Taller # 4

Business Intelligence

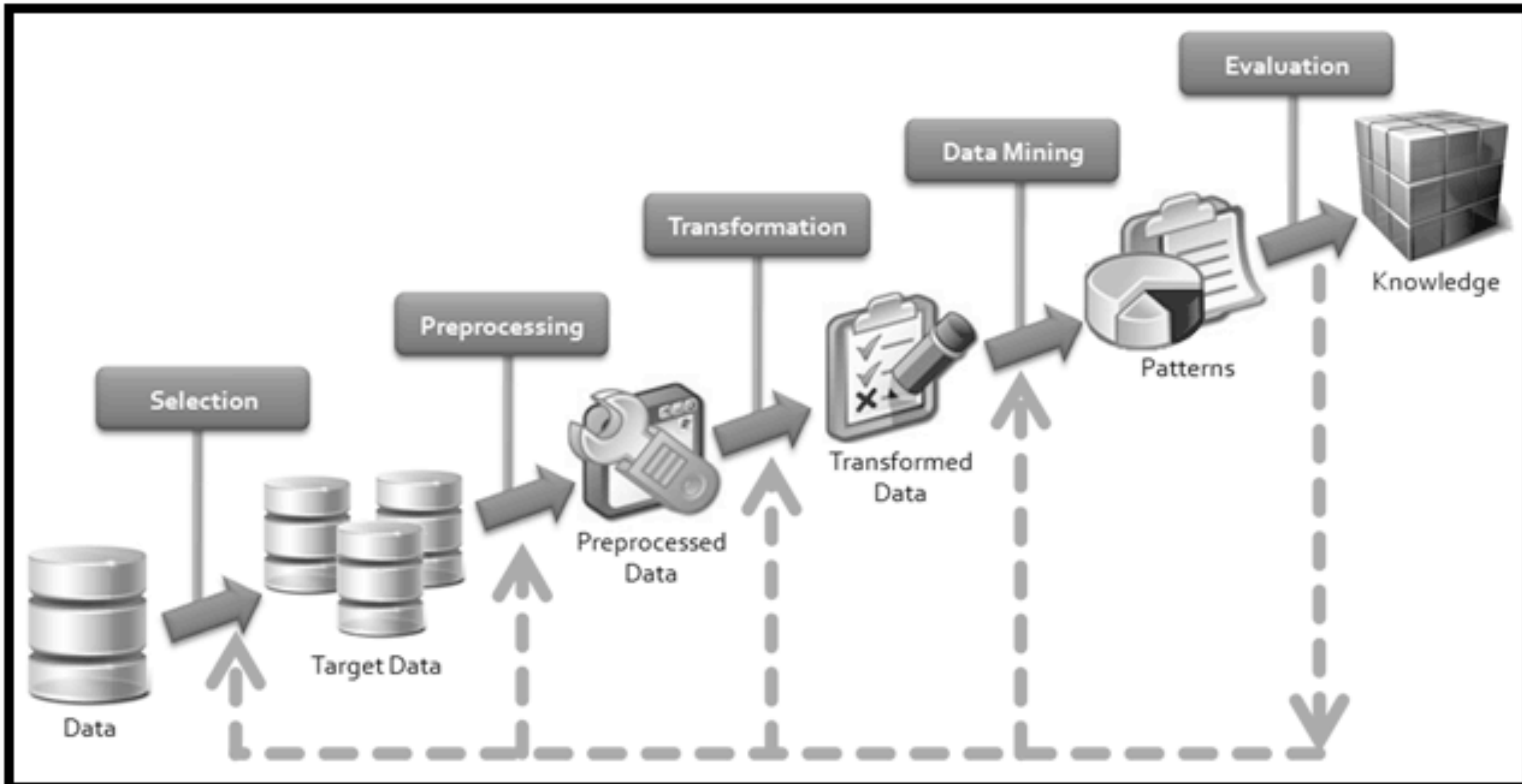
Carlos Reveco
creveco@dcc.uchile.cl

Cinthya Vergara
cvergarasilv@ing.uchile.cl

Agenda

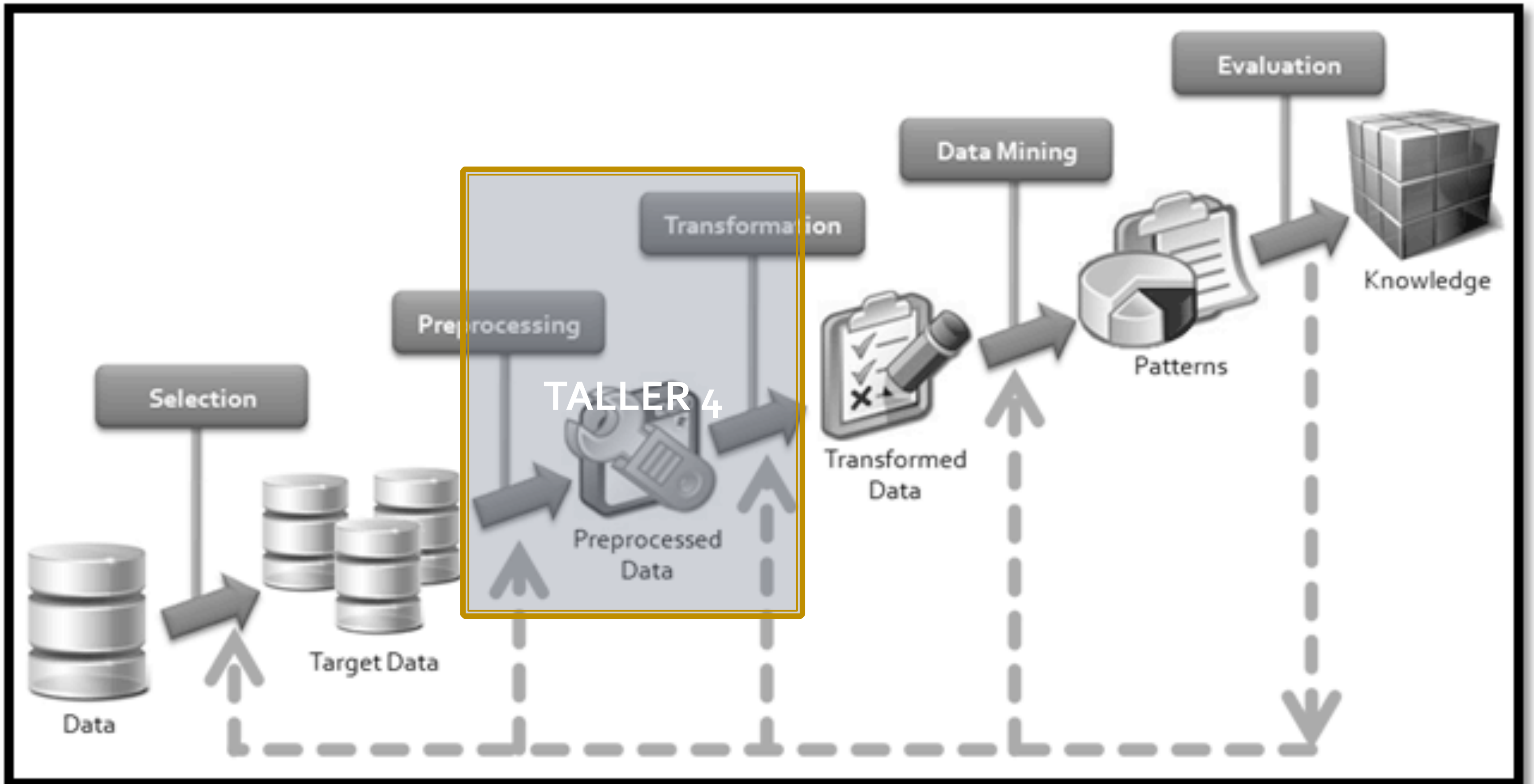
- Taller #4 – Extracción de Características
  - Principal Component Analysis (PCA)
    - Definiciones y modelamiento
