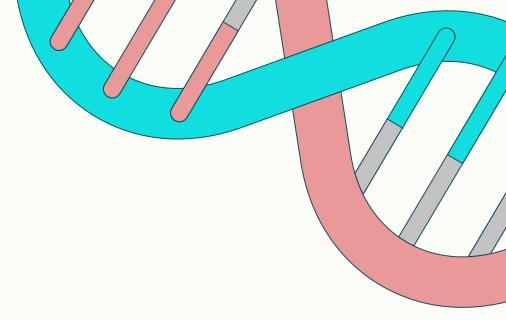
Predicción de cáncer de pulmón

Computación Evolutiva

Lucía de Ancos, Rodrigo Rivas, Camino Rodríguez, Paula Samper, Lucía Yan



Contenidos



OO Descripción del problema

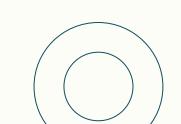
Predicción CancerPulmon

O1 Solución 1
Cromosoma base 3

O2 Solución 2
Características + Pesos

O3 Solución 3 Árboles de decisión

O4 Conclusiones
Resultados



Descripción del problema

Objetivo:

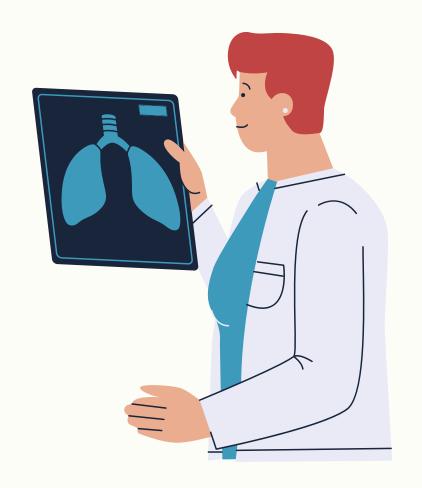
Predicción del CancerPulmon

Dataset:

- 16 columnas
- 284 instancias

Nuestra propuesta:

3 Métodos Evolutivos



Preprocesamiento

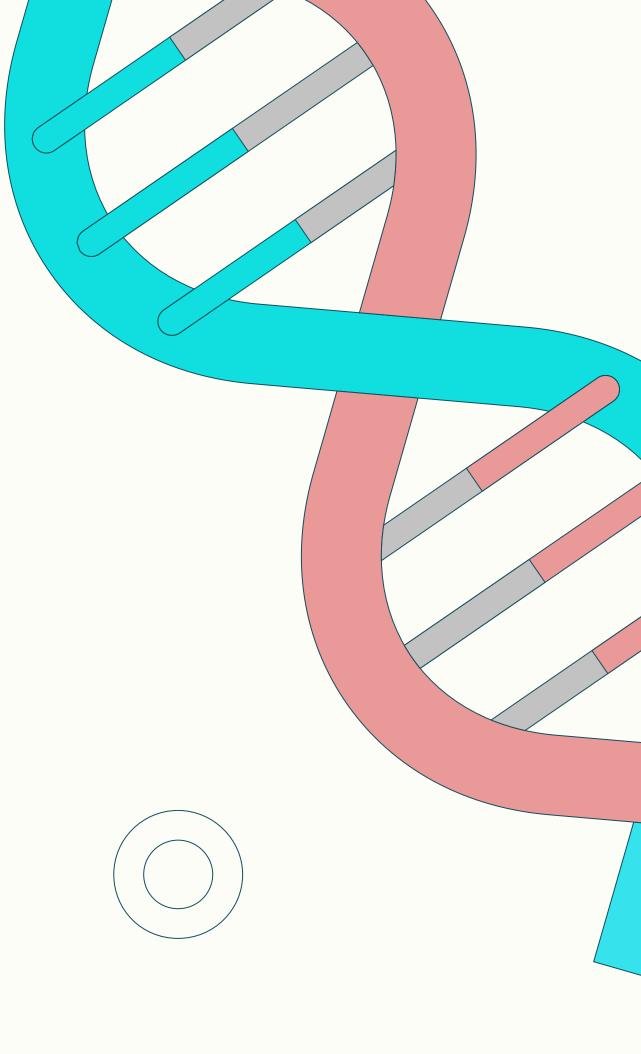
Convertir a binario:

Sexo: M:0, F:1

Edad: < ó > mediana

Dividir en Train (80%) y Test (20%)

