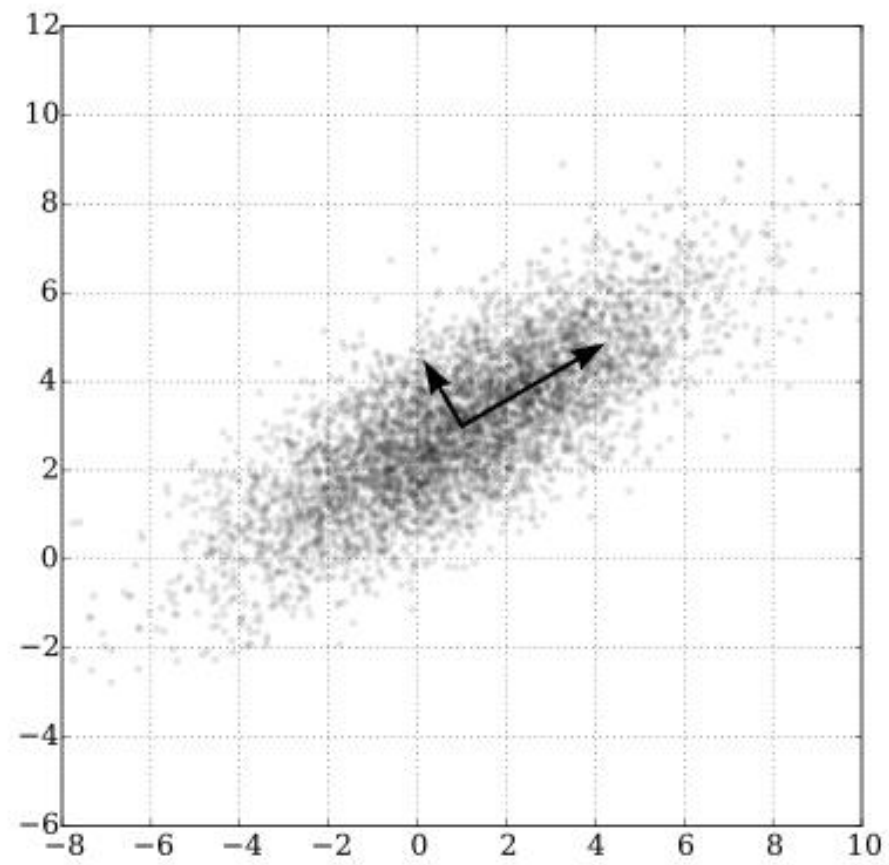




PCA y Reducción de la Dimensionalidad

MACHINE LEARNING

PhD. César Astudillo | Facultad de Ingeniería

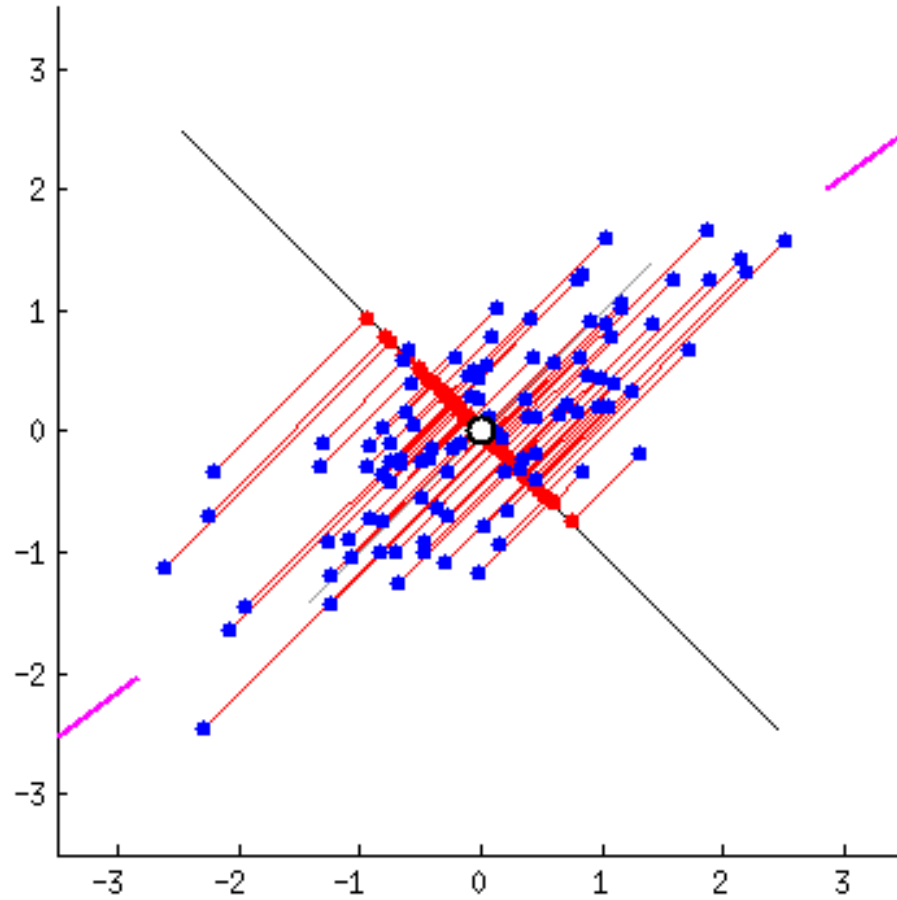


[This Photo](#) by Unknown Author is licensed under [CC BY-SA](#)

¿Qué es PCA?

- **Definición:** Análisis de Componentes Principales (PCA) es una técnica de reducción de dimensionalidad que transforma un conjunto de variables correlacionadas en un nuevo conjunto de variables no correlacionadas llamadas **componentes principales**.
- **Objetivo:** Simplificar la complejidad de los datos mientras se retiene la mayor cantidad de información posible.

¿Cómo funciona PCA?



¿Cómo funciona PCA?

- A través de un mecanismo del Álgebra lineal, encuentra nuevas direcciones (componentes principales) que capturan la mayor variabilidad de los datos.
- Los datos son proyectados a un nuevo espacio con menos dimensiones.

Aplicaciones Comunes de PCA

- Visualización de datos complejos.
- Eliminación de ruido.
- Reducción de multicolinealidad.

Fundamentos Matemáticos de PCA

Concepto Clave:

- Proyección de los datos en nuevas bases ortogonales (componentes principales).

Pasos del Algoritmo:

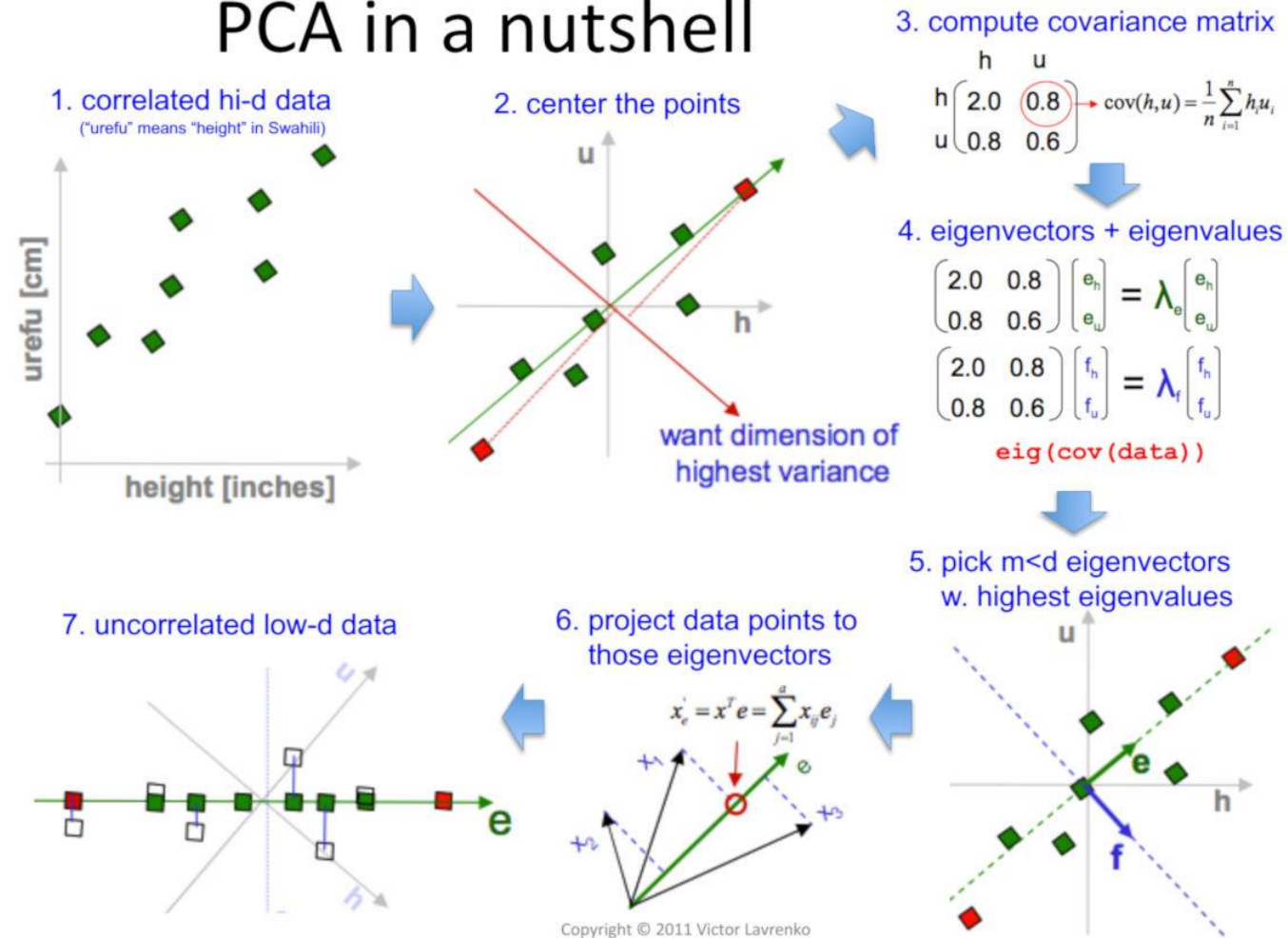
1. Estandarización: Centrar los datos en media 0 y varianza 1.
2. Cálculo de la Matriz de Covarianza: Capturar relaciones lineales entre variables.
3. Descomposición en Autovalores y Autovectores: Los autovectores son las direcciones principales.
4. Selección de Componentes: Mantener los componentes con mayor varianza explicada.

Fórmula Clave:

- $X_{\text{transformado}} = XW$ donde W es la matriz de autovectores.

¿Cómo funciona PCA?

PCA in a nutshell



Interpretación de los Componentes

¿Qué representa cada componente?

- Los componentes principales son combinaciones lineales de las variables originales.

