

Aprendizaje de Maquina No Supervisado

SVD: Estudiantes

En esta ocasión se realizó una transformación SVD a una base de datos con las calificaciones de 10 alumnos. Este dataset ya se había transformado por medio de PCA, por lo que se logro realizar una comparación.

Al aplicar una transformación **SVD de dos dimensiones**, vemos que la perdida es minima. Si observamos el primer renglón del data frame, vemos que la transformación varia por aproximadamente 0.4 tanto de incremento como de reducción.

	Matemáticas	Ciencias	Español	Historia	EdFísica
0	7.0	6.5	9.2	8.6	8.0
1	7.5	9.4	7.3	7.0	7.0
2	7.6	9.2	8.0	8.0	7.5
3	5.0	6.5	6.5	7.0	9.0
4	6.0	6.0	7.8	8.9	7.3
5	7.8	9.6	7.7	8.0	6.5
6	6.3	6.4	8.2	9.0	7.2
7	7.9	9.7	7.5	8.0	6.0
8	6.0	6.0	6.5	5.5	8.7
9	6.8	7.2	8.7	9.0	7.0

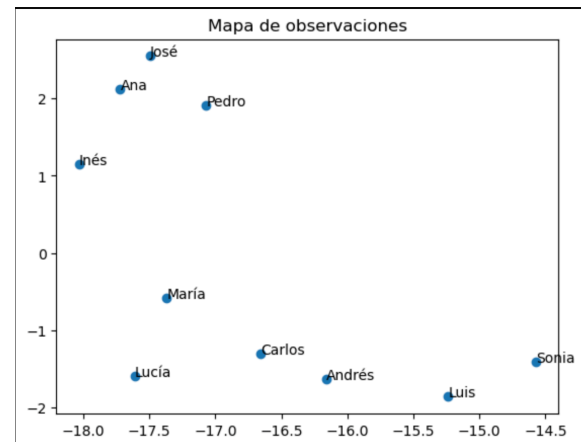
Datos originales

El diagrama generado con este numero de dimensiones es claro, y no presenta problema alguno para ser interpretado. Los alumnos están correctamente separados.

```
NotesReduced2 = UReduced2.dot(SigmaReduced2.dot(VTReduced2))
print("Matriz notes Transformada: ")
print(NotesReduced2)

Matriz notes Transformada:
[[6.65563617 6.93068907 8.45496902 8.64041633 8.61053016]
 [7.50259774 9.15546106 7.47627701 7.62801112 6.39154032]
 [7.66288835 9.06253486 8.0774837 8.24477427 7.24032869]
 [5.61309793 5.65867129 7.41578628 7.58014066 7.72189396]
 [6.05941975 6.24591247 7.7953721 7.96693512 7.99698132]
 [7.83156613 9.60455223 7.73118336 7.88753689 6.55416982]
 [6.35785122 6.70021013 7.95486409 8.12861624 8.02875678]
 [7.86973795 9.80171752 7.53881312 7.68951159 6.21497236]
 [5.48053658 5.67061811 7.01789357 7.17215282 7.18015541]
 [6.86687283 7.53702066 8.13213661 8.30697925 7.93006916]]
```

Datos transformados por medio de SVD a 2 dimensiones.



Scatter plot 2D

Continuando, al realizar una **transformación en tres dimensiones**, se observa como algunos de los valores se mantienen con una pérdida casi similar, pero los valores que solían aumentar, como es el caso del segundo valor en el primer renglón, el incremento es de casi 0.2, mostrando una mejoría en la precisión.

	Matemáticas	Ciencias	Español	Historia	EdFísica
0	7.0	6.5	9.2	8.6	8.0
1	7.5	9.4	7.3	7.0	7.0
2	7.6	9.2	8.0	8.0	7.5
3	5.0	6.5	6.5	7.0	9.0
4	6.0	6.0	7.8	8.9	7.3
5	7.8	9.6	7.7	8.0	6.5
6	6.3	6.4	8.2	9.0	7.2
7	7.9	9.7	7.5	8.0	6.0
8	6.0	6.0	6.5	5.5	8.7
9	6.8	7.2	8.7	9.0	7.0

Datos originales

```
# Reduccion a TRES Dimensiones
n_elements = 3

UReduced3 = U[:, :n_elements]

SigmaReduced3 = Sigma[:n_elements, :n_elements]

VTReducd3 = VT[:n_elements, :]

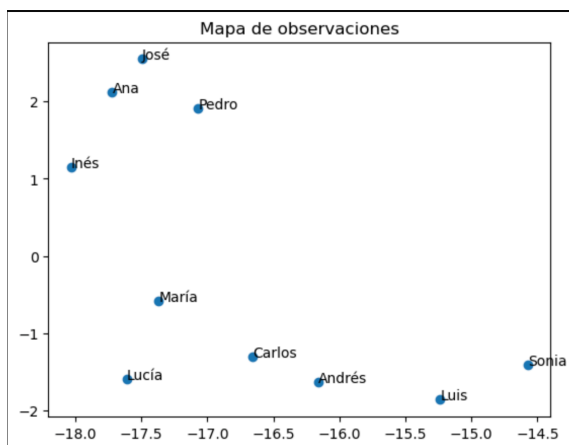
NotesReduced3 = UReduced3.dot(SigmaReduced3.dot(VTReducd3))
print("Matriz notes Transformada: ")
print(NotesReduced3)
```

