# Preguntas Conceptuales.

1° ¿Que pasa si no estandarizamos características con escalas muy diferentes?

Si no se estandarizan, las componentes principales estarán controlados por variables con mayor varianza a nivel de escala. por lo que varias variables relevantes en escala pequeña pueden quedar ignoradas.

2° ¿ cómo determinar el número óptimo de componentes?

Elegir el mínimo número de componentes que expliquen el umbral objetivo, también conservar con un valor > 1.

3° Sí PCA reduce dimensionalidad conservando la mayor parte de la varianza, se puede usar para compensión de reconstrucción aproximada.

Con iris: $n = 150$ muestras, $P = 4$ características

Original: almacena $(150 - 41) = 600$ números

con 2 componentes sería $n \cdot k = 2 \cdot 150 = 300$ números.

con 2 componentes principales sería $P \cdot k = 2 \cdot 4 = 8$

Media $P = 4$

total sería $= 300 + 8 + 4 = 312$ números.

razón de comprensión $= \frac{600}{312} = 1,9231$

Se reduce un 1,9231 del tamaño almacenado.

4. Relación entre PCA y SVD:

Sea X la matriz de datos centrada $(n \times p)$. La SVD de X es

$$X = U \Sigma V^T$$

$U(n \times r)$, $\Sigma(r \times r)$ diagonal con valores singulares $\sigma_i$ y

$V(P \times r)$ vectores singulares pequeños.

Calcular la SVD de la matriz X y usar V y los $\sigma_i$ para obtener direcciones y varianzas.