

{desafío}
latam_

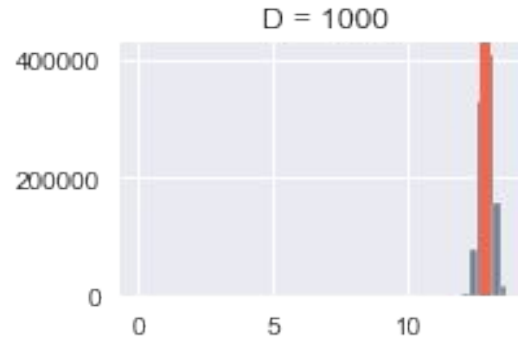
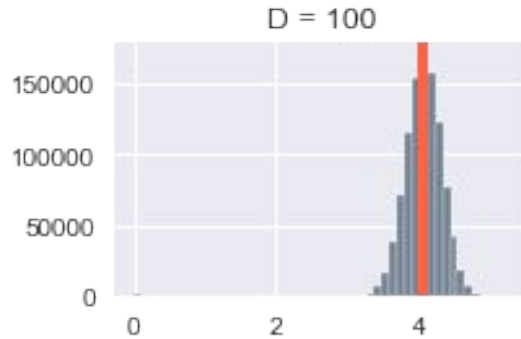
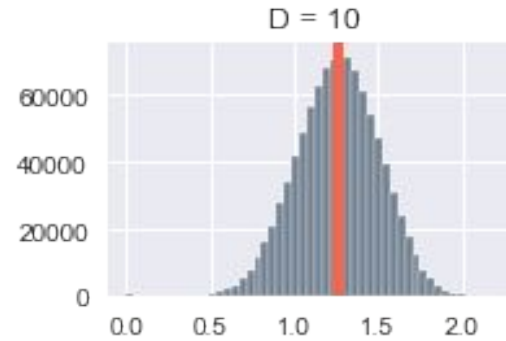
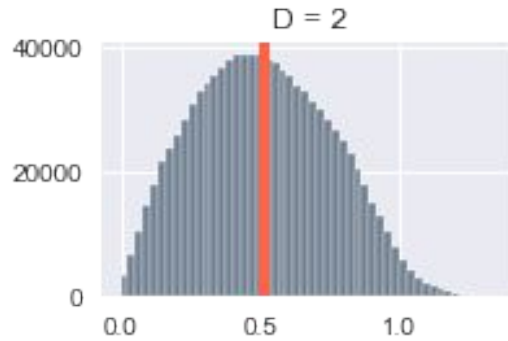
Dimensionalidad _



Aprendizaje No Supervisado

- Definición: ausencia de información previa sobre el vector objetivo.
- Tipos de aprendizaje no supervisado:
 - Reducción de dimensionalidad.
 - Clustering (agrupación)

La maldición de la dimensionalidad



Dimensionalidad desde la Psicometría

Objetivos del Análisis Factorial

- Objetivo: extraer la variabilidad de un conjunto finito de variables y transformarla en una serie de factores latentes.
- Busca identificar una serie de “factores causales” condicionales a un conjunto de datos donde la cantidad de dimensiones es sustancialmente superior a los registros existentes.