• SMOTE: Balanceo de datos

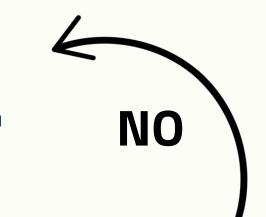


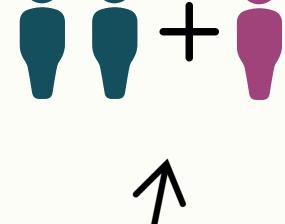


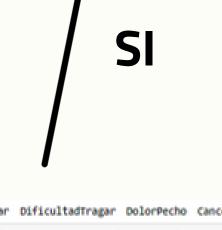


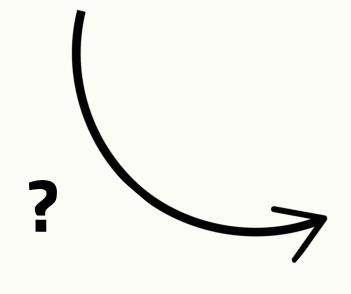
POBLACIÓN INICIAL

Aleatoria









	Sexo	Edad	Fumador	DedosAmarillos	Ansiedad	Hipertension	EnfermedadCronica	Fatiga	Alergia	Silbidos	ConsumidorAlcohol	Tos	DificultadRespirar	DificultadTragar	DolorPecho	CancerPulmon
0	0	1	0	1	1	0	0	1	0	1	1	1	1	1	1	1
1	0	1	1	0	0	0	1	1	1	0	0	0	1	1	1	1
2	1	0	0	0	0	1	0	1	0	1	0	1	1	0	1	0
3	0	1	1	1	1	0	0	0	0	0	1	0	0	1	1	0
4	1	1	0	1	0	0	0	0	0	1	0	1	1	0	0	0
_		-				-					_					
279	1	0	0	1	1	1	0	0	1	1	0	1	0	1	0	1
280	1	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0
281	0	0	1	0	0	0	0	1	1	0	0	0	1	0	1	0
282	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0
283	0	0	0	1	1	0	0	1	0	1	1	1	1	1	1	1
204 10	ws x 16	Leohum	ms.													

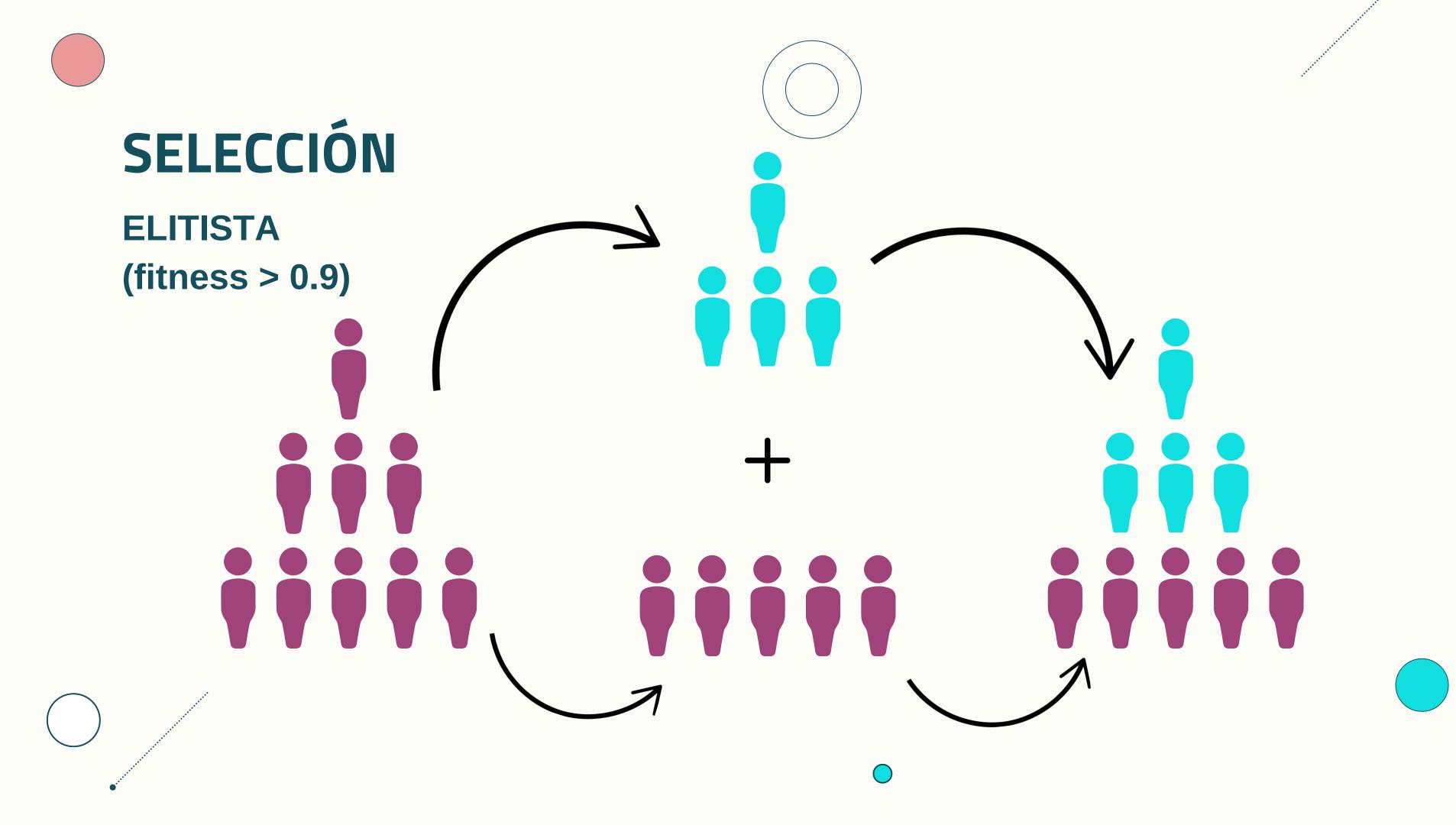
FITNESS

$$fitness = \alpha \times cobertura + \beta \times pureza$$

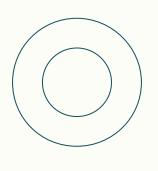
•
$$\alpha = 1 - \frac{i}{n^o iteraciones}; \quad \beta = 1 - \alpha$$