  - Kernel-PCA
    - Funciones de Kernel
  - Independent Component Analysis (ICA)
    - Definiciones y modelamiento

Proceso KDD



Knowledge **D**iscovery in **D**atabases → **KDD**

Proceso KDD



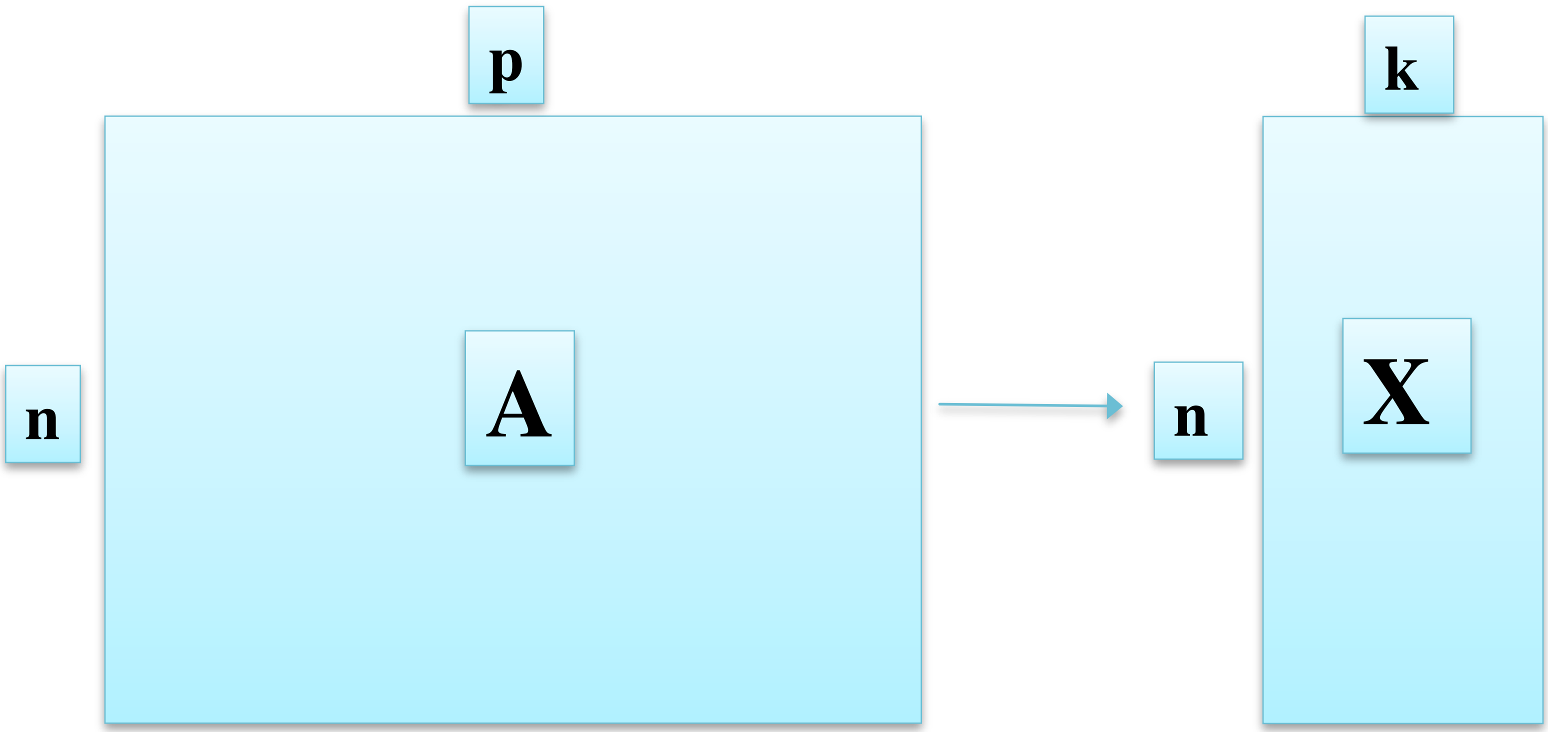
Knowledge **D**iscovery in **D**atabases → **KDD**

