



❖ **Asignatura:**

Aprendizaje Inteligente

❖ **Catedrático:**

FRANCISCO JAVIER LUNA ROSAS

❖ **Tema:**

Segundo Examen Parcial.

❖ **Equipo:**

Michelle Stephanie Adame

Andrés Eloy Escobedo Esparza

❖ **Carrera:**

I.C.I.

❖ **Semestre:**

6º

1. El alumno deberá elaborar un documento que defina, explique e implemente el análisis de componentes principales (PCA).

a) ¿Qué es PCA?

Es un método matemático que se utiliza para reducir el número de variables de forma que pasemos a tener el mínimo número de nuevas variables y que representen a todas las antiguas variables de la forma más representativa posible. Es decir, si se reduce el número de variables a dos o tres nuevas, se pueden representar los datos originales en el plano o en un gráfico de 3-dimensiones y, así, se visualiza de forma gráfica un resumen de nuestros datos. El simple hecho de poder tener los datos de manera visible simplifica mucho el entender qué puede estar pasando y ayuda a tomar decisiones.

b) ¿Qué diferencia existe con el Clustering Jerárquico (CJ)?

Ambos, tanto el PCA como el Clustering Jerárquico son métodos de aprendizaje no supervisado, pero trabajan de formas diferentes. La entrada del algoritmo de clustering jerárquico consiste en medir las similitudes (o diferencias) entre cada par de objetos, y la elección de la medida de similitud puede tener un gran efecto en el resultado. La meta de el algoritmo de clustering es particionar los objetos en grupos homogéneos, tal que las similitudes dentro de un grupo sean mucho mayores y notables sobre las similitudes entre todos los grupos. Por otro lado, el PCA, extrae componentes para representar los patrones de acuerdo las mayores varianzas en el data set, y no trata de maximizar la separación entre los grupos de muestras directamente.

c) ¿Qué diferencia existe con el K-Means?

El PCA busca representar todos los n vectores de datos como una combinación lineal de un pequeño número de vectores propios, y lo hace para minimizar la media cuadrada de la reconstrucción de error. Por otro lado, el K-means busca representar todos los n vectores de datos vía pequeños números de centroides de clusters para representar ellos como combinaciones lineales de un pequeño número de vectores de centroides de cluster donde la combinación de pesos debe ser todos cero excepto por el 1.

d) Aplicaciones de PCA.

Una de las aplicaciones de PCA es la reducción de dimensionalidad (variables), perdiendo la menor cantidad de información (varianza) posible: cuando contamos con un gran número de variables cuantitativas posiblemente correlacionadas (indicativo de existencia de información redundante), PCA permite reducirlas a un número menor de variables transformadas (componentes principales) que

expliquen gran parte de la variabilidad en los datos. Cada dimensión o componente principal generada por PCA será una combinación lineal de las variables originales, y serán además independientes o no correlacionadas entre sí. Las componentes principales generadas pueden utilizarse a su vez en métodos de aprendizaje supervisado, como regresión de componentes principales o partial least squares: ver capítulo Métodos de regularización.

El PCA también sirve como herramienta para la visualización de datos: supóngase que quisiéramos representar n observaciones con medidas sobre p variables ($X=X_1, X_2, \dots, X_p$) como parte de un análisis exploratorio de los datos. Lo que podríamos hacer es examinar representaciones bidimensionales, sin embargo, existen un total de $(p^2)=p(p-1)/2$ posibles representaciones entre pares de variables, y si el número de variables es muy alto, estas representaciones se harían inviables, además de que posiblemente la información contenida en cada una sería solo una pequeña fracción de la información total contenida en los datos.

El PCA puede considerarse como una rotación de los ejes del sistema de coordenadas de las variables originales a nuevos ejes ortogonales, de manera que estos ejes coincidan con la dirección de máxima varianza de los datos.

2. Explicación y análisis.

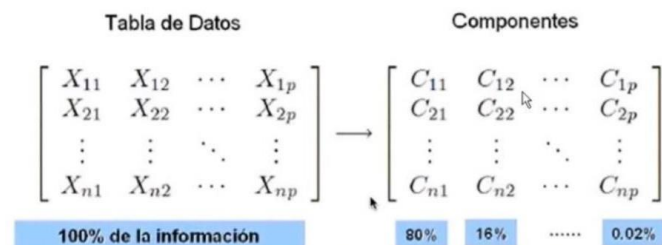


Figura 1. Transformación de la Tabla de Datos en Componentes.

