



UNIVERSIDAD
PANAMERICANA

Reducción de Dimensionalidad

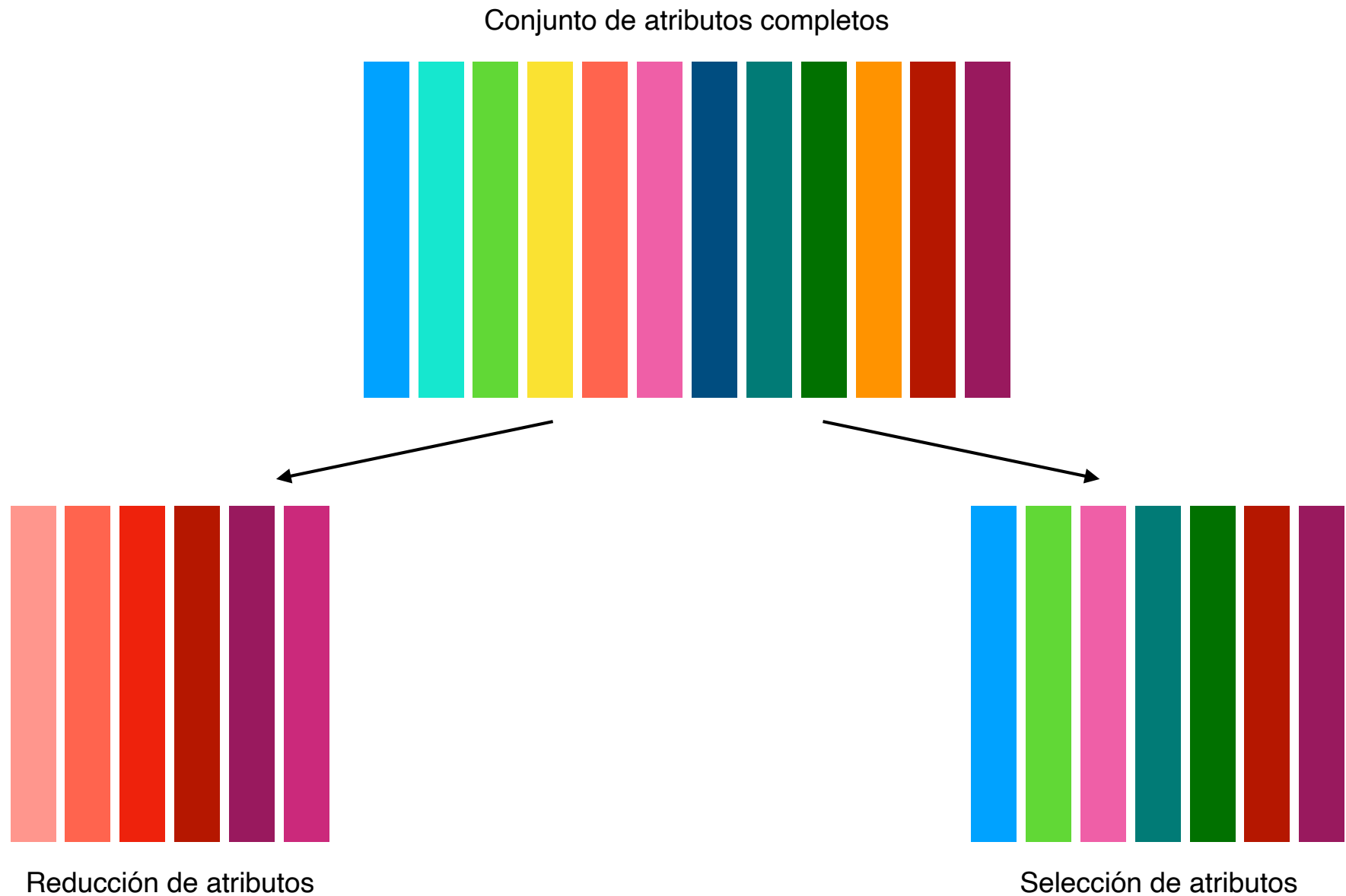
Un problema común en aprendizaje máquina es el **excesivo número de atributos** que tiene un conjunto de datos, lo cual tiene consecuencias tales como:

- Dificultad para analizar los datos
- Dificultad para interpretar los datos
- Visualización de datos casi imposible
- Almacenamiento excesivo de datos

Dos formas de dar solución al problema son:

- Reducción de atributos
- Selección de atributos

Reducción de dimensionalidad



(blank slide)



