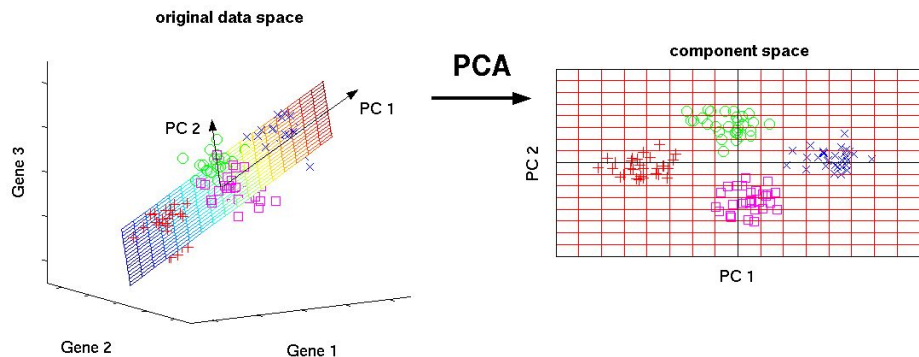


Análisis de Componentes Principales



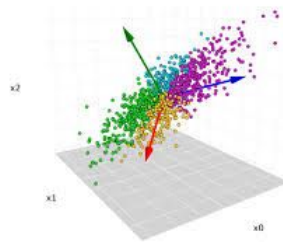
¿Qué es?

- El Análisis de Componentes Principales (Principal Component Analysis, PCA) es un algoritmo de aprendizaje no supervisado.
- Se utiliza comúnmente en el análisis exploratorio de datos.



¿Para qué sirve PCA?

- Reducción de dimensionalidad: reducir el número de variables manteniendo la mayor cantidad de información posible.
- Investigación de subgrupos: identificar subgrupos dentro de las variables u observaciones.
- Proyección de datos en espacios de menor dimensión para facilitar la interpretación y visualización.



Explicación Simple

- PCA transforma un gran número de variables posiblemente correlacionadas en un menor número de variables no correlacionadas llamadas componentes principales.
- Cada componente principal es una combinación lineal de las variables originales.

Cómo funciona PCA

- **Cálculo de la Media:** Calcular la media de cada variable.
- **Restar la Media:** Restar la media de cada variable a los datos originales.
- **Cálculo de la Matriz de Covarianza:** Calcular la matriz de covarianza de los datos ajustados.
- **Cálculo de los Vectores y Valores Propios:** Calcular los vectores propios (direcciones principales) y los valores propios (varianza explicada) de la matriz de covarianza.
- **Selección de Componentes Principales:** Seleccionar los vectores propios con los valores propios más altos.
- **Transformación de los Datos:** Proyectar los datos originales en el espacio de los componentes principales seleccionados.

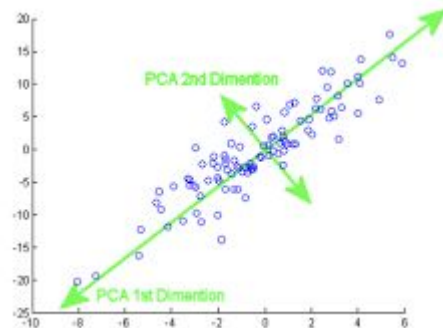
Reducción de Dimensionalidad

- **Componentes Principales:**

- Se ordenan de mayor a menor según la varianza explicada.
- Los primeros componentes principales retienen la mayor parte de la información.

- **Visualización:**

- Las observaciones pueden representarse en un gráfico de 2 o 3 dimensiones si los primeros componentes explican un porcentaje suficiente de la variabilidad.



Usos de PCA

- **Estudio de Correlaciones:**
 - Identificar correlaciones entre variables y reducir el número de variables a medir.
- **Obtención de Factores No Correlacionados:**
 - Utilizar estos factores en métodos de modelización como regresión lineal o logística.
- **Visualización:**
 - Identificar grupos uniformes o atípicos en un espacio bidimensional o tridimensional.

Beneficios del PCA

- Simplificación del análisis de datos complejos.
- Reducción de la redundancia de datos.
- Mejora en la visualización y comprensión de los datos.
- Facilita la identificación de patrones y subgrupos en los datos.



Conclusión

- PCA es una herramienta poderosa para la reducción de dimensionalidad y la exploración de datos.
- Permite simplificar y visualizar datos complejos, manteniendo la mayor cantidad de información posible.