

 <p>Pontificia Universidad <b>JAVERIANA</b> Bogotá</p>	<p>PONTIFICIA UNIVERSIDAD JAVERIANA</p> <p>Maestría en Inteligencia Artificial</p> <p>Aprendizaje de máquina</p> <p>Taller 1 - 202330</p>
---	---

## Taller 1 – Regresión lineal y logística

Lea con atención el enunciado y entregue un archivo .zip que tenga los elementos solicitados en las reglas de juego del curso explicadas en la primera sesión, en particular no olvide incluir:

Jupyter notebook desarrollando los elementos pedidos en cada punto. (Pueden ser varios). Entregar en carpeta *code*.

Reporte escrito con el análisis de cada punto. Recuerde las pautas de informes escritos (IEEE Transactions on Artificial Intelligence). Recuerde escribir introducción a su documento y para cada punto explique su metodología, experimentación, resultados y conclusiones. Realice también conclusiones generales del taller (¿Qué funcionó?, ¿Qué no funcionó?, ¿Qué aprendió?).

**Restricción para este taller: Todos los métodos implementados en el taller deben utilizar solamente la librería NumPy, y todos los cálculos deben estar vectorizados. Solo se permite el uso de la librería Pandas para cargar los datos, su exploración y visualización, y su conversión en matriz de NumPy. No se permite el uso de otras librerías diferentes a las mencionadas anteriormente.**

### Caso de estudio: Costos médicos

Es este caso de estudio usaremos los datos de costos de una compañía de seguros médicos. El dataset cuenta con información acerca de los asegurados (edad, sexo, bmi, hijos, fuma, región) y los costos que tuvo para la compañía esa persona.

Algunos costos en los que incurre la compañía de seguros pueden estar asociados a eventos fortuitos, pero se espera que ciertas poblaciones (personas de más edad o fumadores por ejemplo) impliquen mayores costos para las compañías.

En este caso seguiremos usando el dataset de costos médicos que utilizamos para el primer laboratorio.

### Punto 1 (50%)

Utilizando las variables independientes, cree un modelo de regresión lineal que permita predecir el costo asociado a un paciente dadas sus características. Además de la forma de la hipótesis (selección de variables, interacciones, polinomios, transformaciones) debe probar diferentes hiper-parámetros de regularización.

Items de evaluación:

- Implementación del método 25%
- Experimentación y resultados 10%
- Conclusiones 15%

### Punto 2 (50%)

Utilizando las variables independientes (sin incluir costos), cree un modelo de regresión logística que permita predecir si el paciente fuma o no. Además de la forma de la hipótesis (selección de variables, interacciones, polinomios, transformaciones) debe probar diferentes hiper-parámetros de regularización.

Items de evaluación:

- Implementación del método 25%
- Experimentación y resultados 10%
- Conclusiones 15%