**a). Transformación de la tabla de datos en componentes.
(Figura 1).**

Para calcular los componentes principales se hace lo siguiente :

Se crea una nueva variable estableciendo la combinación lineal
(primer componente)

$$X_1 = X_{a1}$$

donde X son las variables y a_1 son los coeficientes de la combinación lineal

Calculamos su varianza:

$$\frac{1}{n} z_1' z_1 = \frac{1}{n} a_1' X' X a_1 = a_1' S a_1$$

Maximizamos la varianza a partir de un Lagrangiano:

$$M = a_1' S a_1 - \lambda (a_1' a_1 - 1)$$

Segundo componente:

Se generan ahora 2 combinaciones lineales y se maximizar la varianza de ambas

$$\phi = a_1' S a_1 + a_2' S a_2 - \lambda_1 (a_1' a_1 - 1) - \lambda_2 (a_2' a_2 - 1)$$

**b). Reducción de la dimensión de los individuos (plano de
similitud de individuos en dos dimensiones).**

Al crearse la tabla de componentes tenemos en las 2 primeras
columnas el 96% de la información y estas se grafican y así
puede observar el cluster.

c). Reducción de la dimensión de las variables (circulo de correlación).

Se hace lo mismo a la tabla de variables que a la tabla de datos y al graficarse se muestra la correlación de variables

Interpretación:

Si X^i y X^j están cercanas entre si, entonces X^i y X^j son fuerte y positivamente correlacionadas

Si el ángulo entre X^i y X^j es cercana a los 90° entonces NO existe ninguna correlación entre ambas variables.

Si X^i y X^j están opuestas al vértice (origen) entonces existe una correlación fuerte y negativa entre X^i y X^j

d). Calidad de la representación de individuos y variables (cosenos cuadrados).

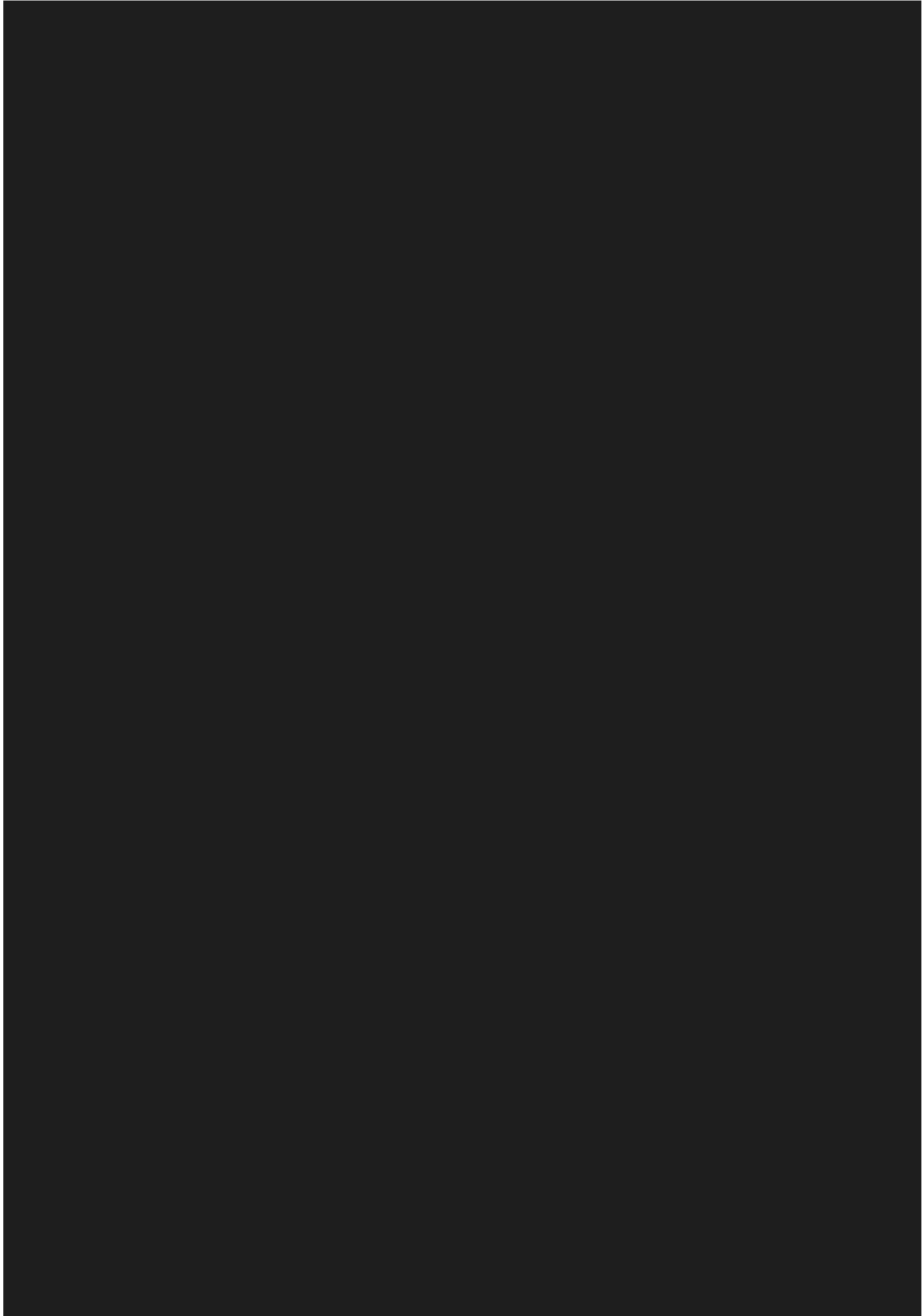
La representación de las variables se hace con cosenos cuadrados que es la suma del cuadrante 1 y el cuadrante 2 multiplicado por 100 y esa representación da la ubicación de los individuos en el plano de similitud

