

Hoja de Prácticas: PCA

Álgebra aplicada

Domingo López Rodríguez

Índice de Contenidos

Ejercicios Teóricos y de Reflexión	3
Ejercicios Prácticos: Dataset Wine	4

Instrucciones

Realiza los siguientes ejercicios utilizando R. Entrega un documento (PDF generado con Quarto) que incluya el código, los resultados y una breve explicación razonada de cada paso, conectando los resultados con los conceptos de Álgebra Lineal vistos en la asignatura.

Ejercicios Teóricos y de Reflexión

Ejercicio 1: Espacios Vectoriales y Dimensión

Considerando un dataset con 13 variables numéricas (como el dataset `wine` que veremos luego):

1. ¿Cómo describirías el espacio vectorial en el que reside cada observación antes de aplicar el PCA? ¿Cuál es la dimensión de este espacio?
2. Si decidimos retener solo las 3 primeras componentes principales, ¿qué estamos haciendo desde el punto de vista del Álgebra Lineal (subespacios, proyecciones)?

Ejercicio 2: Ortogonalidad y Diagonalización

El PCA genera nuevas componentes principales que forman una nueva base para el espacio de los datos.

1. ¿Por qué esta nueva base es ortogonal (o incluso ortonormal)?
2. ¿Cómo se relaciona la construcción de esta base con el proceso de **diagonalización** de una matriz simétrica (la matriz de covarianza)?

Ejercicio 3: Autovalores y Varianza

Los autovalores λ_i de la matriz de covarianza nos indican la varianza explicada en cada dirección.

1. Si la suma de todos los autovalores es S , ¿qué representa el ratio λ_i/S ?
2. ¿Por qué el autovector asociado al **mayor** autovalor es el más importante para el PCA?

Ejercicios Prácticos: Dataset Wine

Para los siguientes ejercicios, utilizaremos el dataset `wine` (del paquete `HDclassif`), que contiene análisis químicos de 178 vinos italianos divididos en 3 cultivares.

Ejercicio 4: Carga y Preparación

Carga el dataset `wine`, elimina la primera columna (que contiene la clase) y realiza una breve exploración estadística. ¿Crees que las variables están en la misma escala? Justifica si es necesario estandarizar.

```
data(wine)
wine_data <- wine[, -1]
# Tu código aquí...
```

Ejercicio 5: Ejecución del PCA

Realiza el PCA sobre `wine_data` asegurando que los datos se escalen (`scale = TRUE`).

1. ¿Cuánta varianza explican las dos primeras componentes?
2. ¿Cuántas componentes necesitarías para retener al menos el 80% de la información original? Dibuja el *Scree Plot*.

Ejercicio 6: Interpretación de Pesos (Loadings)

Examina la matriz de rotación (`rotation`).

1. Identifica las 3 variables que más contribuyen (positiva o negativamente) a la PC1.
2. Basándote en esto, ¿qué crees que intenta "medir" la primera componente en el perfil químico de un vino?

Ejercicio 7: Visualización: El Biplot

Genera un biplot del análisis.

1. Observa las flechas de `Total phenols` y `Flavanoids`. ¿Qué puedes decir sobre su correlación?
2. Identifica si hay algún grupo de vinos que se separe claramente del resto. Colorea los

puntos por la variable `class` original para verificarlo.

Ejercicio 8: Reconstrucción (Opcional Avanzado)

Toma la primera observación del dataset. Multiplica sus valores escalados por la primera columna de la matriz de rotación. Compara el resultado con el primer "score" de esa observación en `acp$x`. ¿Qué has calculado exactamente?