Taller #4

Extracción de Características

Extracción de Atributos
Reducción de dimensionalidad

- Idea principal:



Extracción de Atributos
Reducción de dimensionalidad

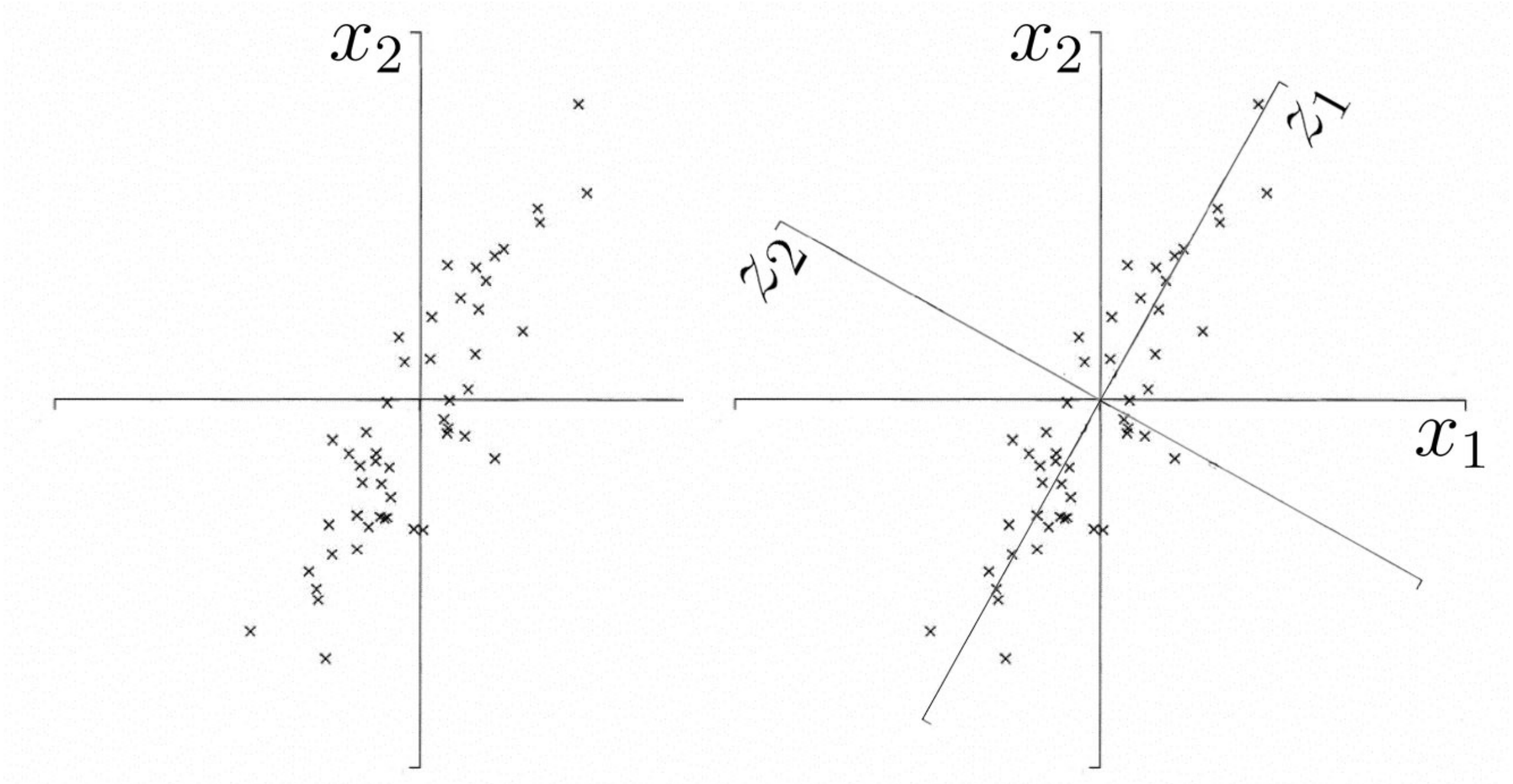
- Conjunto inicial con **p** variables
- Objetivo: Encontrar un nuevo conjunto de **k** variables (sintetizadas o compuestas del conjunto inicial), que representen la misma información.
- Tener cuidado con:
  - Claridad en la representación
  - Sobre-simplificar la información o perdida de información relevante.

Extracción de Atributos
Reducción de dimensionalidad

- Algunos métodos:
  - Análisis de componentes principales
    - **PCA, Kernel PCA**
  - Análisis Factorial
  - Linear Discriminant Analysis, Kernel Discriminant Analysis
  - Análisis de componentes independientes
    - **ICA**, FastICA
    - Exploratory Projection Pursuit (EPP)
  - Análisis de variables latentes
    - Singular Value Decomposition (SVD)
  - Escalamiento multidimensional
    - Sammon's mapping, Encoding/Decoding NeuralNets, etc
  - Self Orginizing Maps (SOMs)

Principal Component Analysis PCA

- Idea principal:



Principal Component Analysis PCA (2)

- En PCA deseamos determinar un nuevo espacio de k variables representado por la combinación lineal no correlacionada (ortogonal) de las p variables originales
- En general, cumplan con:
  - **Maximizar la varianza** de los datos en cada componente principal
  - Todas las **componentes son independientes (ortogonales)** entre ellas.

