

Redes Neuronales y Algoritmos Evolutivos

Algoritmos Genéticos

Ejercicio 1

Se desea utilizar un Algoritmo Genético (AG) para obtener las reglas de clasificación a partir de los ejemplos de la siguiente tabla (*curso.xls*)

Practica	Activ-Distancia	Activ-Presencial	Calificacion
bien	alta	alta	aprobado
bien	baja	alta	desaprobado
muy_bien	alta	alta	aprobado
regular	alta	alta	aprobado
regular	alta	baja	desaprobado
regular	baja	baja	desaprobado
muy_bien	baja	baja	aprobado
bien	baja	baja	desaprobado
bien	alta	baja	aprobado
regular	baja	baja	desaprobado
bien	alta	baja	aprobado
muy_bien	alta	alta	aprobado
regular	baja	baja	desaprobado
regular	alta	alta	aprobado

Se ha decidido no someter el consecuente de la regla a evolución sino asociar todos los individuos de la población con una misma clase y ejecutarlo tantas veces como clases haya.

Por simplicidad se buscará obtener únicamente las reglas correspondientes a la clase “Calificacion=Aprobado”.

- Utilizando los atributos Practica, Activ-Distancia y Activ-Presencial construya una población inicial formada por 10 reglas aleatorias utilizando una codificación entera de longitud fija.
- Elija cuatro reglas de la población anterior y represéntelas utilizando una codificación
 - Entera de longitud variable.
 - Binaria de longitud fija
- Usando la población de a), implemente manualmente el proceso de obtención de la próxima generación reemplazando los dos peores individuos (reglas) por los dos hijos de los dos mejores.
 - La aptitud del individuo (regla) se obtiene promediando sus valores de soporte y confianza.
 - El cruce (o crossover) es de un punto y NO hay mutación.
- Indique la regla (fenotipo) correspondiente al mejor individuo de la población obtenida en c), junto con su fitness.
- Repita c) y d) dos veces mas.
- ¿Cuál es la importancia de la mutación en el proceso evolutivo? Ejemplifique su respuesta utilizando los individuos de su población que, luego del punto e), ya se encuentran en la 4ta. generación.

Ejercicio 2

El archivo *Cancer.xls* contiene 683 muestras de cáncer de mama. Todos los atributos son numéricos y toman valores enteros entre 1 y 10.

- **Grosor** : Grosor del grupo de células
- **U_size** : Uniformidad del tamaño de la célula
- **U_Forma** : Uniformidad referida a la forma de la célula
- **Adhesion** : Adhesión marginal
- **Size** : Tamaño de una célula epitelial simple
- **Nucleos** : Nucleos aislados
- **Cromatina** : Cromatina Blanda
- **Nucleolitos** : Nucleolitos normales

Las muestras se encuentran agrupadas en dos clases posibles

Clase 2 : BENIGNO

Clase 4 : MALIGNO

A partir de esta información, programe en MatLab un algoritmo genético para obtener al menos dos reglas de clasificación para cada clase discretizando cada atributo primero en 4 intervalos y luego en 7 intervalos.

Deberá utilizar cromosomas de longitud fija formados únicamente por la información correspondiente al antecedente de la regla. Esto implica que el proceso se repetirá para cada clase. Con respecto a los operadores genéticos se utilizará cruce de un punto y mutación uniforme.

La representación será entera de longitud fija, es decir que se utilizará un dígito decimal para cada atributo. Su solución deberá controlar las soluciones inválidas que puedan generarse por la aplicación de los operadores genéticos.

Como función de fitness se utilizará el promedio entre el soporte y la confianza de la regla.

Se probarán dos estrategias de selección y reemplazo distintas

- a) En cada generación sólo se crearán dos hijos a partir de dos padres seleccionados por el método de la ruleta. Los nuevos individuos reemplazarán a dos existentes seleccionados también por el método de la ruleta entre el 75% peor de la población (es decir que el mejor 25% pasa intacto de una generación a la siguiente).
- b) En cada generación sólo se crearán dos hijos a partir de dos padres seleccionados por el método del sobrante estocástico con reemplazo. Cada uno de los nuevos individuos competirá en un torneo binario probabilístico, con probabilidad $p=0.7$, con un individuo seleccionado aleatoriamente del peor 75% de la población.

Ejemplo

Las siguientes reglas fueron obtenidas utilizando la representación indicada luego de 100 generaciones para cada clase. El tamaño de población empleado en cada caso fue de 70 individuos y se discretizó en 4 intervalos. Las probabilidades de cruce y mutación fueron de 0.65 y 0.05 respectivamente. La selección y el reemplazo corresponden a la opción a)

SI (1 < U_size <= 2) ENTONCES BENIGNO (SOP 0.5944 - CONF 0.9713 – FITNESS 0.7829)

SI (1 < Nucleos <= 2) AND (1 < Nucleolos <= 2) ENTONCES BENIGNO

(SOP 0.5754 - CONF 0.9899 – FITNESS 0.7827)

SI (8 < Nucleos <= 10) ENTONCES MALIGNO (SOP 0,2299 – CONF 0,9691 – FITNESS 0,5995)

SI (8 < Grosor <= 10) ENTONCES MALIGNO (SOP 0,1801 – CONF 0,9685 – FITNESS 0,5743)