



Presentación Tarea 1

Análisis de Componentes

Principales

Machine Learning

Integrantes: Bruno Fonseca, Valentín Grau, Valeria Rodríguez

Profesor: Tomás Fontencilla

Código

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from sklearn.datasets import load_breast_cancer
```

```
#Cargamos datos
data = load_breast_cancer()
df = pd.DataFrame(data.data, columns=data.feature_names)
df
```

```
#Es necesario normalizar los datos,
# ya que pca trabaja mejor con datos centrados y escalados.
df_std = StandardScaler().fit_transform(df)
df_std= pd.DataFrame(df_std, columns=df.columns)
```

```
#Creamos la matriz de correlacion entre cada columna
corr = df_std.corr()
corr
```

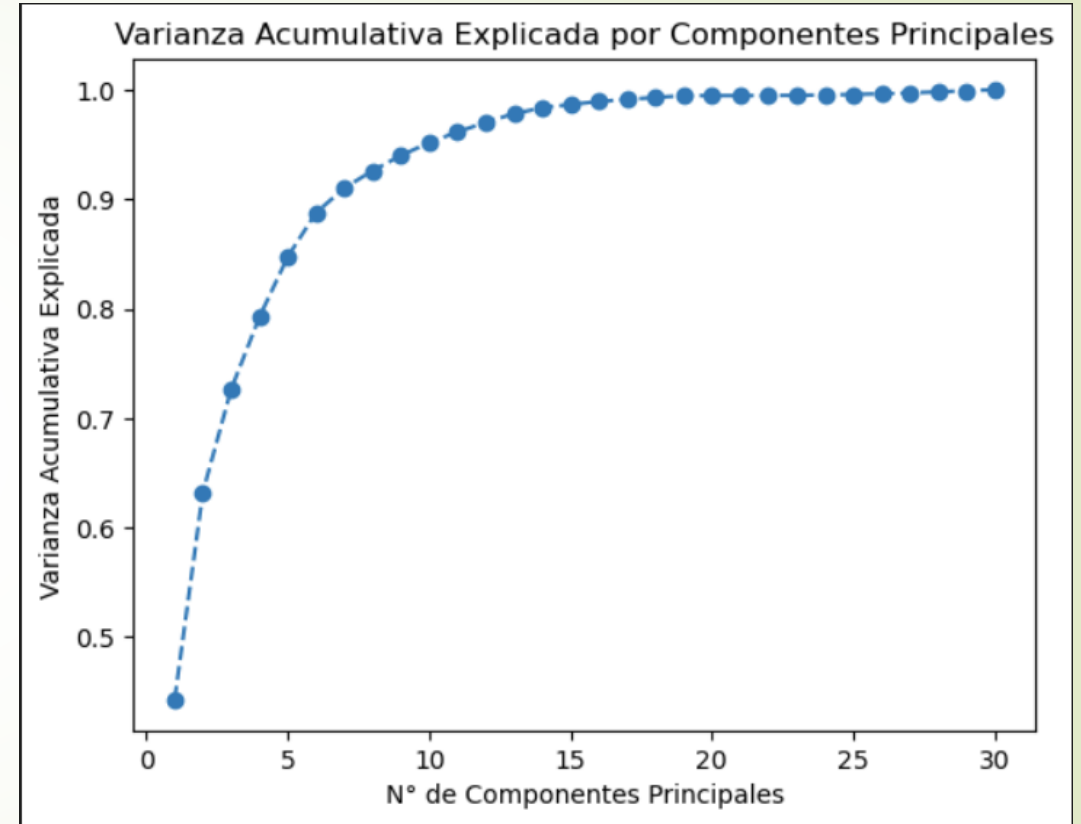
```
#Calculamos valores y vectores propios de la matriz de correlacion
values, vectors = np.linalg.eig(corr)
```

Código

```
#Juntamos en pares y ordenamos de manera descendente segun valor propio
pares = [(np.abs(values[i]), vectors[:, i]) for i in range(len(values))]
pares.sort(key=lambda x: x[0], reverse=True)
pares
```

```
#Calculamos la varianza explicada por cada componente, y la acumulada.
var_total = sum(values)
var_expl = [value / var_total for value in values]
var_cum = np.cumsum(var_expl)
var_cum
```