Principal Component Analysis PCA (3)

- Dado el conjunto de n observaciones de **p** variables

$$x = (x_1, x_2, \ldots, x_p)$$

- Definimos la primera componente principal según la siguiente transformación lineal

$$z_1 \equiv a_1^T x = \sum_{i=1}^{p} a_{i1} x_i \, , a_1 = (a_{11}, a_{21}, \ldots, a_{p1})$$

- Donde el vector a_1 es elegido de tal manera que se maximice la varianza de z1

$$\mathrm{var}[z_1]$$

- Sujeto a que $a_1^T a_1 = 1$

Principal Component Analysis PCA (4)

- De esta manera, la k-esima componente principal de la muestra se puede representar por

$$z_k \equiv a_k^T x = \sum_{i=1}^{p} a_{ik} x_i , a_k = (a_{1k}, a_{2k}, \ldots, a_{pk})$$

- Donde el vector a_k es elegido maximizando la varianza

$$\text{var}[z_k]$$

- Sujeto a que

$$\text{cov}[z_k, z_l] = 0, k > l \geq 1$$

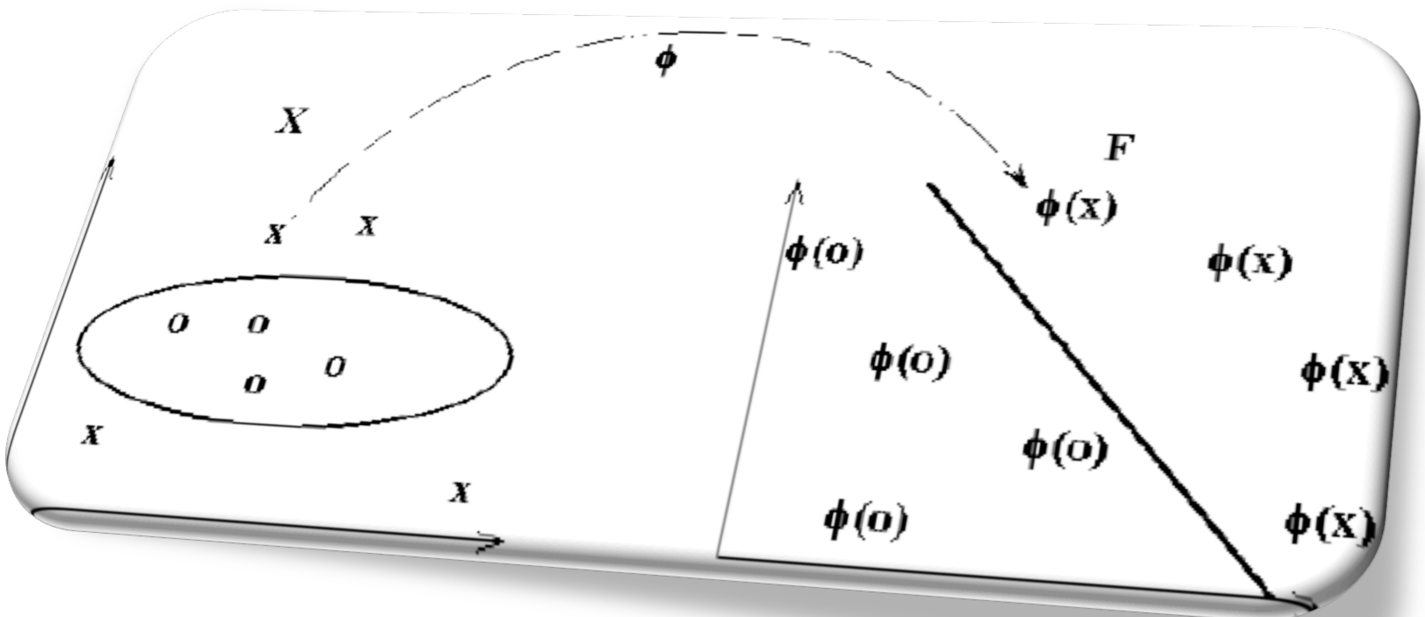
$$a_k^T a_k = 1 \qquad a_k^T a_l = 0, \forall k, l, k \neq l$$

Principal Component Analysis PCA (5)

- La **primera componente principal** es aquella que representa la **máxima variabilidad** con respecto a los atributos originales.
- Las **siguientes componentes**, están ordenadas según la **ortogonalidad de la componente principal anterior**.
- Es importante notar que se **asume linealidad** en la relación entre los atributos (dada la transformación lineal que se asume a-priori)
- PCA se puede potenciar utilizando **Kernel methods para incorporar relaciones no-lineales**.

Kernel Methods
"kernel trick"

- Cuando buscamos un hiperplano en un conjunto de datos no linealmente separable, definimos una transformación que mapea los datos en otro espacio. ("Kernel Trick")



Kernel Methods
Función de Kernel

- Se define la función de "mapeo":

$$\begin{aligned}\phi: \quad & \mathbb{R}^n \rightarrow F \\ & x \rightarrow \phi(x)\end{aligned}$$

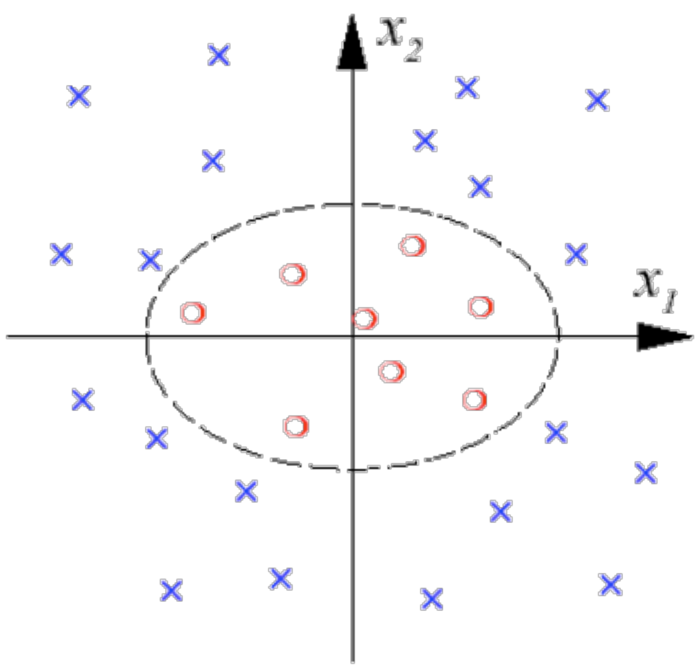
- En base a la anterior, se define la función de kernel:

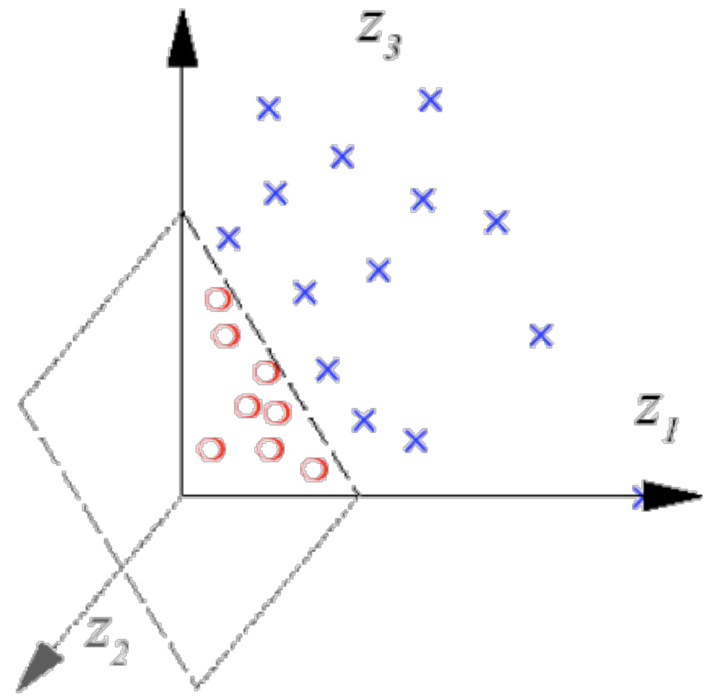
$$k(x, x') := (\phi(x) \cdot \phi(x'))$$

$$\begin{aligned}k: \quad & \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R} \\ & (x, x') \rightarrow k(x, x')\end{aligned}$$

Kernel Methods
Función de Kernel

$$\Phi : \mathbb{R}^2 \to \mathbb{R}^3$$
$$(x_1, x_2) \mapsto (z_1, z_2, z_3) := (x_1^2, \sqrt{2}\, x_1 x_2, x_2^2)$$





Fuente: Kernel Tricks, means and ends. Bernhard Schölkopf, Max Planck Institute for Biological Cybernetics

Kernel Methods
Funciones de Kernel

- Kernel polinomial

$$k(x, x') = (x \cdot x')^d$$

- Kernel base radial (RBF)

$$k(x, x') = exp(-\gamma||x - x'||^2), \gamma > 0$$

- Kernel base radial gaussiana (GRBF)