Matriz notes Transformada:

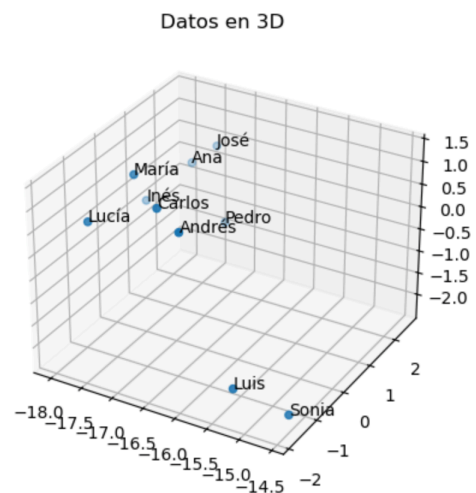
```
[[6.65245886 6.72762726 8.69042319 9.08512051 8.09898421]
 [7.50649243 9.40437093 7.1876611 7.08289996 7.01858503]
 [7.66453175 9.1675645 7.95569976 8.01475998 7.50491554]
 [5.6205042 6.13200674 6.8669445 6.54353881 8.91430342]
 [6.05471948 5.94551744 8.14368606 8.62479847 7.24023707]
 [7.83115817 9.57847935 7.76141538 7.94463634 6.48848799]
 [6.35253602 6.36051578 8.34874639 8.87254488 7.17301112]
 [7.86825556 9.7069775 7.64866605 7.89699168 5.97630674]
 [5.49031524 6.29557293 6.29324613 5.80350542 8.75451889]
 [6.86123771 7.17687926 8.54972766 9.09568679 7.02281405]]
```

Datos transformados por medio de SVD a 3 dimensiones.

La visualización en **3D** en esta ocasión, no es la más favorable, pues algunos de los valores se sobreponen. Los grupos siguen estando **correctamente separados**, pero en términos de practicidad, no presentan una mejoría en comparación a la visualización 2D.



Scatter plot 2D

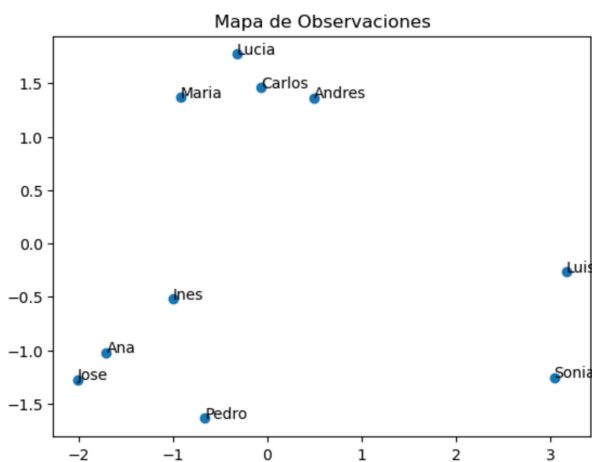


Scatter plot 3D

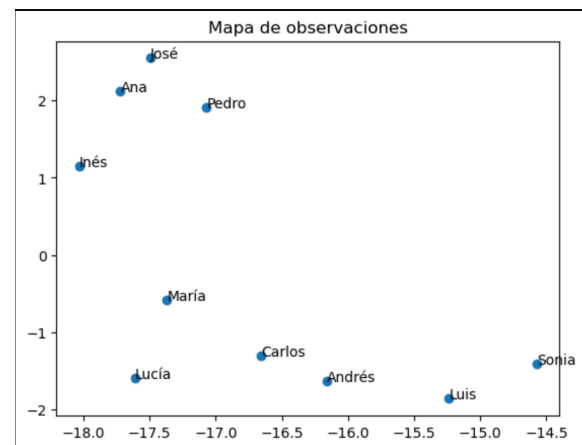
Realizando una **comparación con la descomposición por PCA**, notamos pequeñas diferencias. La primera, no siendo crítica, es el acomodo de los grupos, pues basta con girar la descomposición por SVD 90 grados a la derecha para tener la misma distribución que PCA.

La segunda diferencia que puede ser un poco más crítica, es que, si bien **los datos están agrupados de la misma manera**, SVD en esta ocasión, tendió a no mostrar tanta separación con el grupo de **Luis y Sonia**, un grupo notoriamente separado del resto en la descomposición por PCA.

Además, SVD tiene un poco más de separación dentro de cada grupo, a como lo muestra PCA. Comparando la escala de los Ejes, parece indicar un 'Zoom' a los datos, lo cual no está mal, pero la separación entre **Luis y Sonia** se mantiene como un caso que podría reflejar las pérdidas que hubo al momento de la transformación SVD, o bien, una mejora en contra de PCA.



Descomposición por PCA



Descomposición por SVD