

Proyecto Final

Métodos Numéricos y Optimización

Equipo1

Paulina Gómez-Mont Wiechers 113018

Rodrigo Floriano Verástegui 111954

Carlos Alberto García 181987

1. Descripción del proyecto

Este proyecto aborda el problema descrito en la competencia de Kaggle llamada “Reducing Commercial Aviation Fatalities”. El propósito de esta competencia es identificar si los pilotos de aviones presentan algún estado cognitivo que pueda ser peligroso en diversas simulaciones de vuelo, así como cuáles son sus causas.

La información disponible está conformada por datos psicológicos de 18 pilotos que fueron sujetos a distracciones en un ambiente de simulación de vuelo. El conjunto de entrenamiento está conformado por experimentos controlados, mientras que el conjunto de prueba consiste en datos en vuelos completos.

Con el objetivo de aislar el estado cognitivo de los pilotos ante situaciones, los pilotos fueron expuestos a distracciones con la intención de llevarlos a alguno de los siguientes estados cognitivos:

- **Atención Canalizada (CDA)** - Es la concentración en una tarea excluyendo todas las demás. Se induce obligando a los pilotos a jugar videojuegos.
- **Atención Desviada (DA)** - Es la desviación de la atención debido a acciones o procesos de pensamiento asociados a la toma de decisiones. Se induce resolviendo periódicamente problemas matemáticos antes de regresar a la tarea monitoreada.
- **Sorpresa (SS)** - Es inducido al presentar al sujeto de estudio cortes de películas en donde el observador pueda asustarse/sorprenderse.

Cada experimento consistió en monitorear en un par de pilotos de cada grupo (**crew**) expuestos *a priori* o inducidos desde el **estado base A** a situaciones o experimentos que pudieran inducir a alguna condición cognitiva en los pilotos. Por ello, en cada observación se tiene solamente la observación de un ejercicio asociado a los estados CDA, DA o SS. En el evento final se observa si el individuo alcanzó alguno de los estados CDA, DA, SS o A, en donde A implica el fallo en conseguir la condición cognitiva esperada. Por ejemplo, en el experimento CDA, los pilotos estaban en un estado base A o en el estado CDA y se les indicó jugar videojuegos, por lo que se espera que presenten la condición CDA o la ausencia de ella (A).

Por otro lado, el conjunto de prueba contiene una simulación de vuelo completa, durante la cual se levantó información del piloto, en donde podrían experimentar cualquiera de los estados CDA, DA, SS o A pero nunca más de uno a la vez.

2. Objetivo

El objetivo del proyecto, desde la perspectiva del análisis y optimización numérica, es aplicar un **regresión logística multinomial** en donde se pretende clasificar cada evento observado en las categorías CDA, DA, SS o A. Los problemas de optimización de esta forma parten de minimizar la **devianza** generada por el modelo propuesto y por los datos disponibles. En particular, la descripción del modelo es la siguiente:

Sean $g \in \{1, \dots, K\}$ las clases a las que pertenecen las observaciones $x^{(i)}$, $i = 1, \dots, N$ obtenidas de una vector aleatorio X . Las probabilidades de pertenecer a una clase, condicional a una observación x son: $p_g(x) = P(G = g|X = x)$, $g \in \{1, \dots, K\}$ en donde:

$$\begin{aligned}
p_1(x) &= \exp(\beta_{0,1} + \beta_{1,1}x_1 + \dots + \beta_{p,1}x_p) / Z \\
p_2(x) &= \exp(\beta_{0,2} + \beta_{1,2}x_2 + \dots + \beta_{p,2}x_p) / Z \\
&\dots \\
p_{K-1}(x) &= \exp(\beta_{0,K-1} + \beta_{1,K-1}x_2 + \dots + \beta_{p,K-1}x_p) / Z \\
p_K(x) &= 1/Z \\
Z &= 1 + \sum_{j=1}^{K-1} \exp(\beta_0^j + \beta_1^j x_1 + \dots + \beta_p^j x_p)
\end{aligned}$$

Por lo anterior, el problema de **minimización de devianza** mencionado está descrito como:

$$\min_{\beta} D(\beta) = -2 \sum_{i=1}^N \log p_{g(i)}(x^{(i)})$$

en donde $\beta = (\beta_0^1, \beta_1^1, \dots, \beta_p^1, \beta_0^2, \beta_1^2, \dots, \beta_p^2, \dots, \beta_0^{K-1}, \beta_1^{K-1}, \dots, \beta_p^{K-1})$.

3. Propuesta:

Para el proyecto se utilizarán dos métodos de optimización que utilizarán técnicas de cómputo numérico. Los dos métodos a utilizar (y sus respectivas rutinas) son:

1. **Método de descenso en gradiente.**

- 1.1 Cálculo de gradiente a través de diferencias centradas.
- 1.2 Método de bisección para encontrar tamaño óptimo del paso.

2. **Método de descenso con dirección de Newton**

- 2.1 Cálculo de gradiente a través de diferencias centradas.
- 2.2 Cálculo de matriz Hessiana por diferencias.
- 2.3 Método de bisección para encontrar tamaño óptimo del paso.
- 2.4 Solución de sistemas lineales por factorización QR de la matriz de segundas Hessiana.

Se propone realizar el proyecto utilizando el lenguaje de programación **R**. Dada la capacidad de procesamiento local y el tamaño del problema, se ha comenzado a explorar la opción de utilizar herramientas adicionales como **Docker**, **Spark**, **AWS** o **Dask**.