

## Tarea 1

KEVIN GARCÍA<sup>1,a</sup>, ALEJANDRO VARGAS<sup>1,b</sup>

<sup>1</sup>DEPARTAMENTO DE ESTADÍSTICA, UNIVERSIDAD DEL VALLE, CALI, COLOMBIA

---

### 1. Antecedentes trabajo de grado

### 2. Efron & Hastie (2017)

#### 2.1. Redes neuronales y aprendizaje profundo(cap 18)

#### 2.2. Bosques aleatorios e impulso(cap 17)

### 3. Lantz (2013) o Torgo (2017)

#### 3.1. Aprendizaje probabilístico - clasificación usando bayes ingenuo(cap 4 - Lantz 2013)

El aprendizaje probabilístico, expuesto por Lantz (2013) en el cuarto capítulo de su libro, es un tema que nos pareció muy interesante ya que utiliza la estadística bayesiana, más específicamente el teorema de Bayes, para tratar de predecir o clasificar, modelando la incertidumbre por medio de probabilidades y utilizando internamente pruebas de hipótesis.

Investigando un poco a fondo sobre este tema, nos dimos cuenta que es un método muy utilizado en la actualidad en diversos contextos por su uso práctico y los buenos resultados en la mayoría de los casos. Por ejemplo, se puede utilizar para algo muy trivial, como clasificar una persona en hombre o mujer basándonos en las características de sus medidas: peso, altura y número de pie; pero, también se puede usar para problemas más complejos y muy útiles como por ejemplo, el que plantea el autor, filtrar el spam de los teléfonos celulares para eliminarlos automáticamente y que esto no se convierta en una molestia para los usuarios. Este último ejemplo, nos llamó mucho la atención y de aquí que este tema ocupara la

---

<sup>a</sup>Universidad del Valle. E-mail: kevin.chica@correounivalle.edu.co

<sup>b</sup>Universidad del Valle. E-mail: jose.alejandro.vargas@correounivalle.edu.co

primera posición, ya que los llamados "spam" son en la mayoría de casos cadenas de texto haciendo publicidad o informándonos acerca de un "premio" que ganamos sin siquiera participar, entre otros. Con la proliferación de datos no estructurados, la clasificación de texto o la categorización de texto ha encontrado muchas aplicaciones en la clasificación de temas, análisis de sentimientos, identificación de autoría, detección de correo no deseado, etc. A pesar de que actualmente hay muchos algoritmos de clasificación disponibles, el Bayes ingenuo sigue siendo uno de los clasificadores más antiguos y populares. Por un lado, la implementación del bayes ingenuo es simple y, por otro lado, requiere menos cantidades de datos de entrenamiento. Sin embargo, el Bayes ingenuo se desempeña pobremente en comparación con otros clasificadores en la clasificación de texto. Como resultado, esto hace que el ingenuo clasificador de Bayes sea inutilizable a pesar de la simplicidad e intuición del modelo. Pero, creemos que sería muy útil tanto para nosotros como para el grupo en general, saber como funciona este método y ver los posibles usos que este tiene en la vida cotidiana, ya que esto funcionaría como introducción al tema de algoritmos clasificadores y posiblemente nos facilite a comprender la lógica y el funcionamiento de los nuevos algoritmos de clasificación que hay en la actualidad.

### **3.2. Clasificación usando reglas y arboles de decisión(cap 5 - Lantz 2013)**

## **References**

Lantz, B. (2013), *Machine learning with R*, Packt Publishing.