

# Clasificación inteligente sinergia entre lógica difusa y algoritmos genéticos

José Adrián Rodríguez González

Noviembre 2024

## 1. Introducción

Para este mes, se corrigió el primer modelo que se tenía, sin embargo, para poder corregir los errores, se desarrolló otro modelo con una mira distinta.

## 2. Problema del Iris.

Este conjunto de datos es uno de los más clásicos, consiste en 3 especies de flores, cada una posee 4 características.

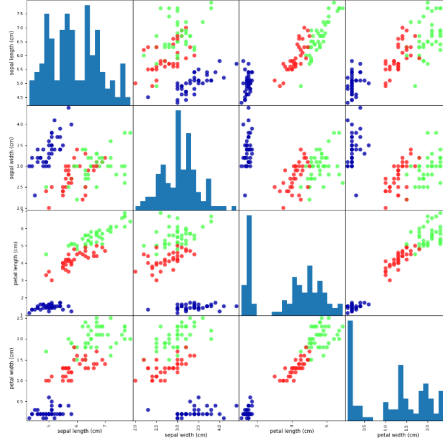


Figura 1: Cruza y mutación

Este problema se basa del paper de (Ishibuchi et al., 1995), en el que usa este conjunto de datos para probar su selección. En este caso, se siguió el proceso descrito del mes pasado, solamente que se cambió la forma de clasificar, anteriormente lo que se realizaba era separar las clases dentro del conjunto de prueba de salida para calcular los  $\alpha$ , sin embargo, las clases como se mencionó

en octubre dependen de la clase asociada a la regla máxima, así que se cambió el proceso. Con este enfoque Michigan se puede clasificar la clase sin tener que ponderarse a una punto. Los procesos del algoritmo genético cambiaron, ahora se realiza una selección por torneo, la cruce es a través de una máscara binaria, en la que se intercambia aleatoriamente el índice de los dos padres. La mutación, utiliza un 0.1 de tasa y compara los hijos con los padres para observar quienes son los individuos que se quedan para la siguiente generación. Además, se añadió el subconjunto *Don't care*, que es una forma de reducir la cantidad de patrones en nuestro sistema basado en reglas. Sin embargo, el problema que se observó en este modelo, es que convergía a las pocas generaciones. Para evitar que convergiera tan rápido el modelo se reintroduce una población aleatoria cada cierta generación. Se probó con 1000 individuo por 20 generaciones, y a las pocas generaciones tendían a una regla. Sin importar las formas en como se ejecutaba el experimento, siempre convergía al conjunto  $[0,0,2,0]$  con un CF de 1, para la clase 1, y las reglas  $[0,0,5,0]$ , con un CF de 0.8678 y clase de 3, y  $[0,0,4,4]$ , para la clase 2.

Regla	Fitness	CF (Factor de Confianza)	Clase
$[0, 0, 5, 0]$	8	0.8678	3
$[0, 0, 2, 0]$	8	1.0000	1
$[0, 0, 4, 4]$	3	0.7149	2

Cuadro 1: Resumen de las mejores reglas.

Los individuos se repiten muchas veces, por lo que se necesitará realizar otro tipo de estrategias. Una de las métricas que se utilizó para medir la eficiencia del modelo fue el accuracy y, particularmente como lo menciona (Ishibuchi et al., 1995), el modelo puede sobre ajustarse, dando un accuracy o muy cercano al 100 %. En nuestro modelo propuesto, dió un accuracy de 99 %, lo que puede significar que nuestro modelo se sobreajusta a nuestro conjunto de datos, por lo que sería punto de estudio, el usar validaciones cruzadas o *Leaving one out*, como técnica para corroborar y poder corregir el sobreajuste de nuestro modelo

### 3. Conclusión

Después de haber mejorado el modelo con las otras estrategias, se implementará en el modelo para la base de datos de cáncer de mama.

### Referencias

Ishibuchi, H., Nozaki, K., Yamamoto, N., and Tanaka, H. (1995). Selecting fuzzy if-then rules for classification problems using genetic algorithms. *IEEE Transactions on Fuzzy Systems*, 3(3):260–270.