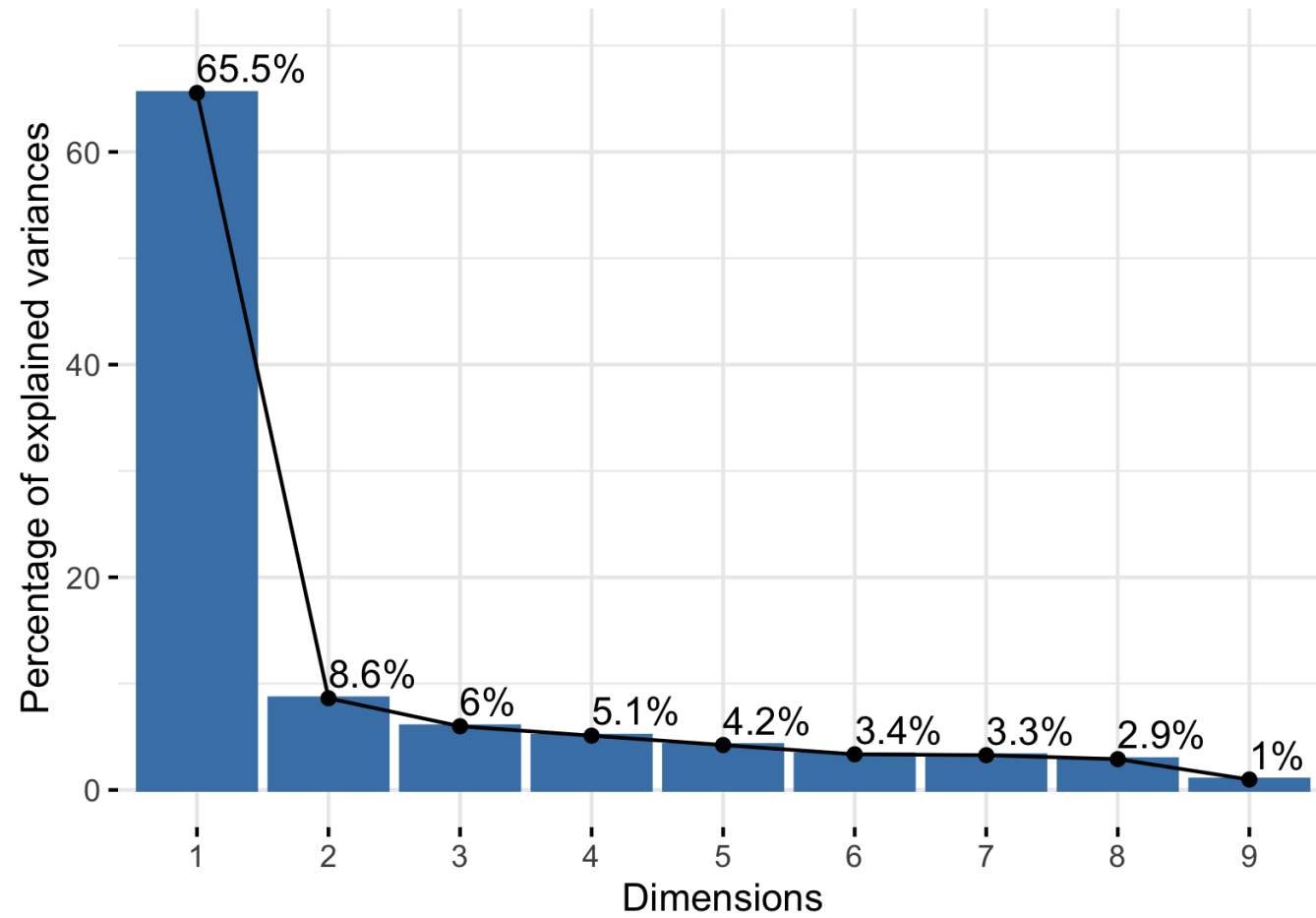
Varianza Explicada:

- El primer componente explica la mayor parte de la varianza.
- Los siguientes explican varianzas decrecientes.

Cómo Determinar Cuántos Componentes Usar:

- **Scree Plot:** Visualización del codo de la curva.
- **Proporción de Varianza Explicada:** Seleccionar el número que captura un % alto de varianza.

Scree Plot



Varianza Acumulada

| | Eigenvalue | % Variance | Cumulated % Variance |
|-----|-------------------|-------------------|-----------------------------|
| PC1 | 8.21 | 39.12 | 39.12 |
| PC2 | 2.99 | 14.26 | 53.38 |
| PC3 | 2.81 | 13.36 | 66.74 |
| PC4 | 2.34 | 11.12 | 77.86 |
| PC5 | 1.37 | 6.53 | 84.39 |

Ventajas y Limitaciones de PCA

Ventajas:

- Reducción efectiva de dimensionalidad.
- Elimina la multicolinealidad.
- Mejora la visualización de los datos.

Limitaciones:

- Solo captura relaciones lineales.
- Sensible a la escala de los datos.
- Difícil de interpretar cuando se reducen muchas dimensiones.



PCA y Reducción de la Dimensionalidad

MACHINE LEARNING

PhD. César Astudillo | Facultad de Ingeniería