$$ullet cobertura = rac{filas_cubiertas}{filas_totales}$$

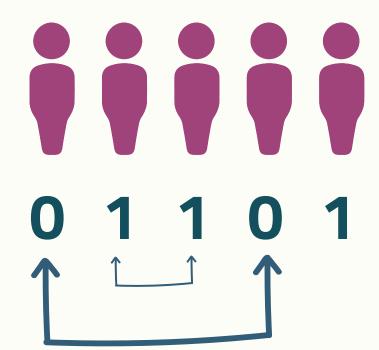
$$egin{aligned} ullet & pureza = 1 - gini \ & = 1 - entropía \end{aligned}$$



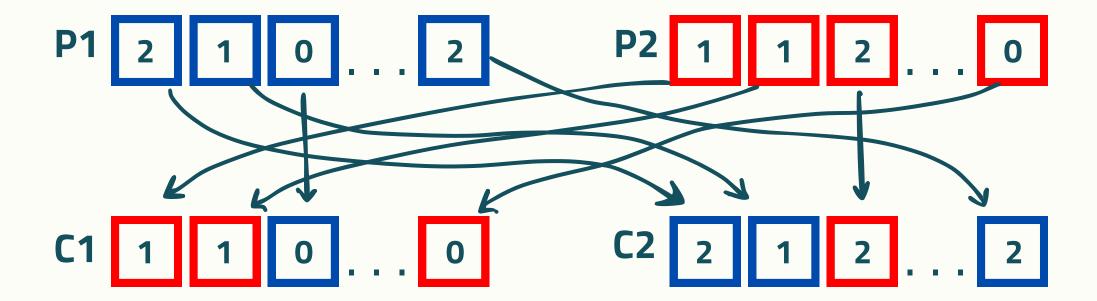








Recombinación uniforme



MUTACIÓN

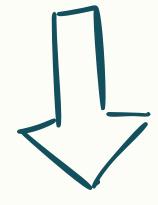
random_prob > mutation_rate



$$mutation_rate = 1 - rac{i}{n^oiteraciones}$$

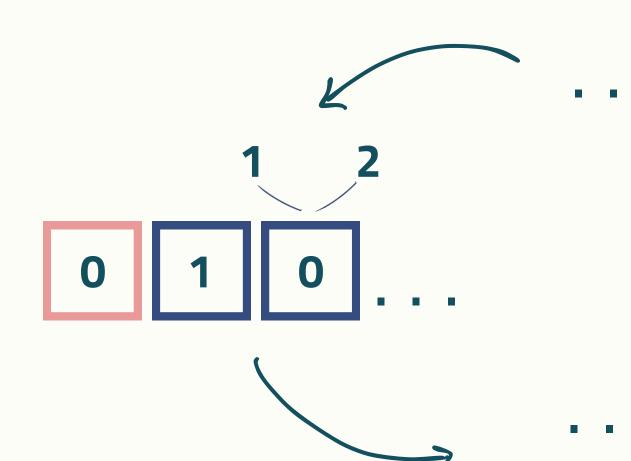
Antes





Después

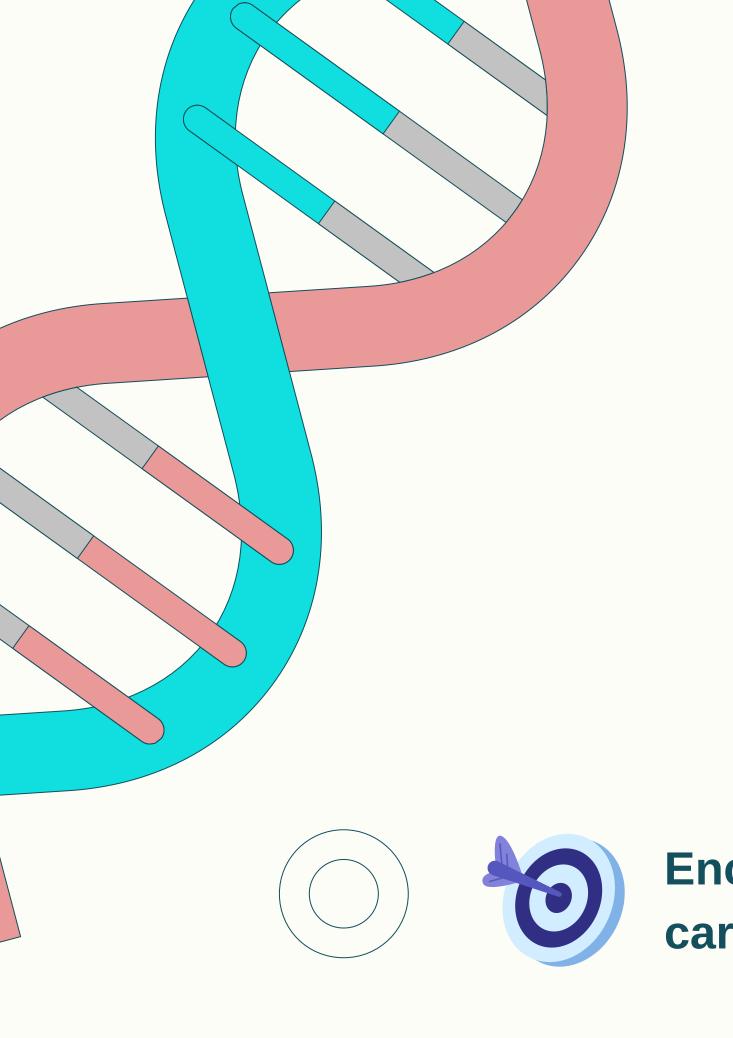