e). Interpretación de los clusters utilizando el principio de dualidad.

Se basa en los gráficos de plano de similitud de individuos en dos dimensiones y circulo de correlación, se comparan y se analizan las relaciones entre ellos

3. Implemente el algoritmo de PCA en Python, para calcular la matriz de componentes principales de los individuos (ver Figura 2), el dataset que se utilizará, es el dataset visto en clase de notas escolares.

- **Código.**



```
• plt.step(range(5), cum_var_exp, where='mid', linestyle='--', label='Vacumulada')
• plt.ylabel('RVarianza')
• plt.xlabel('PC')
• plt.legend(loc='best')
• plt.tight_layout()
• plt.show()
• #Paso 6 Calcule matriz de componentes principales
• matrix_w = np.hstack((eig_pairs[0][1].reshape(5,1),
•                       eig_pairs[1][1].reshape(5,1)))
• print('Components matrix:\n', matrix_w)
• Y = X_std.dot(matrix_w)
```

- Capturas.

```
PS C:\Users\ANDY\Desktop\Examen Luna 2P> & C:/Users/ANDY/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/ANDY/Desktop/Examen Luna 2P/P
CA.py"
Matematicas Ciencias Espanol Historia EdFisica
Lucia 7.0 6.5 9.2 8.6 8.0
Pedro 7.5 9.4 7.3 7.0 7.0
Ines 7.6 9.2 8.0 8.0 7.5
Luis 5.0 6.5 6.5 7.0 9.0
Andres 6.0 6.0 7.8 8.9 7.3
Ana 7.8 9.6 7.7 8.0 6.5
Carlos 6.3 6.4 8.2 9.0 7.2
Jose 7.9 9.7 7.5 8.0 6.0
Sonia 6.0 6.0 6.5 5.5 8.7
Maria 6.8 7.2 8.7 9.0 7.0
[[7. 6.5 9.2 8.6 8. ]
 [7.5 9.4 7.3 7. 7. ]
 [7.6 9.2 8. 8. 7.5]
 [5. 6.5 6.5 7. 9. ]
 [6. 6. 7.8 8.9 7.3]
 [7.8 9.6 7.7 8. 6.5]
 [6.3 6.4 8.2 9. 7.2]
 [7.9 9.7 7.5 8. 6. ]
 [6. 6. 6.5 5.5 8.7]
 [6.8 7.2 8.7 9. 7. ]]
```