Análisis Factorial

$$X_1 = \lambda_{10} + \lambda_{11}f_1 + \dots \lambda_{1k}f_k + \varepsilon_1$$

$$X_2 = \lambda_{20} + \lambda_{21}f_1 + \dots \lambda_{2k}f_k + \varepsilon_2$$

⋮

$$X_N = \lambda_{n0} + \lambda_{n1}f_1 + \dots \lambda_{nk}f_k + \varepsilon_n$$

Carga Latente
Constante

Carga Latente
del Ítem

Ítem

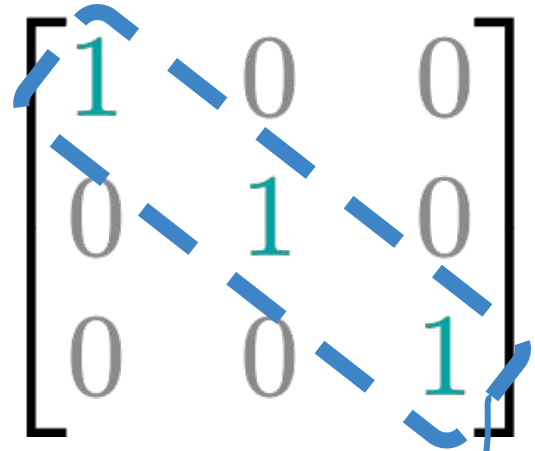
Error de Medición

Pasos de Implementación

- Definimos una batería de preguntas:
 - Observamos la escala de los datos. Si las escalas son heterogéneas, podemos estandarizar.
- Eliminamos / imputamos valores perdidos.
- Generamos métricas de validación.
- Evaluamos la cantidad de factores.
- Inferimos características de los factores

Prueba de Esfericidad de Bartlett

- ¿Es nuestra matriz de correlaciones una matriz de identidad?
- Comprobamos mediante una prueba de hipótesis:
 - Hipótesis Nula: La matriz de correlaciones es una matriz de identidad.
 - Hipótesis Alternativa: La matriz de correlaciones no es una matriz de identidad.

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$


Diagonal Principal

Prueba Kaiser-Meyer-Olkin

- ¿Son las correlaciones parciales entre dos ítems cercanas a 0?

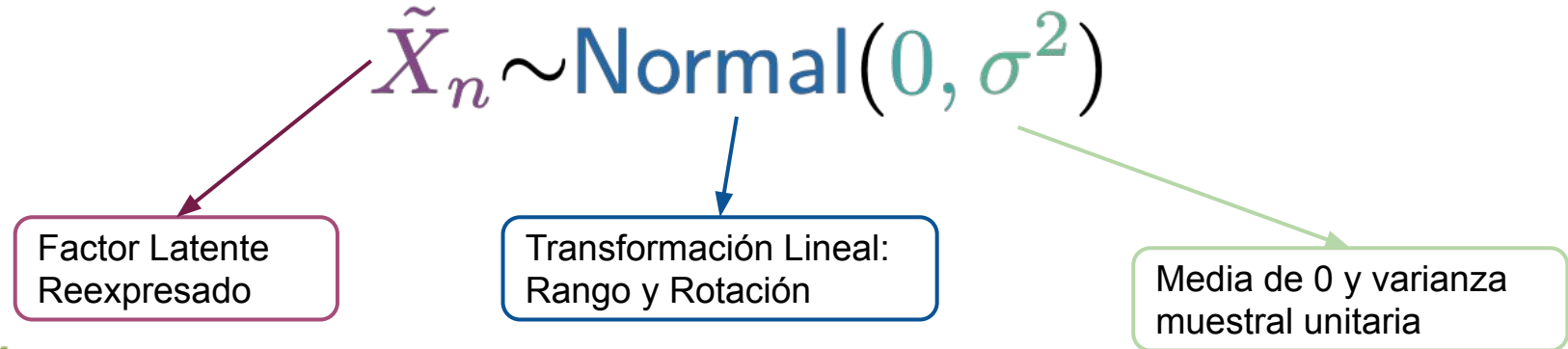
$$\text{KMO} = \begin{cases} > .7 & \text{Existe por lo menos un factor latente} \\ < .7 & \text{No existen factores latentes} \end{cases}$$

Definiendo cantidad de factores

- ¿Qué tantos factores necesitamos?
- Analizamos los eigenvalues de la descomposición de la matriz.
- Criterio: $\lambda > 1$
- Las cargas no tienen identificación única:
 - Rotación Ortogonal: No permiten correlacionar factores.
 - Rotación Oblicua: Permite correlacionar factores.
- Podemos utilizar un scree plot para definir la cantidad de factores.