n° iteraciones = 500

init_size	accuracy
300	92.98%
500	91.22%
1000	91.22%

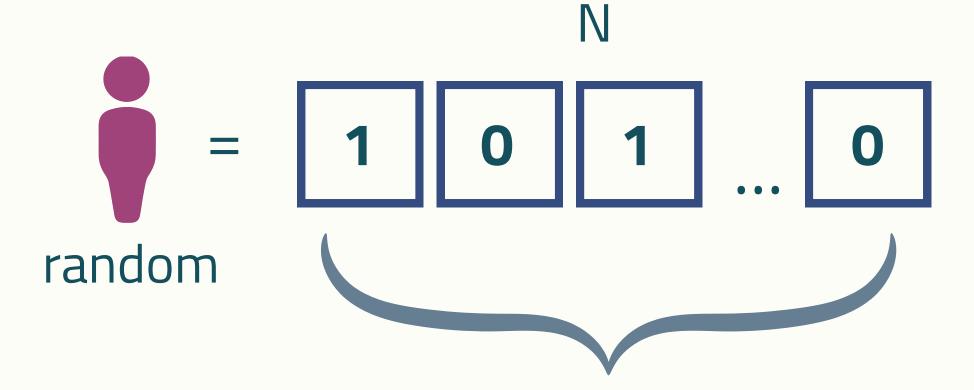


O2 Solución 2

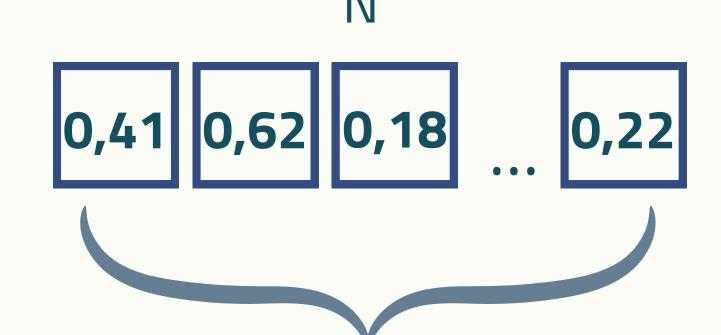
Pesos

Encontrar un subconjunto óptimo de carácterísticas y sus correspondientes pesos

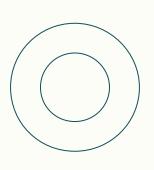
POBLACIÓN INICIAL



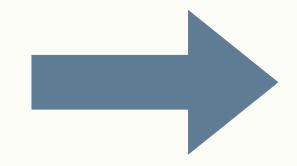




Conjunto de pesos Asignan importancia



NORMALIZACIÓN

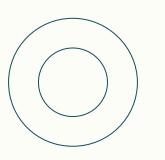


Correcta representación de la influencia de cada característica