$$k(x, x') = exp(-\frac{||x - x'||^2}{2\sigma^2})$$

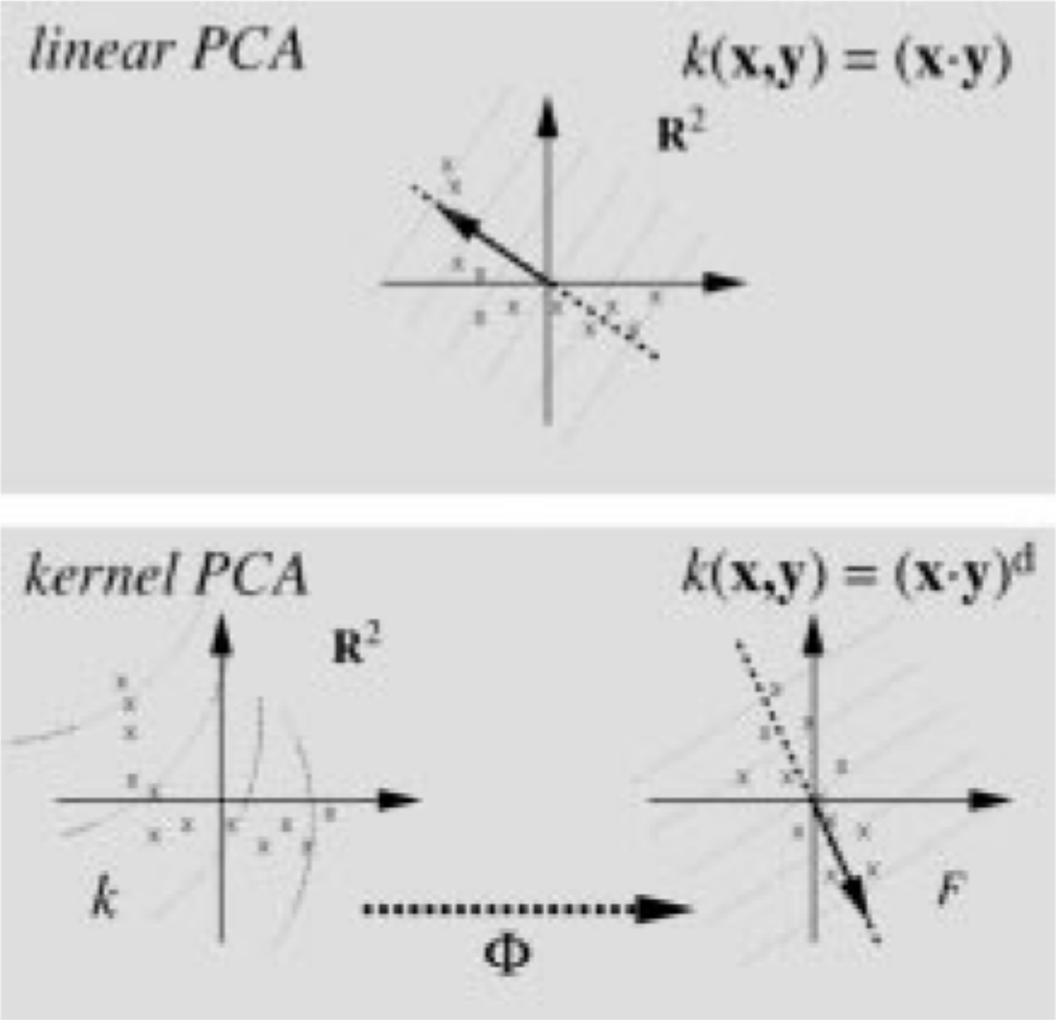
- Kernel tangente hiperbólica

$$k(x, x') = \tanh(\kappa x \cdot x' + c)$$

Kernel-PCA

- Utilizando el **“kernel Trick”**, la transformación lineal de PCA se asume en el espacio característico, **no lineal con respecto a los atributos originales**.
- **No es necesario determinar la función de mapeo** ya que basta con utilizar la matriz de Kernel para resolver el algoritmo PCA anteriormente descrito.
- De esta manera, podemos **capturar dependencias no lineales entre los atributos originales** de la base de datos.
- **Desventaja**: debemos asumir a-priori cual es la interacción no lineal, i.e. debemos elegir la **función de kernel**.

Kernel-PCA (2)

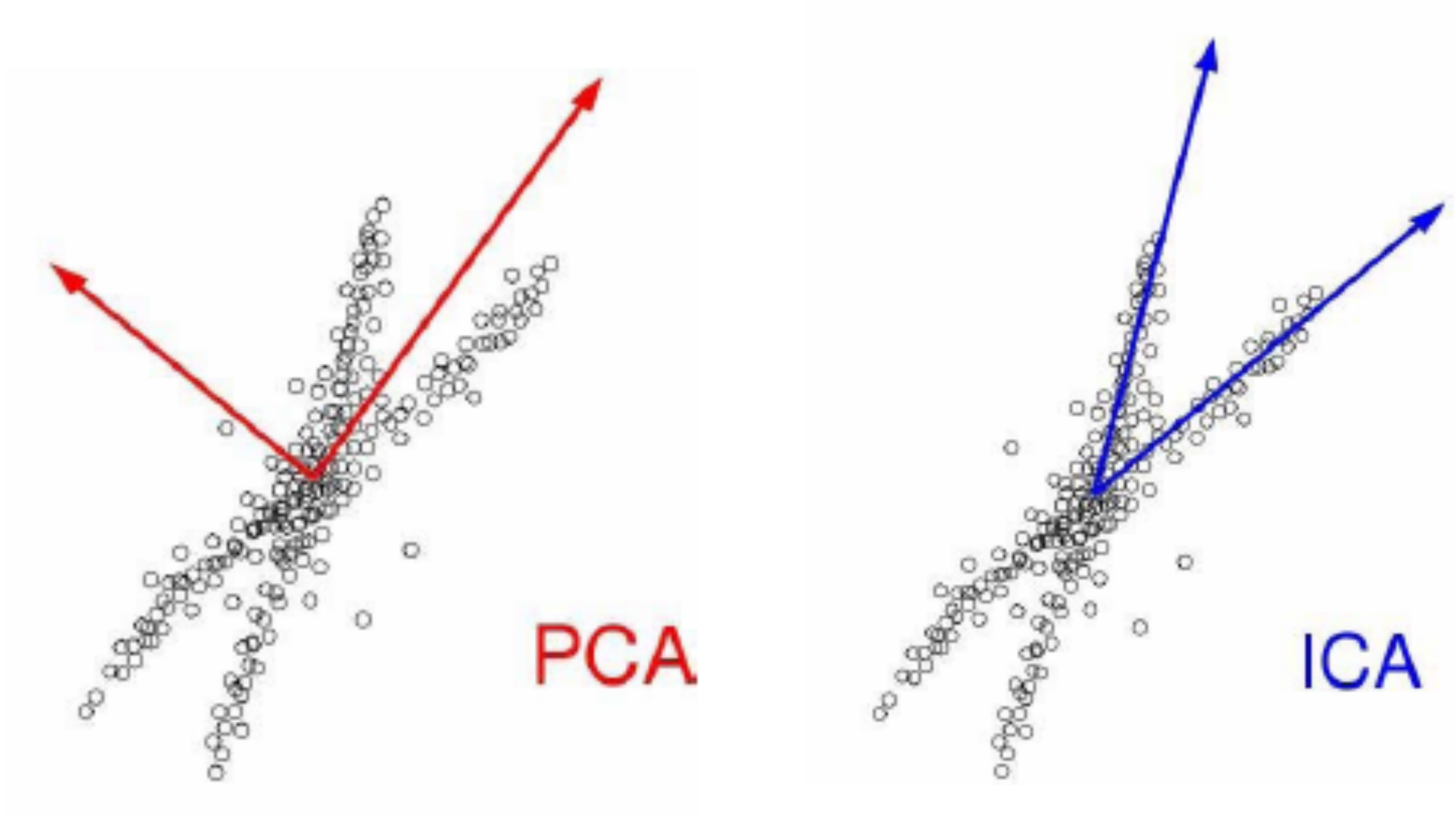


linear PCA
$k(\mathbf{x},\mathbf{y}) = (\mathbf{x}\cdot\mathbf{y})$
$\mathbf{R}^2$
kernel PCA
$k(\mathbf{x},\mathbf{y}) = (\mathbf{x}\cdot\mathbf{y})^d$
$\mathbf{R}^2$
k
Φ
F

Independent Component Analysis ICA

- Técnica estadística computacional que permite extraer factores latentes que entreguen una **representación "interesante" del conjunto de información inicial**.
- Al igual que PCA, asume combinaciones lineales entre los atributos originales, pero se diferencia en que se desea **minimizar la información mutual (mutual information)**

Independent Component Analysis ICA (2)



Independent Component Analysis ICA (3)

- A grandes rasgos, en ICA se desea determinar un nuevo conjunto de atributos z

$$z_k \equiv a_k^T x = \sum_{i=1}^{p} a_{ik} x_i , a_k = (a_{1k}, a_{2k}, \ldots, a_{pk})$$

- Tal que permita **minimizar la información mutua entre todos los atributos z**

$$I(z) = \sum_{i=1}^{k} H(z_i) - H(z)$$

Independent Component Analysis ICA (4)

- El principal objetivo de ICA es **buscar la independencia estadística** entre los atributos extraídos.

$$f_{12\ldots k}(z_1, z_2, \ldots, z_k) = f_1(z_1) \cdot f_2(z_2) \cdot \ldots \cdot f_k(z_k)$$

- FastICA es una implementación de ICA que asume una forma funcional y una serie de parámetros sobre los elementos a utilizar en el problema de optimización.