Interpretación de factores

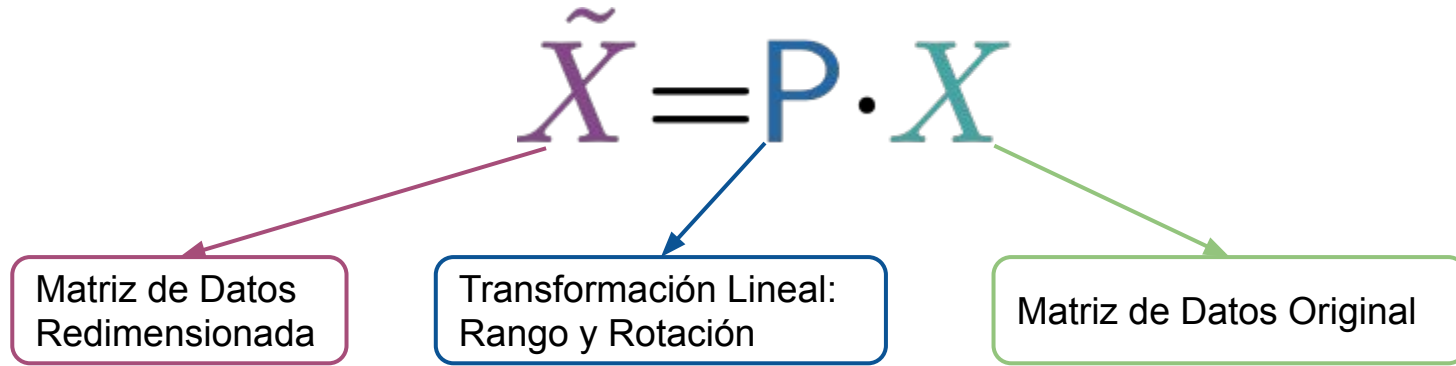
- Nos fijamos en la magnitud y dirección de las cargas:
 - Magnitud: Predominancia en el factor latente.
 - Dirección: Asociación con el factor latente.
- También podemos extraer la varianza explicada por cada dimensión latente.
 - Punto a considerar: los métodos de aprendizaje no supervisado no están exentos del trueque entre sesgo y varianza.
- También se pueden inferir los puntajes predichos del factor:



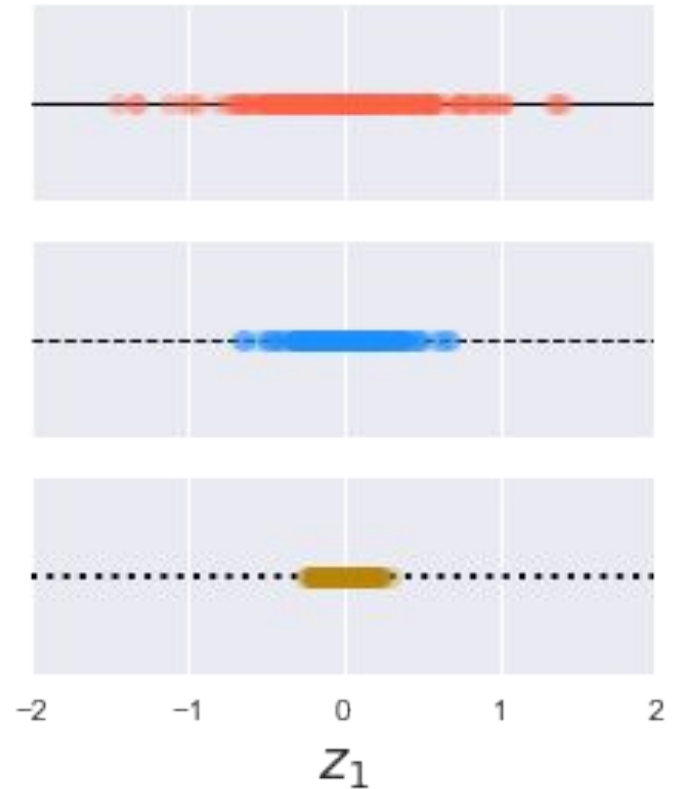
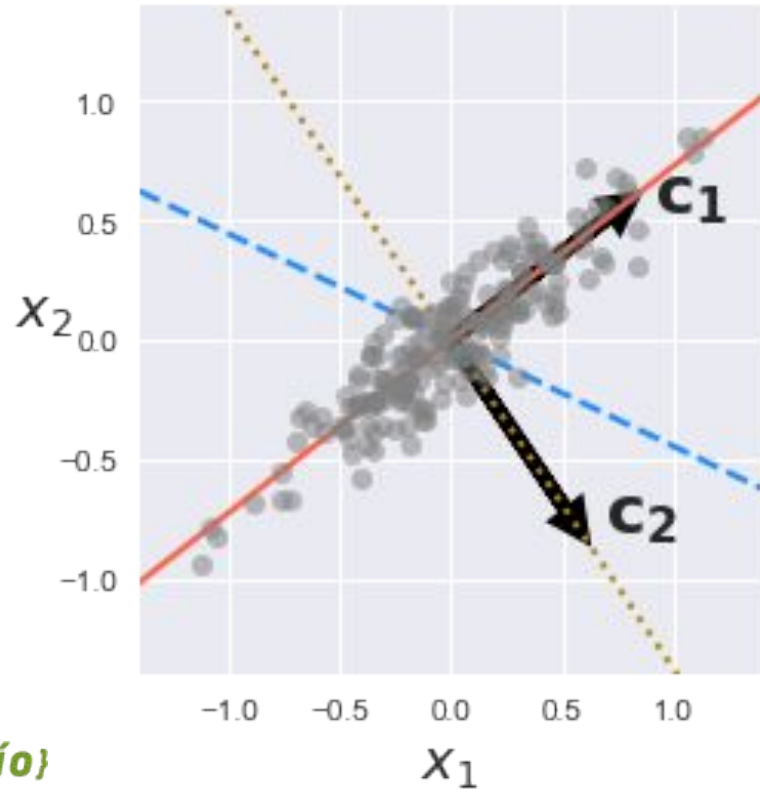
Dimensionalidad desde Machine Learning