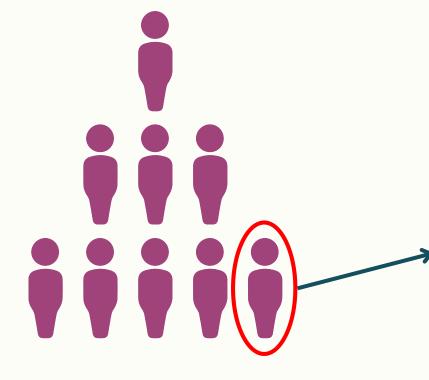
Comparaciones justas

FUNCIÓN DE FITNESS



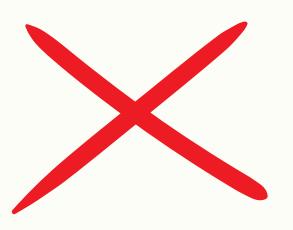






Acuracy, con todo el dataframe



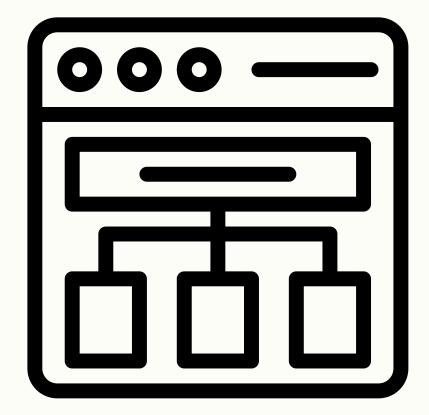


METODOS DE SELECCION



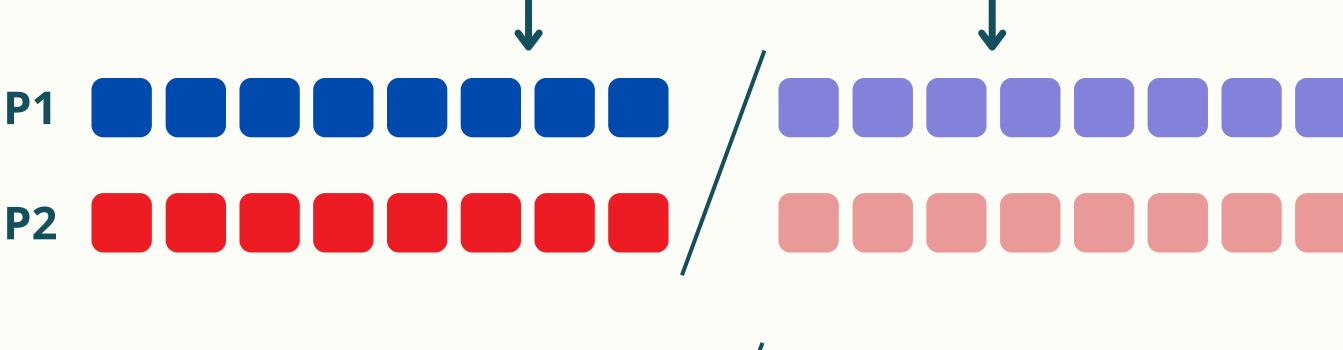


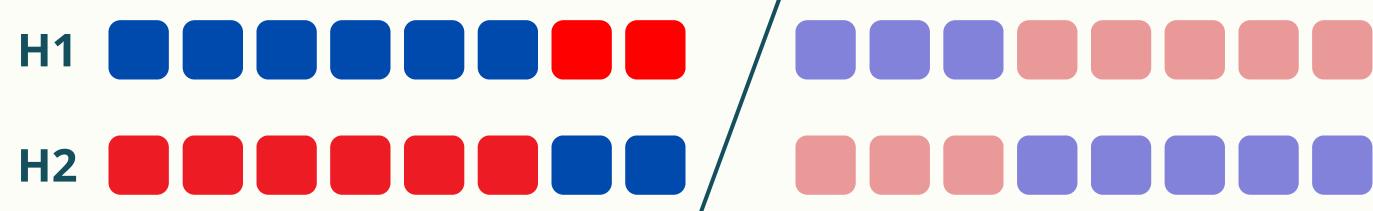




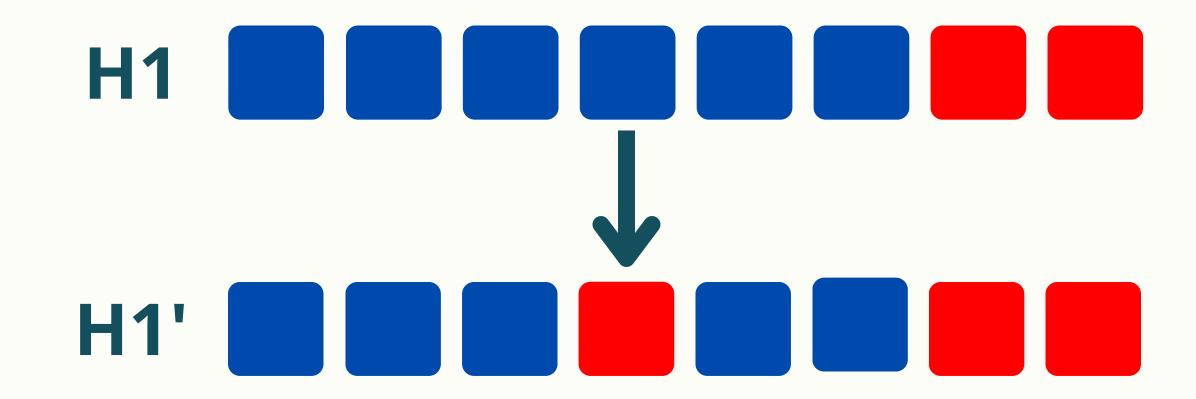
Jerárquico

CRUCES





MUTACIONES



Exploración y Explotación

¿Tasa de mutación? ¿Frecuencia de mutación?