Taller #4

Ejercicio práctico PCA con Rapid Miner

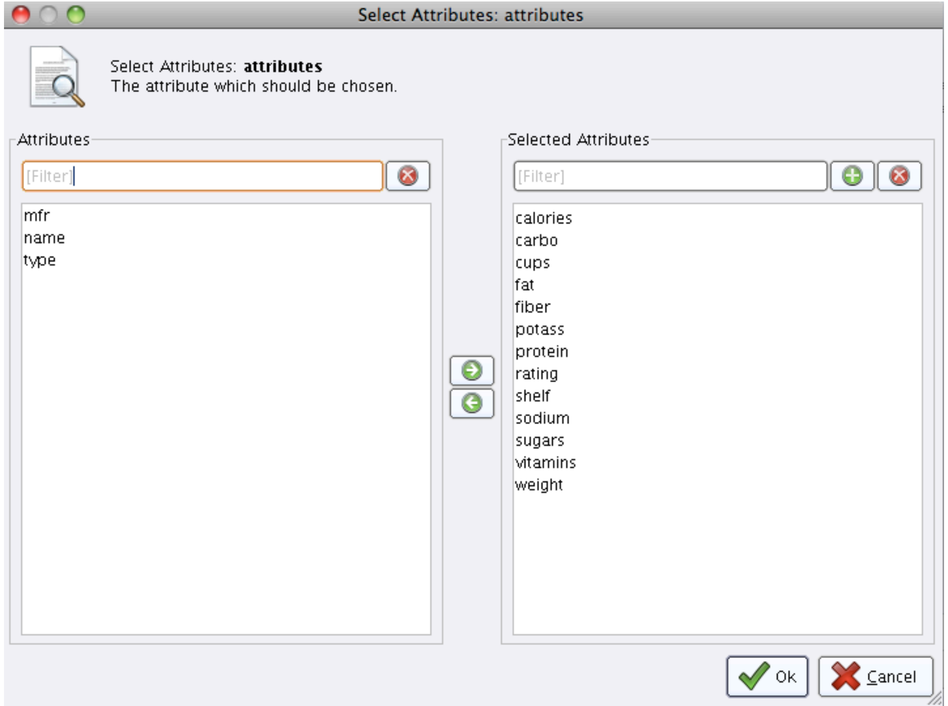
Paso 1

- Ingresar Data
  - Descargar archivo cereales.xls de U-cursos

- Revisar archivo en Rapid Miner
  - Este archivo incluye información nutricional de 77 cereales.
  - Tiene 15 variables, incluyendo 13 numéricas.
  - El objetivo es reducir la cantidad de atributos mediante PCA.

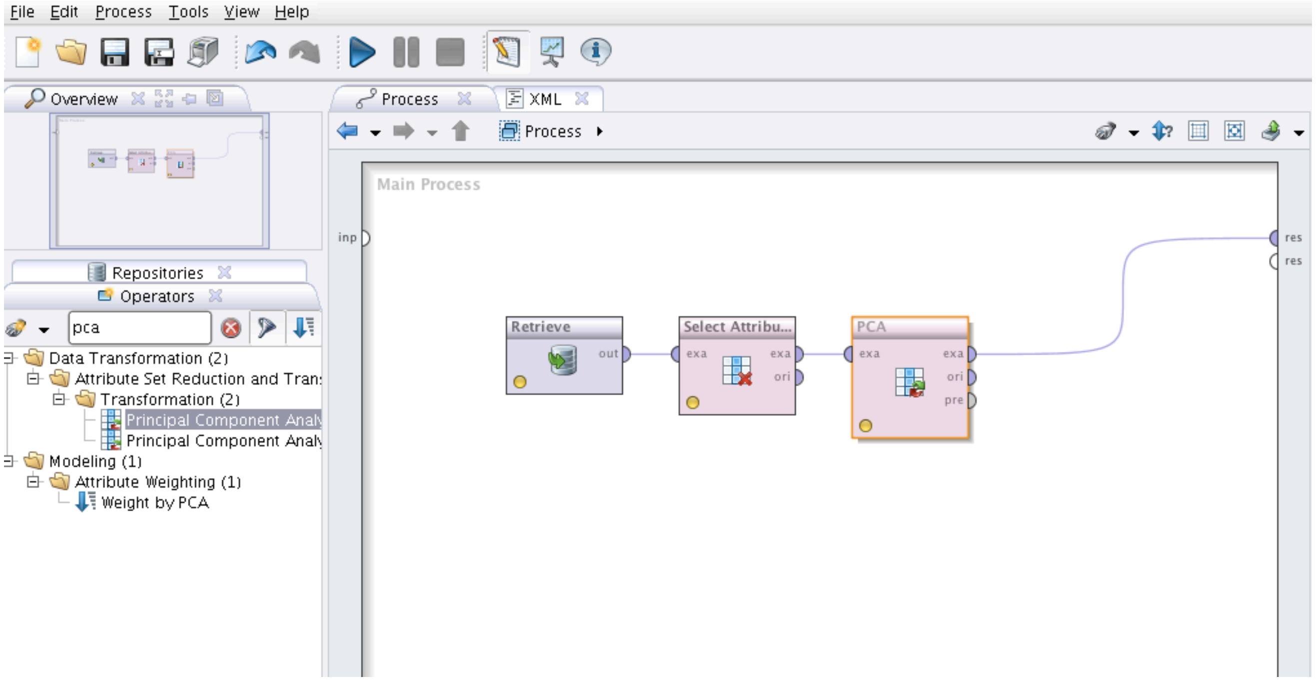
Paso 2

- Eliminar Variables Categóricas
- Usar operador “select Atributtes”
- No seleccionar columnas nombre, mfr y type



