

# Taller #4 de Métodos Computacionales

## FISI 2028, Semestre 2014 - 10

Profesor: Jaime Forero

Miércoles 26 de Febrero, 2014

### Importante

- Todos los programas que solucionan esta tarea deben encontrarse en un repositorio en github con un commit final hecho antes del medio día del viernes 14 de Marzo. Cada literal debe resolverse con un notebbook de ipython por separado.
- La nota máxima de este taller es de 140 puntos. Los puntos indicados en cada literal solamente se otorgan si el da los resultados esperados según la descripción de cada punto.

### 1. Homicidios

En el siguiente repositorio:

<https://github.com/finiterank/homicidios/>

se encuentra el archivo `data/homicidios.1990.a.2013.csv` que contiene los datos de homicidios en todos los municipios de Colombia entre 1990 y 2013.

- a) (25 puntos) Realice un descomposición en PCA de todos los historiales de homicidios expresados en la tasa de homicidios por cada 100 mil habitantes. ¿Es posible reducir la dimensionalidad de los datos?
- b) (25 puntos) Expresa ahora cada una de las historias de tasa de homicidios como la lista de coeficientes que son necesarios para expresarlas como combinación lineal de los autovectores encontrados en el ítem anterior. Haga un análisis de K-clustering sobre estos datos para  $K = 2, 3, 4, 5$ . ¿Es posible decir que los datos se separan en clusters?

### 2. Pobreza

En el siguiente repositorio:

[https://github.com/finiterank/mapa\\_pobreza\\_colombia](https://github.com/finiterank/mapa_pobreza_colombia)

se encuentra el archivo `data/pobreza.csv` con datos multidimensionales de pobreza que incluyen acceso a salud, calidad escolar, empleo informal, etc.

- a) (25 puntos) Realice un estudio PCA sobre las variables *Analfabetismo*, *Inasistencia Escolar*, *Alta dependencia económica*, *Sin Fuente Agua Mejorada*. ¿Es posible reducir la dimensionalidad de los datos?
- b) (25 puntos) ¿Existen clusters en el espacio de esas mismas variables?

### 3. Saber, Pobreza y Homicidios

(20 puntos)

En el siguiente repositorio

[https://github.com/finiterank/saber\\_notebooks](https://github.com/finiterank/saber_notebooks)

Se encuentra el archivo `rank.2011.csv` que tiene resultados sobre las pruebas Saber en todos los municipios de Colombia para el año 2011.

Para este punto también tenga en cuenta los datos de los puntos anteriores.

- a) (40 puntos) Tomando en cuenta los resultados de las pruebas saber (puntajes de matemáticas y lenguaje), los datos de *Alta dependencia económica*, *Inasistencia Escolar*, *Sin Fuente Agua Mejorada*. ¿Hay alguna diferencia en la distribución de estas variables entre los diferentes clusters definidos de acuerdo a los homicidios?