n° iteraciones = 300

init_size	accuracy
300	84,21%
500	86,96%
1000	84,21%

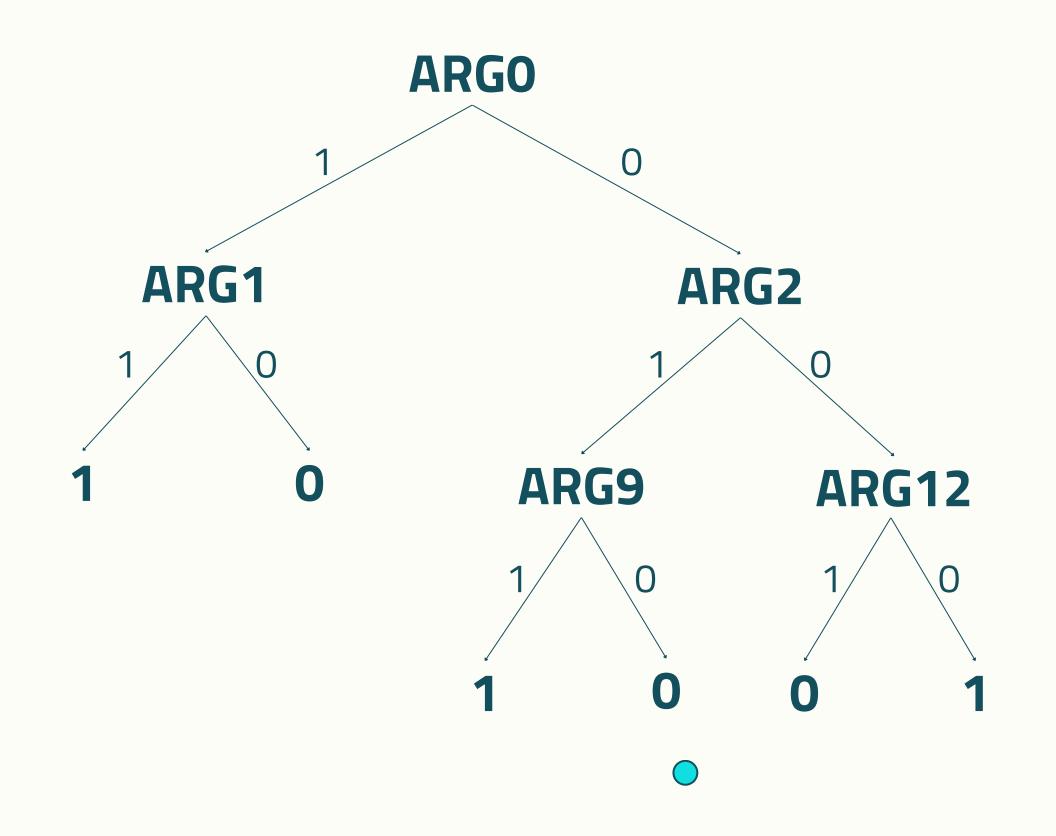
O3 Solución 3

Árboles de Decisión

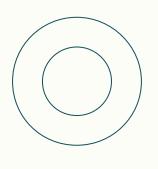


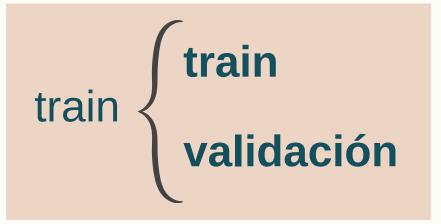


ÁRBOLES DE DECISIÓN



ALGORITMO



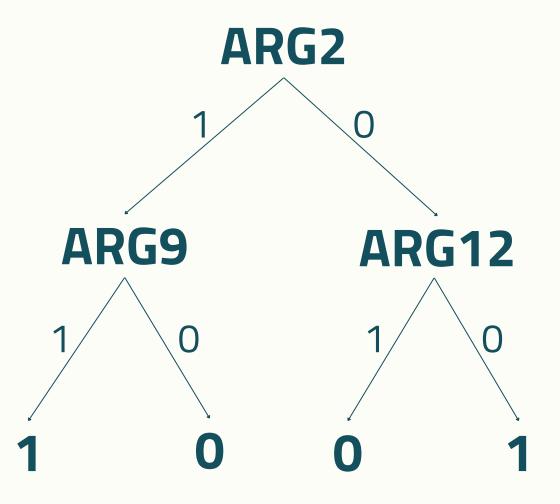


- Población inicial
 - La profundidad máxima de los árboles es el número de atributos.
- Evaluación (función de fitness)

 Accuracy del árbol sobre los datos de train.
- Selección y crossover, o mutación
 O bien los individuos se reproducen, o bien se
 muta un nodo de un individuo de la población.
- Evaluación (función de fitness)

 Evaluación de los nuevos individuos.
- Selección del mejor individuo

 Accuracy del los árboles sobre los datos de validación.