```
#Graficamos la varianza explicada segun componentes principales
plt.plot(range(1, len(var_cum) + 1), var_cum, marker='o', linestyle='--')
plt.grid()
plt.xlabel('N° de Componentes Principales')
plt.ylabel('Varianza Acumulativa Explicada')
plt.title('Varianza Acumulativa Explicada por Componentes Principales')
plt.show()
```

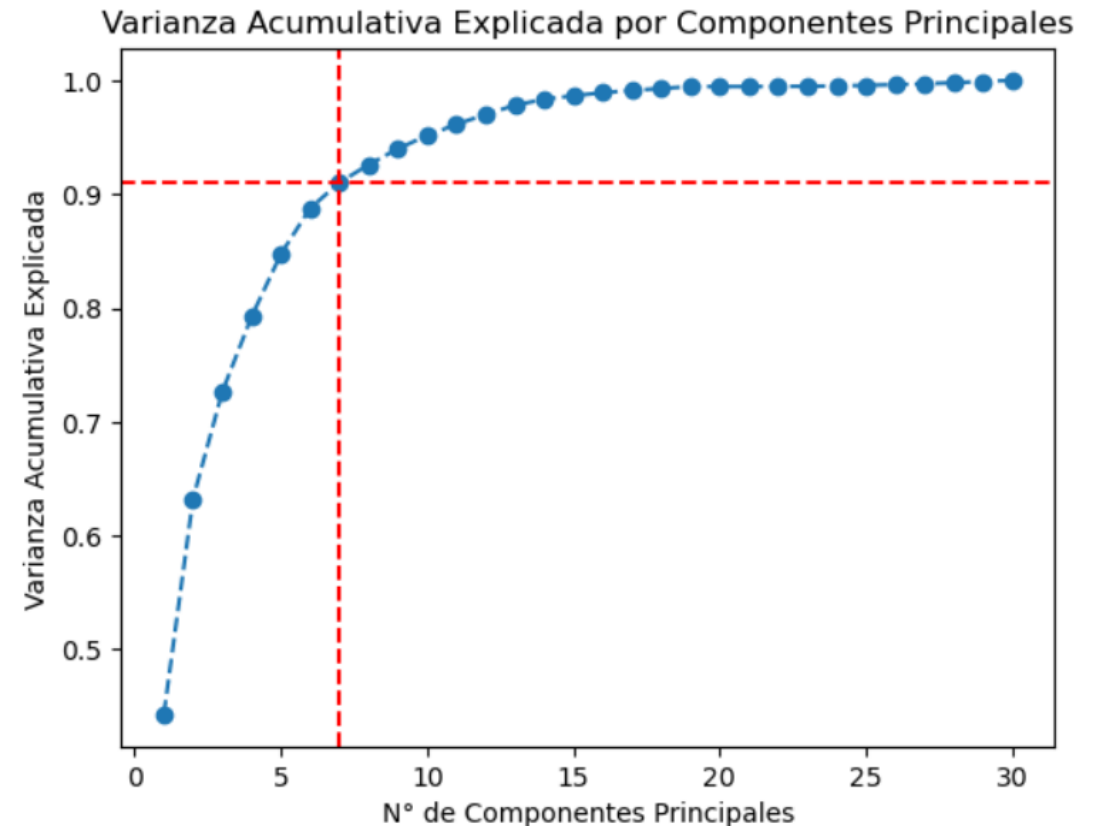


Conclusiones

Al analizar el gráfico de varianza acumulada explicada, optamos por definir 7 componentes principales como adecuados para un posterior análisis, ya que explica el 91% de la varianza acumulada.

Se considera adecuado ya que se logra reducir significativamente la dimensionalidad de los datos y al mismo tiempo se logra mantener una representación efectiva de la estructura inicial de los datos.

```
#Graficamos la varianza explicada segun componentes principales
plt.plot(range(1, len(var_cum) + 1), var_cum, marker='o', linestyle='--')
posicion_vertical = 7
plt.axvline(x=posicion_vertical, color='red', linestyle='--', label='Línea Vertical')
posicion_horizontal = 0.9100953006967304
plt.axhline(y=posicion_horizontal, color='red', linestyle='--', label='Línea Horizontal')
plt.xlabel('N° de Componentes Principales')
plt.ylabel('Varianza Acumulativa Explicada')
plt.title('Varianza Acumulativa Explicada por Componentes Principales')
plt.show()
```



Presentación Tarea 1

Análisis de Componentes Principales

Machine Learning

Integrantes: Bruno Fonseca, Valentín Grau, Valeria Rodríguez

Profesor: Tomás Fontencilla