UNIVERSIDAD
PANAMERICANA

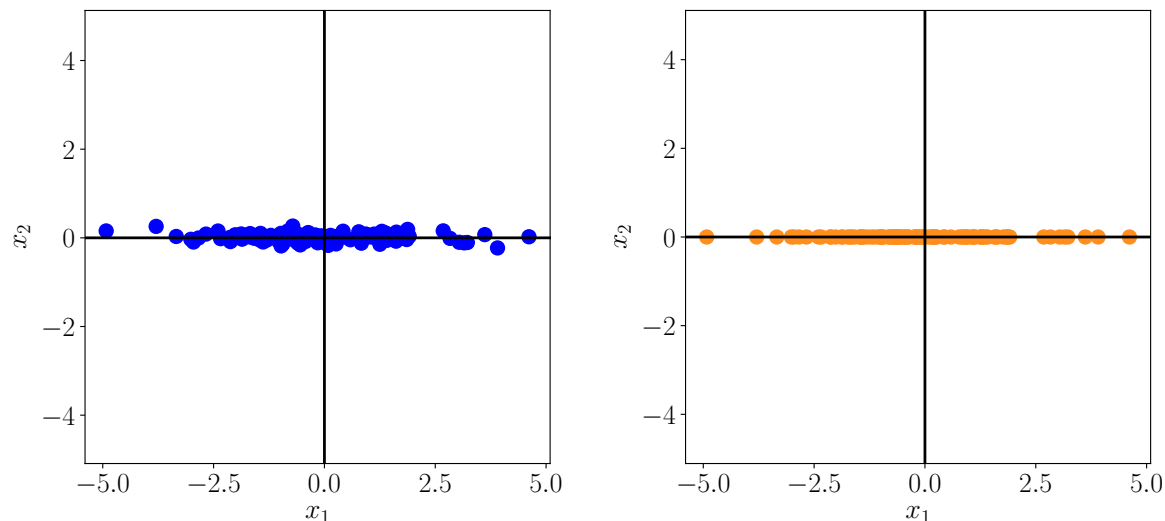
Análisis de componentes principales

Análisis de componentes principales

La técnica de **análisis de componentes principales (PCA, en inglés)** hace una transformación ortogonal sobre un conjunto de datos, posiblemente correlacionados, que deriva en un conjunto de atributos linealmente no correlacionados.

Como cualquier técnica de reducción de atributos, se busca una transformación de datos en n dimensiones a un conjunto de datos de k dimensiones, tal que $k < n$.

$$T : \mathbb{R}^n \rightarrow \mathbb{R}^k$$



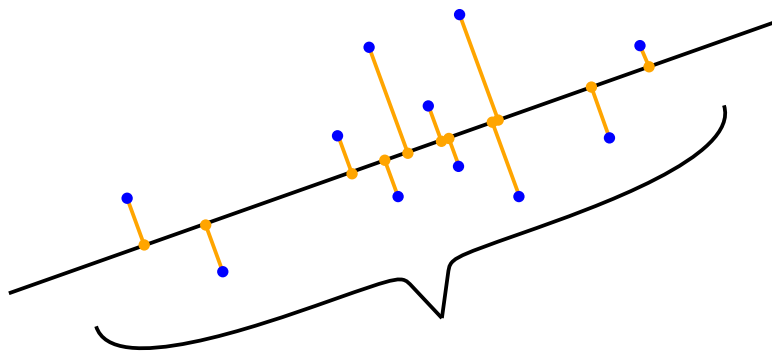
Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

Análisis de componentes principales

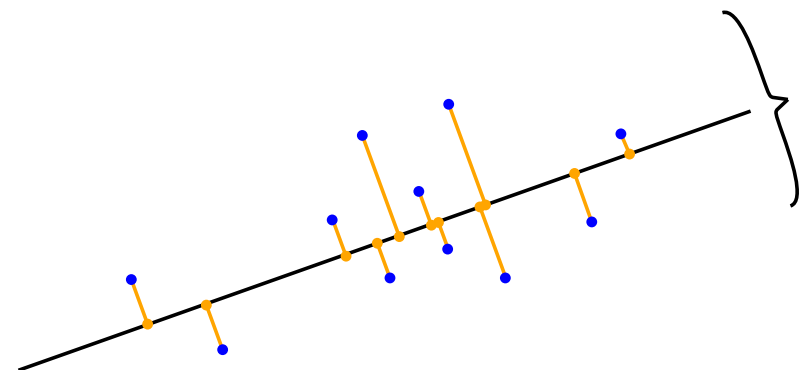
La técnica de **análisis de componentes principales (PCA, en inglés)** hace una transformación ortogonal sobre un conjunto de datos, posiblemente correlacionados, que deriva en un conjunto de atributos linealmente no correlacionados.

