

# Machine Learning para Ciencias Sociales

## Unsupervised Learning

Fecha de Entrega: 4 de Marzo (primer día de exposiciones)

Profesor: Pedro Marco Achanccaray Diaz

Cualquier duda o comentario escribir a: Pedro Achanccaray (pachanccarayd@uni.pe)

## Principal Component Analysis – PCA

### Ejercicio 1.

Generar una base de datos simulada con 60 observaciones y 50 atributos (usar la función *rnorm()* para generar la base de datos):

- Realizar PCA en las 60 observaciones.
- Graficar las 2 primeras componentes principales
- Mostrar el porcentaje de información contenida en cada componente principal
- Mostrar el porcentaje acumulado de información.

## *k*-means

### Ejercicio 2a.

Dadas las siguientes observaciones:

- Graficar las observaciones.
- Realizar *k*-means con  $k=2$ .
- Graficar las observaciones asignando un color para cada cluster.

| Obs. | $X_1$ | $X_2$ |
|------|-------|-------|
| 1    | 1     | 4     |
| 2    | 1     | 3     |
| 3    | 0     | 4     |

|   |   |   |
|---|---|---|
| 4 | 5 | 1 |
| 5 | 6 | 2 |
| 6 | 4 | 0 |

### Ejercicio 2b.

Usando la base de datos del Ejemplo 2:

- Realizar  $k$ -means con  $k=5$  y variar el método para la distancia: euclidean, maximum, manhattan, canberra, pearson, correlation, spearman, kendall.
- Comparar los resultados obtenidos usando los diferentes métodos para la distancia.

## Hierarchical Clustering

### Ejercicio 3.

Usando la base de datos USArrests:

- Realizar Hierarchical Clustering con los siguientes tipos de *linkage*: Complete, Single y Average.
- Graficar los Dendrogramas para cada tipo de *linkage*.
- Cortar los Dendrogramas a cierta altura para generar 3 clusters.
- Mostrar que estados pertenecen a cada cluster.