SELECCIÓN

Selección por rango

Probabilidad de ser elegido para la reproducción

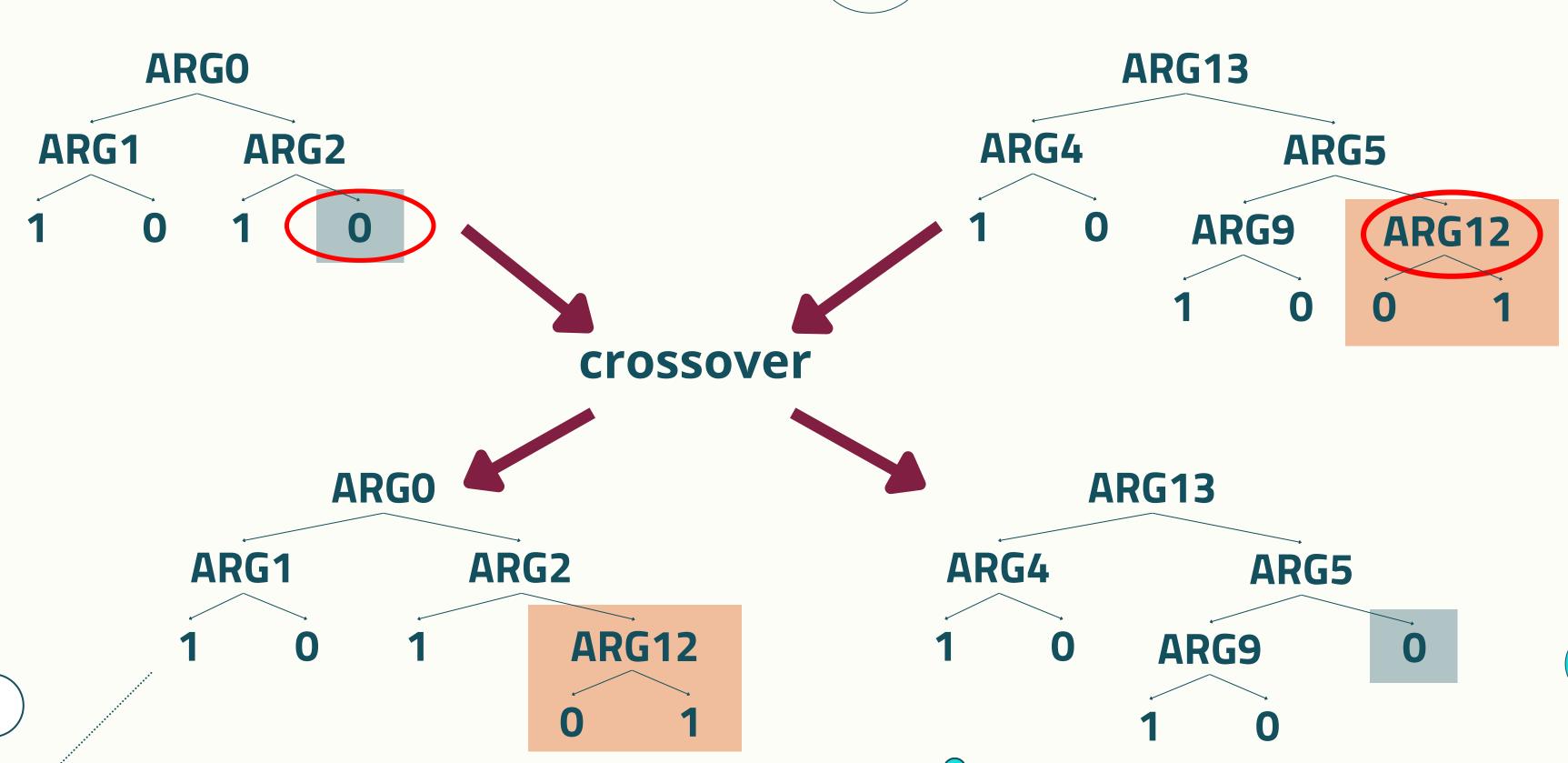


por su fitness.

Elitismo

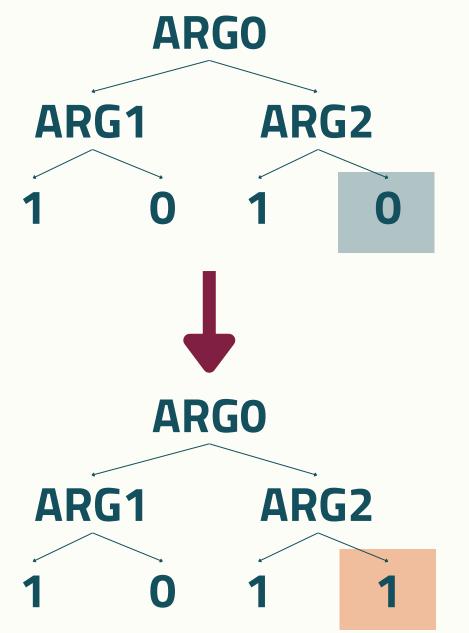
Se conserva el mejor 20% de la población

