

## 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 clasificador bayesiano ingenuo 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 existen en la actualidad.

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

La clasificación usando reglas y arboles de decisión, expuesto por Lantz (2013) en el quinto capítulo de su libro, al igual que el tema anterior, nos pareció interesante ya que básicamente, es un sistema de aprendizaje supervisado que aplica la estrategia “divide y vencerás” para hacer la clasificación, implementando métodos y técnicas para la realización de procesos inteligentes, representando así el conocimiento y el aprendizaje, con el propósito de automatizar tareas. Su utilización y metodología es algo semejante al tema anterior, ya que dado un conjunto de datos se fabrican diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema.

*“Por su estructura son fáciles de comprender y analizar; su utilización cotidiana se puede dar en diagnósticos médicos, predicciones meteorológicas, controles de calidad, y otros problemas que necesiten de análisis de datos y toma de decisiones”, lo que quiere decir que los árboles de decisión pueden ser usados en cualquier ámbito sin importar que sea laboral o personal, siempre y cuando implique toma de decisiones con cierto grado de incertidumbre* (Calancha Zuniga, Carrión Bárcena, Cori Vargas, & Villa Torres, 2010, pág. 2).

Es evidente que obtener un mecanismo como este sería muy útil debido a que podríamos predecir comportamientos futuros a partir de los comportamientos observados en el pasado. Por ejemplo, a partir de los síntomas que tenemos observados en enfermos anteriores, y sabiendo ya si han desarrollado o no cierta enfermedad, podríamos extraer patrones que nos permitieran predecir si un paciente

nuevo, aquejado de ciertos síntomas, desarrollará o no la misma enfermedad, lo que nos permitiría adelantarnos a su tratamiento. Otro ejemplo que es muy usado en la actualidad se encuentra en el mundo de los créditos bancarios, a partir de los comportamientos de los clientes antiguos con respecto a la morosidad o no de sus pagos del crédito concedido, podemos inferir qué nuevos clientes pueden ser los más convenientes para la concesión de un crédito, es decir, cuáles de ellos tienen más probabilidad de hacer frente al pago del mismo y cuáles más probabilidad de dejarlo sin pagar.

Para nosotros sería importante presentar este tema en seminario, ya que es una técnica muy utilizada por sus buenos resultados, y con ella, se abre paso a otra técnica mucho más potente que son los bosques aleatorios, los cuales utilizan una serie de árboles de decisión con el fin de mejorar la tasa de clasificación; por lo tanto, exponer este tema y ver sus posibles aplicaciones puede ser muy útil para todos, ya que en la vida laboral nos podemos encontrar tranquilamente con problemas de este tipo, que deben ser solucionados utilizando esta metodología.

## References

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