Análisis de Componentes Principales

- El Análisis de Componentes Principales es un método de transformación lineal ortogonal que mapea datos en un nuevo sistema de coordenadas, maximizando la varianza de las extrapolaciones

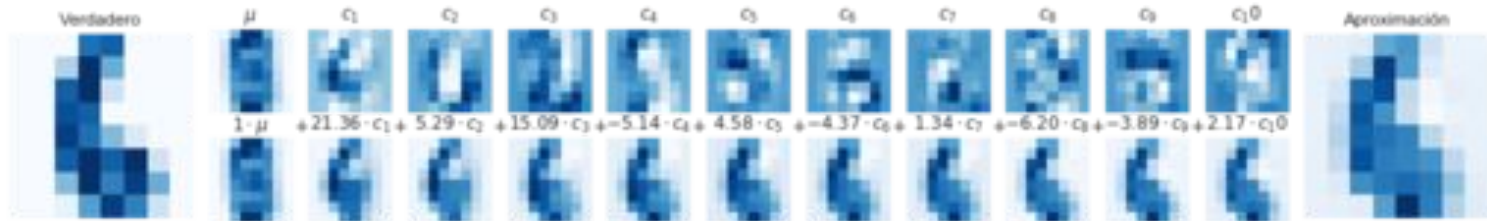


Análisis de Componentes Principales



Análisis de Componentes Principales

$$\text{imagen}(x) = \mu + \sum_{i=1}^{10} \lambda_i x_i$$



Agrupación desde Machine Learning

Rudimentos de la Agrupación

- Objetivo: Dividir los datos en una serie de grupos en base a una medida de similitud entre los puntos.
- Limitante no supervisada: No tenemos certeza sobre la cantidad de grupos

KMeans

- Pasos de KMeans:
 - Asignar una cantidad de clusters.
 - Inferir la media para cada cluster determinado en el espacio.
 - Para cada observación en el espacio, asignar el cluster más cercano.
 - Actualizar la media inferida para todos los puntos.
 - Iterar hasta que se satisfaga algún criterio.

Inercia del Modelo

- Medir la distancia promedio entre cada punto y el centroide mediante la suma de distancias (euclídeas) cuadráticas.
- Buscamos obtener una reducción sustancial de las distancias cuadráticas entre los puntos y sus centroides.

{desafío}
latam_

*Academia de
talentos digitales*

www.desafiolatam.com