CROSSOVER

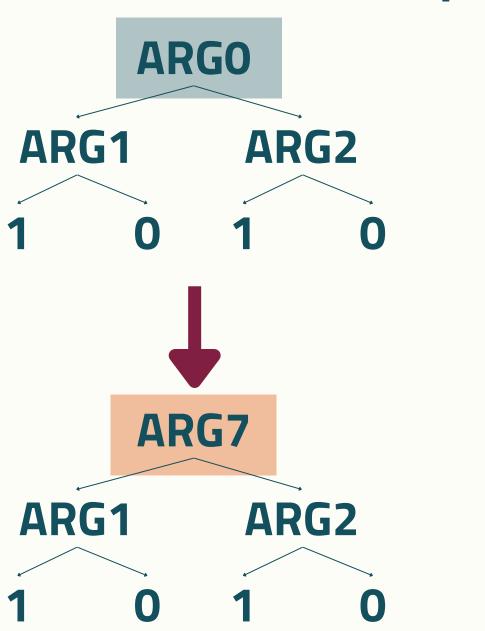


MUTACIÓN

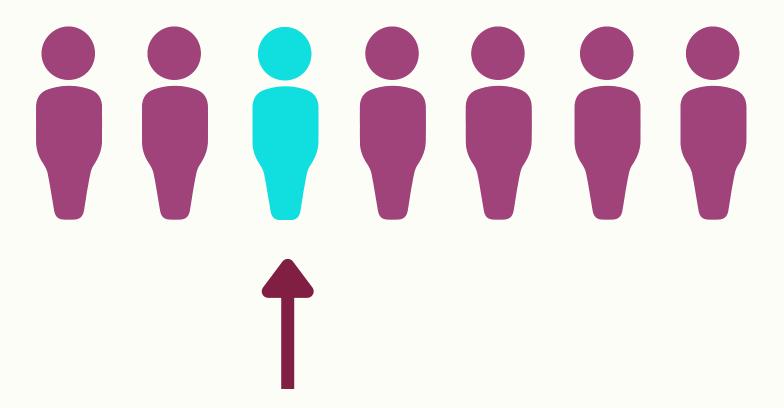
Mutación terminal simple



Mutación funcional simple

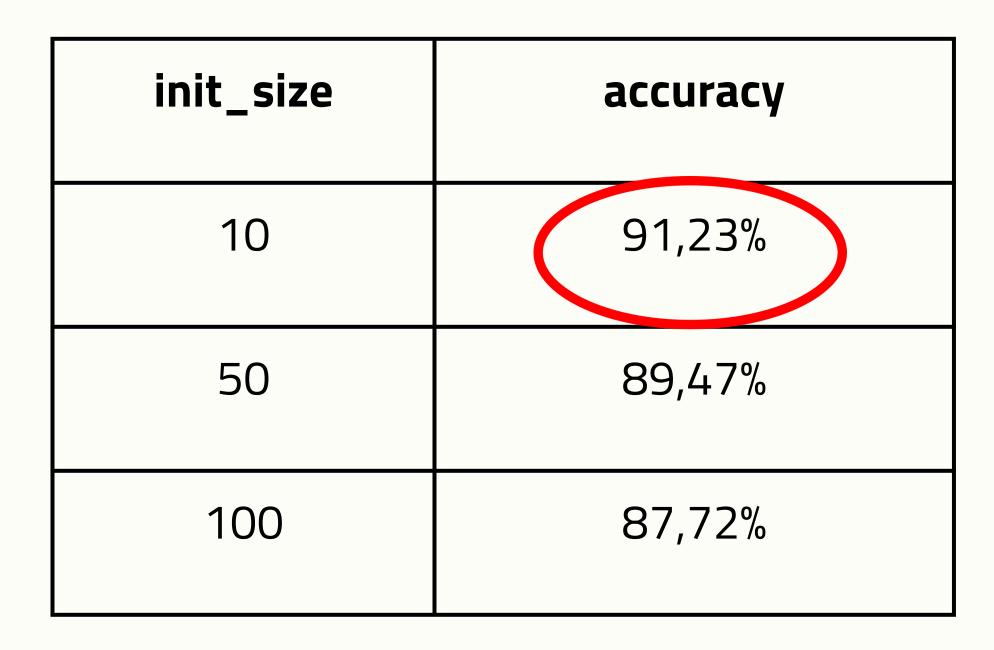


SELECCIÓN DEL MEJOR INDIVIDUO



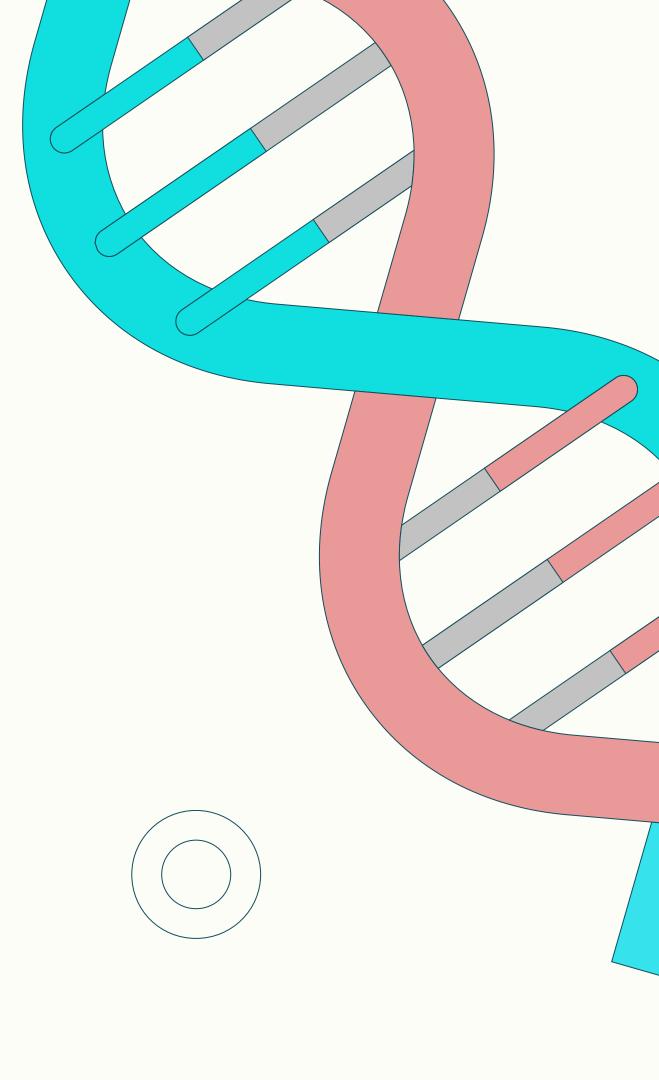
Selección del mejor individuo de la población calculando el accuracy de cada individuo en el **conjunto de validación**.

RESULTADOS



O4 Conclusiones

Resultados





Conclusiones

Solución	accuracy
1	92.98%
2	86.96%
3	91,23%



- Las tres soluciones consiguen el objetivo, clasifican con precisión alta
- Trade-off precisión y velocidad convergencia
- Ayuda a la enfermedad de cáncer pulmonar.

