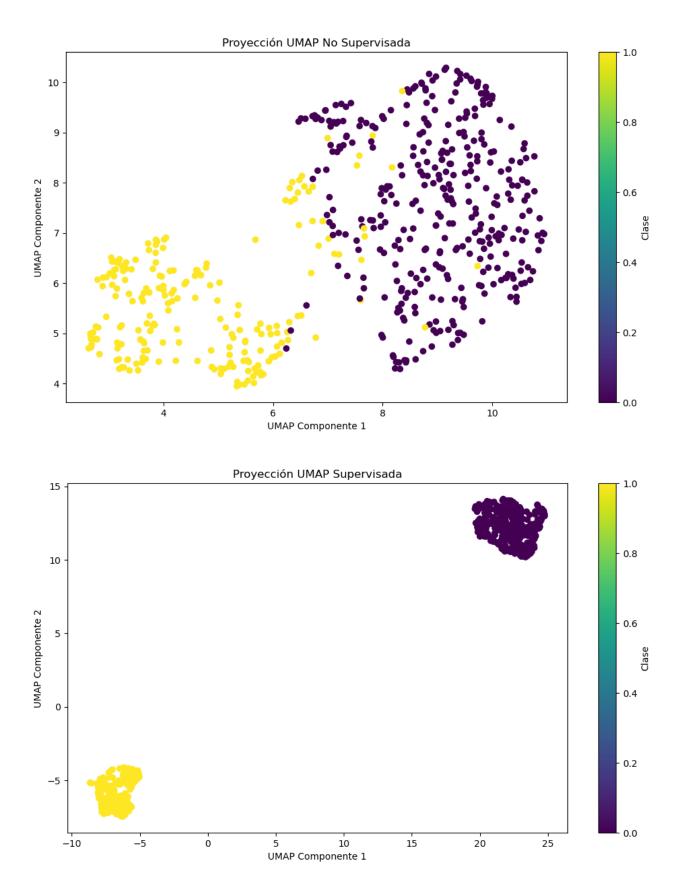


¿Es posible identificar si hay ejes en el espacio original más importantes que otros?

Si claro note que la componente 1 es la que tiene tiene mayor varianza. Y si obtenemos los valores de las componentes sus eigenvector, notaremos que hay unos valores más altos que otros. Recuerde que cada componente del eigenvector corresponde a un tipo de "peso" de esa característica en el espacio original. Por lo quiere decir que si tiene mayor "peso" esa componente del eigenvector quiere decir esa característica aporta mucho en la varianza de los datos. Lo cuál quiere decir que es una característica importante para lograr identificar la clase.

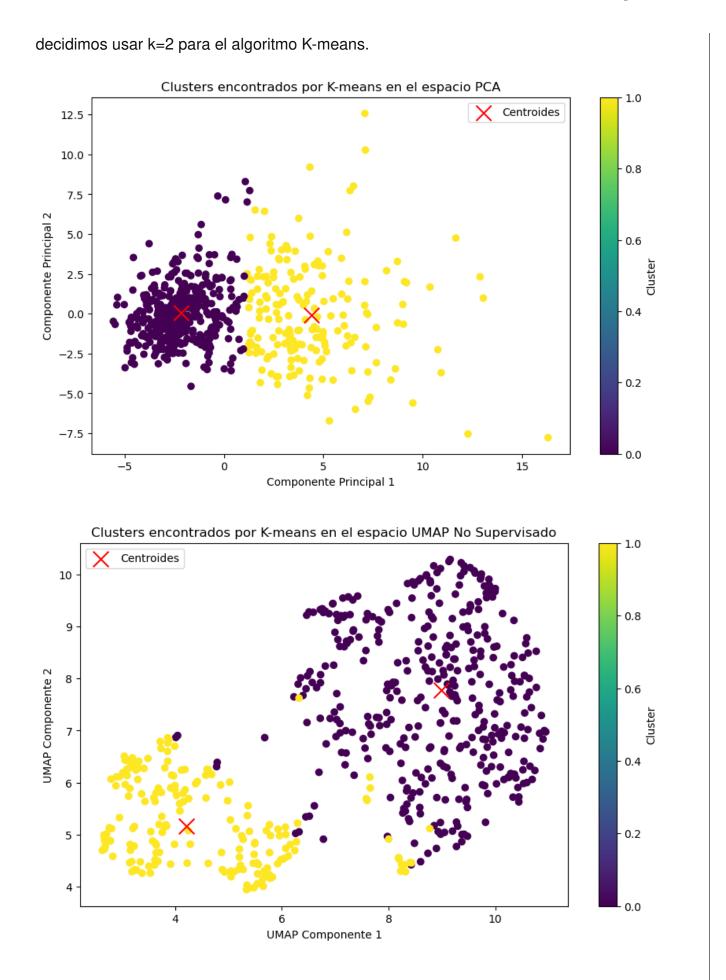
¿Es posible visualizar conglomerados en el subespacio bidimensional encontrado con PCA?