Principio de la técnica PCA:

- La transformación generada es una proyección espacial.
- Tiene un espacio que maximiza la variabilidad de los datos proyectados sobre él.
- Los datos proyectados no deben estar correlacionados.
- Los datos originales tienen media cero y covarianza S .



Perspectiva: maximización de la variabilidad

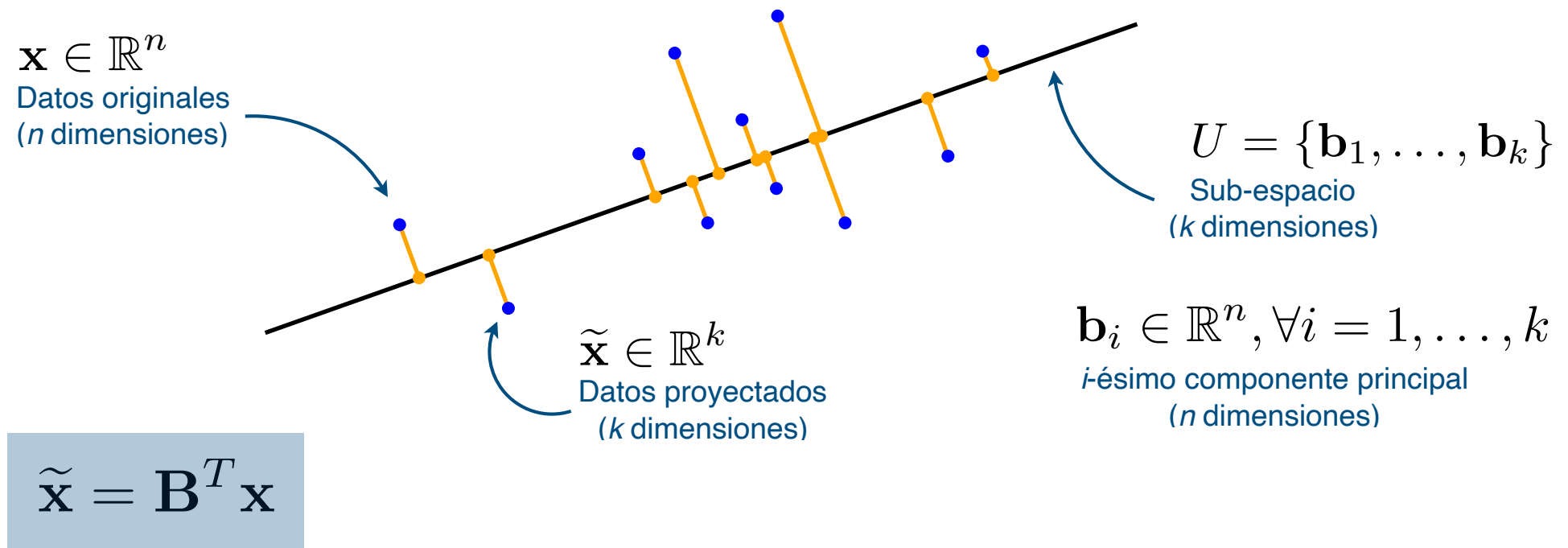


Perspectiva: minimización de las diferencias

Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

Análisis de componentes principales

La técnica de **análisis de componentes principales (PCA, en inglés)** hace una transformación ortogonal sobre un conjunto de datos, posiblemente correlacionados, que deriva en un conjunto de atributos linealmente no correlacionados.



Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

Análisis de componentes principales

La técnica de **análisis de componentes principales (PCA, en inglés)** hace una transformación ortogonal sobre un conjunto de datos, posiblemente correlacionados, que deriva en un conjunto de atributos linealmente no correlacionados.

Consideraciones de la técnica PCA:

- La transformación tiene una base ortonormal que permite independencia de atributos.
- Cada dato proyectado es una combinación lineal de los atributos originales.

Matriz de
covarianza

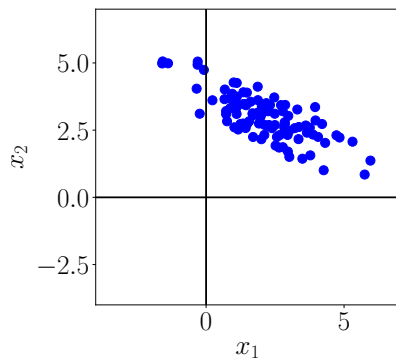
$$\mathbf{S} = \frac{1}{N} \mathbf{X} \mathbf{X}^T \in \mathbb{R}^{n \times n}$$