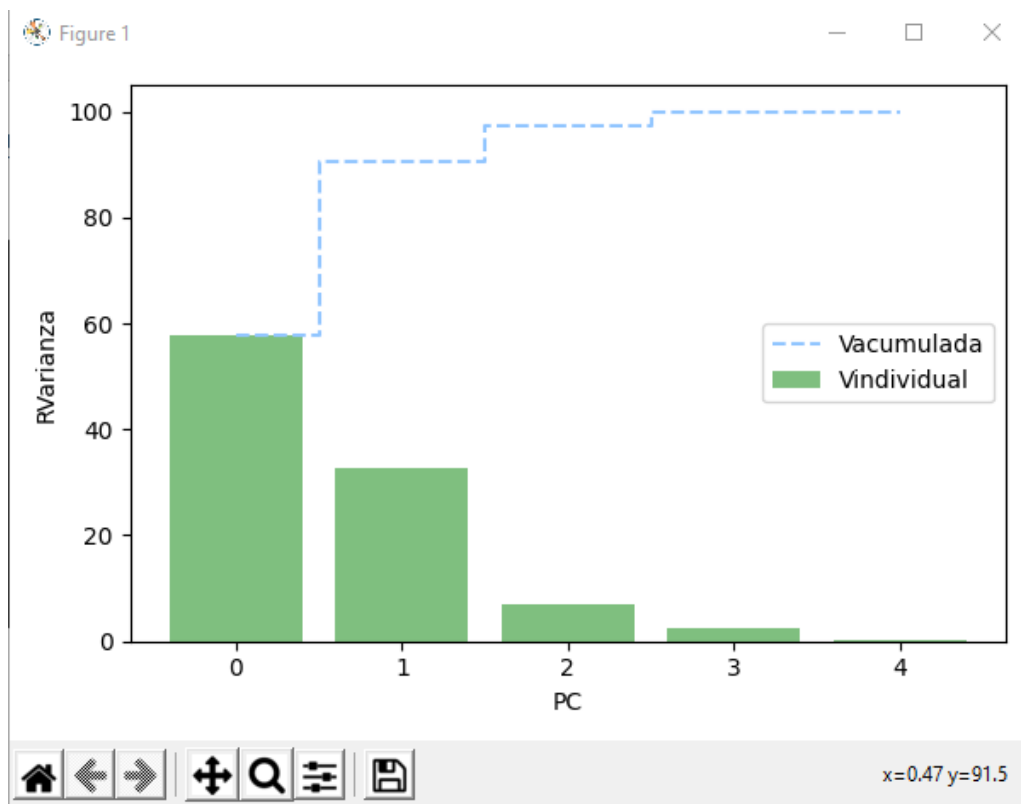
```
Covarianza Matrix:
[[ 1.11111111  0.94897642  0.42730471  0.23021584 -0.87462521]
 [ 0.94897642  1.11111111 -0.0222802 -0.02393269 -0.76413395]
 [ 0.42730471 -0.0222802  1.11111111  0.9121291 -0.40615935]
 [ 0.23021584 -0.02393269  0.9121291  1.11111111 -0.56444591]
 [-0.87462521 -0.76413395 -0.40615935 -0.56444591  1.11111111]]

Eigenvectors
[[-0.52664397 -0.2704963 -0.43820071 -0.62387762 -0.26121779]
 [-0.42493622 -0.50807221 -0.04049491  0.32538951  0.67362724]
 [-0.35914704  0.56208159 -0.56227583  0.48374732 -0.07008647]
 [-0.35269747  0.58648985  0.39418032 -0.42043348  0.44664495]
 [ 0.53730181  0.09374599 -0.57862603 -0.30679407  0.52305619]]

Eigenvalues
[3.21472186 1.80961158 0.38510672 0.00987933 0.13623607]
Autovalores en orden descendente:
3.2147218593532703
1.8096115833081277
0.3851067205716997
0.13623606621916967
0.009879326103283882
```


Components matrix:

```
[[-0.52664397 -0.2704963 ]  
 [-0.42493622 -0.50807221]  
 [-0.35914704  0.56208159]  
 [-0.35269747  0.58648985]  
 [ 0.53730181  0.09374599]]
```



4. Conclusión.

Esta técnica es muy utilizada para describir un conjunto de datos en términos de nuevas variables o componentes no correlacionadas. Los componentes se ordenan por la cantidad de varianza original que describen, por lo que la técnica es útil para reducir la dimensionalidad de un conjunto de datos.

Con este trabajo nos damos cuenta de la importancia de tener conocimientos de álgebra lineal, como son las operaciones con matrices y las características y propiedades de los vectores propios. También aprendimos sus semejanzas con otros métodos de aprendizaje no supervisado como el clustering jerárquico y el método de k-means.

5. Bibliografía.

- [1] RPubs - Análisis de componentes principales (PCA). (2019, 27 septiembre). PCA. https://rpubs.com/Cristina_Gil/PCA
- [2] What is the relation between k-means clustering and PCA? (2015, 23 noviembre). Cross Validated. <https://stats.stackexchange.com/questions/183236/what-is-the-relation-between-k-means-clustering-and-pca#:~:text=PCA%20is%20used%20to%20project%20the%20data%20onto%20two%20dimensions.&text=Then%2C%20K-means%20can%20be,supervised%20methods%20in%20machine%20learning>.
- [3] O. (2020, 7 septiembre). ¿Qué es el Clustering? GraphEverywhere. <https://www.grapheverywhere.com/que-es-el-clustering/>. GraphEverywhere.