Si note que en el gráfico anterior como se encuentra una conglomerado en la parte izquierda superior de cáncer benigno, y un conglomerado en la parte derecha superior de cáncer maligno.

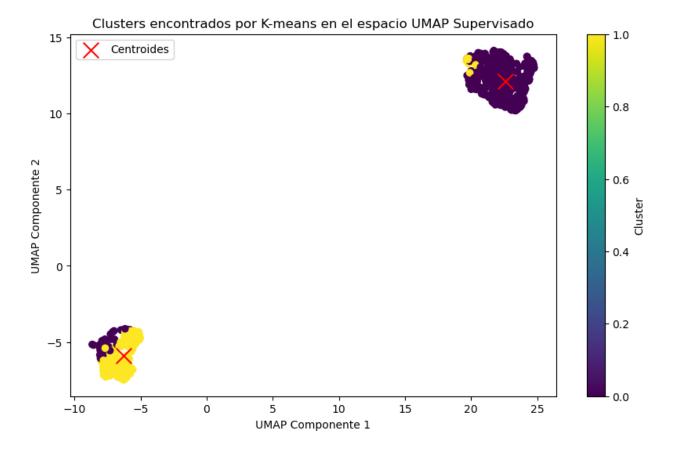


De los experimentos anteriores, analice si se pueden observar conglomerados aparentes, y qué número de ellos podría ser interesante evaluar.

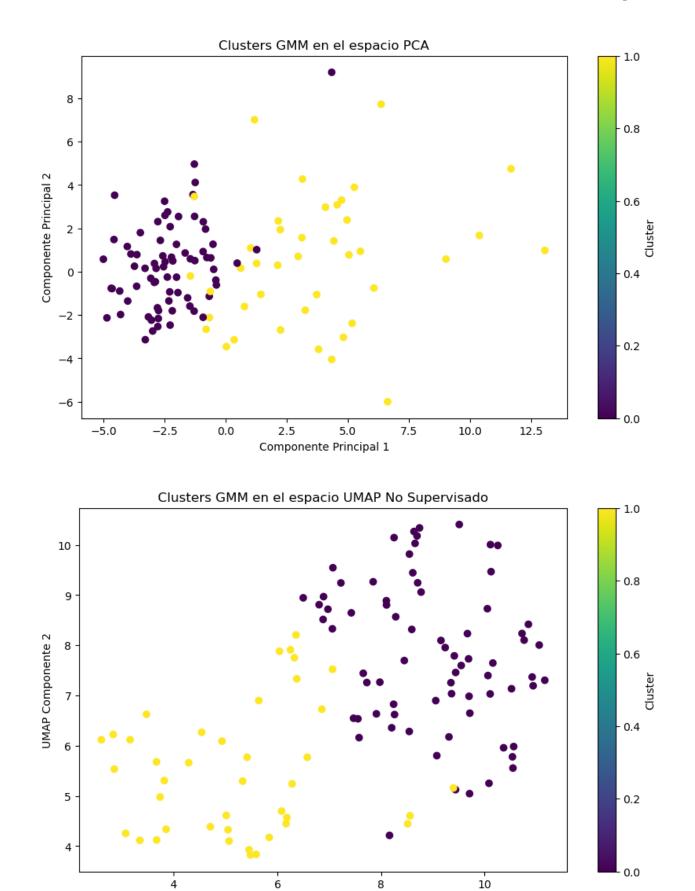
Como consistentemente vemos dos conglomerados principales en los gráficos,



La separación de clases mejora significativamente en comparación con el UMAP no supervisado sin K-means.



La supervisión adicional en UMAP podría estar introduciendo una complejidad que K-means no maneja bien, resultando en clústers menos definidos.



7 of 17 5/18/24, 19:47

UMAP Componente 1