

Benemérita Universidad Autónoma de Puebla  
Facultad de Ciencias de la Computación

# Máquinas de aprendizaje

## Reporte: Otro ejemplo de regresión logística



# BUAP

**Docente: Abraham Sánchez López**

**Alumno**

Taisen Romero Bañuelos

**Matrícula**

202055209

## Regresión logística

El problema de las enfermedades cardíacas es muy cercano a nosotros como mexicanos, cada año aumenta a la par que los índices de obesidad, estrés y otras variables que considero pudieron haber aportado a este análisis breve.

La regresión logística es una técnica eficaz cuando tratamos de clasificar variables (en categorías, obviamente). El problema planteado busca detectar si una persona tiene probabilidades considerables de padecer una enfermedad cardíaca con base en unos cuantos predictores. También se plantea que la regresión logística no es de mucha ayuda ya que devuelve variables continuas y no categóricas, pero, ¿Qué hay de los árboles?, es decir, también realizan la clasificación devolviendo variables categóricas. Pues, podría deberse a la interpretabilidad de los datos, también al tamaño del dataset (en este caso es de 10,000 observaciones) e incluso a cómo se relacionan unas variables con otras. Los árboles son útiles para detectar relaciones complejas entre múltiples variables, y siendo franco, nuestras variables del corazón no parecen ser demasiado complejas. Pero, hablando en favor de los árboles, si consideramos la delicadeza de este problema de salud quizá si convenga considerar a los árboles para tener más exactitud (y ya que estamos, incluir otras variables para poder capturar más interacciones entre variables).

Retomando lo visto en el PDF, hay un hilo del que todavía hay que tirar. Se menciona que el café está asociado a una disminución de las probabilidades de padecer enfermedades cardíacas pese a que en un inicio parecía tener una mayor incidencia en las enfermedades cardíacas. En estadística, un **confusor** es una variable que está relacionada tanto con el predictor como con la variable de respuesta. En nuestro caso, los bebedores de café tienden a gastar más en comida rápida (posiblemente por un estilo de vida más acelerado o por preferencias personales). Y el gasto en comida rápida está directamente asociado con un mayor riesgo de enfermedades cardíacas. Cuando se observa la relación entre café y enfermedades cardíacas sin controlar otras variables, parece que el café está asociado con un mayor riesgo. Pero cuando el modelo ajusta por el gasto en comida rápida, se revela que no es el café el que está aumentando el riesgo, sino los hábitos alimenticios asociados con quienes consumen café.

Entonces, a partir de una variable podemos empezar a hacer un análisis de las personas más profundo, en relación a su estilo de vida. Las personas que consumen café pueden tener otros hábitos (como hacer ejercicio regularmente) que no se están midiendo en este análisis y que podrían contribuir a una mejor salud cardíaca, o bien pueden no hacerlo. Este análisis en favor del café considera que la persona en cuestión se modere al consumir dicha bebida, pero aquí es cuando la experiencia médica de mi

familia me hace cuestionar si en realidad es correcto que el café tiene un efecto protector sobre la salud cardiaca. Ese conocimiento empírico me ha hecho observar que las personas bebedoras de café, de hecho, tienen más riesgo de padecer enfermedades cardiacas ya que hablando nuevamente del estilo de vida, el café se usa como un recurso para sostener el cansancio y estrés que fisiológicamente ya no está siendo sostenible. Realmente es una explicación larga, pero considero importante no dejarse llevar por esa aparente satisfacción de haber entendido un confusor, porque bien podría tratarse del confusor de otro confusor (por así decirlo).