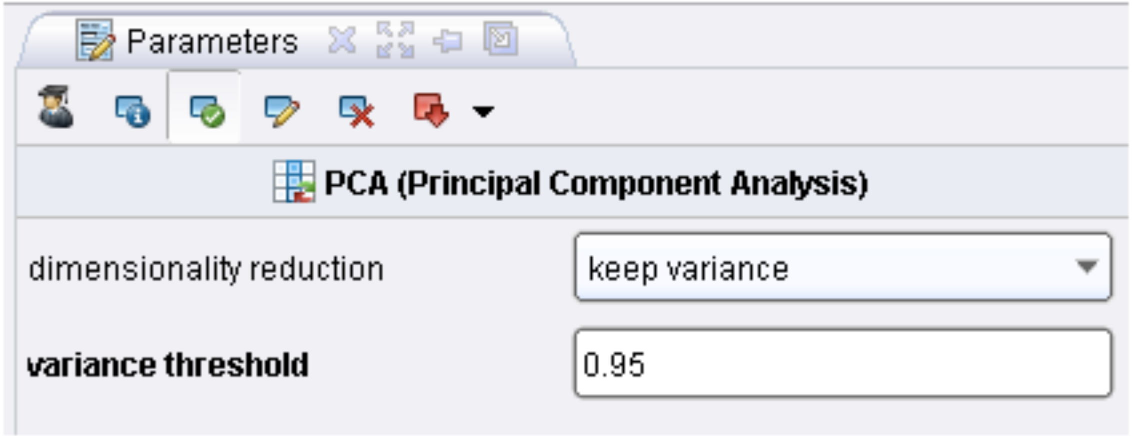
Paso 3

- Elegir Operador PCA



Paso 3

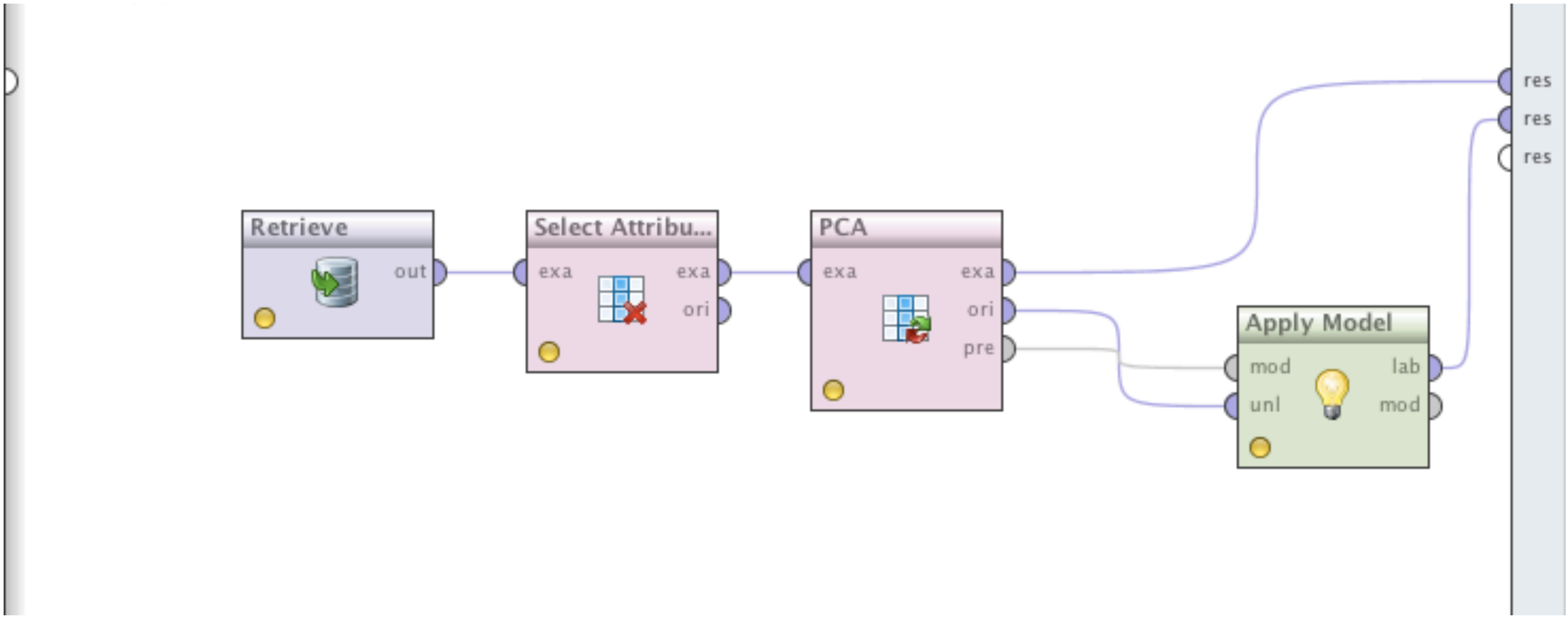
- Utilizar un umbral del 95% de la varianza.



Paso 4

- Aplicar Modelo
  - Utilizar operador “Apply Model”
  - del Operador PCA “**ori**”ginal a Apply Model “**unl**"abled
  - Del operador PCA, "**pre**"processing a Apply Model "**mod**"el
  - Del operador Apply Model, output "**lab**” a "**res**” y output "**mod**" a "**res**" port

Ver Resultados



Solucionar Errores

- Las componentes principales son las de mayor valor y no considera que están en diferentes escalas.
- Solución: Normalizar
- Re-interpretar resultados.

Taller # 4

Business Intelligence

Carlos Reveco
creveco@dcc.uchile.cl

Cinthya Vergara
cvergarasilv@ing.uchile.cl