Representación de
la transformación

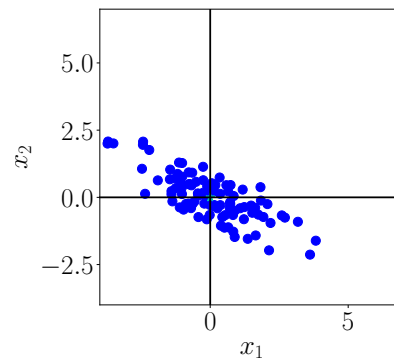
$$\mathbf{S} \mathbf{b}_i = \lambda_i \mathbf{b}_i, \forall i = 1, \dots, k$$

$$\begin{aligned} & \max_{\mathbf{b}_1} \mathbf{b}_1^\top \mathbf{S} \mathbf{b}_1 \\ & \text{subject to } \|\mathbf{b}_1\|^2 = 1. \end{aligned}$$

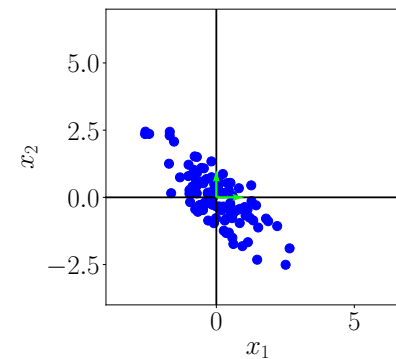
Análisis de componentes principales



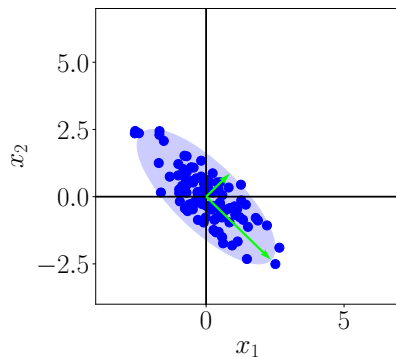
(a) Original dataset.



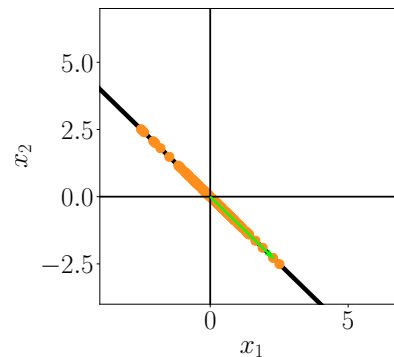
(b) Step 1: Centering by subtracting the mean from each data point.



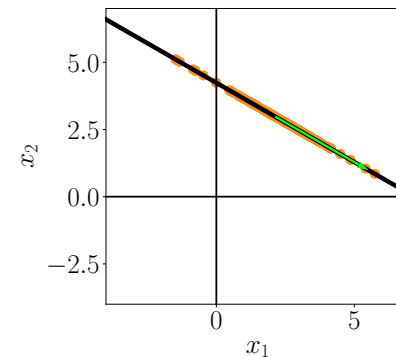
(c) Step 2: Dividing by the standard deviation to make the data unit free. Data has variance 1 along each axis.



(d) Step 3: Compute eigenvalues and eigenvectors (arrows) of the data covariance matrix (ellipse).



(e) Step 4: Project data onto the principal subspace.



(f) Undo the standardization and move projected data back into the original data space from (a).

Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

(blank slide)



UNIVERSIDAD
PANAMERICANA

Eliminación Recursiva de Atributos

Eliminación recursiva de atributos

La **eliminación recursiva de atributos (RFE, en inglés)** es un método de reducción de dimensiones que selecciona un subconjunto de atributos que preservan el desempeño de un modelo de aprendizaje. Además, el método es capaz de jerarquizar el poder de predicción que tiene cada atributo en el subconjunto generado.

Modelo de
aprendizaje

(sólo supervisado)

Jerarquización
de atributos

(criterio de jerarquía)

Eliminación
de atributos

(mínimo valor en criterio
se elimina)

$$c_i = (w_i)^2, \forall i = 1, \dots, n_t$$

Pesos asociados al
discriminador del modelo

Número de atributos
en la iteración t

Guyon, I., et al., Gene Selection for Cancer Classification using Support Vector Machines, *Machine Learning* 46: 389–422, 2002.

Eliminación recursiva de atributos – *algoritmo simple*

1. Inicializar el conjunto de atributos, S
2. Inicializar el subconjunto de atributos, R
3. Mientras no esté vacío S :
 - i. Entrenar un modelo supervisado usando $X(:, S)$, $y(:, S)$
 - ii. Evaluar el criterio de jerarquía de cada atributo en S
 - iii. Elegir los k atributos con menor criterio y actualizar R
 - iv. Eliminar los k atributos de S
4. Regresar el subconjunto de atributos R (atributos eliminados)

Guyon, I., et al., Gene Selection for Cancer Classification using Support Vector Machines, *Machine Learning* 46: 389–422, 2002.

(blank slide)