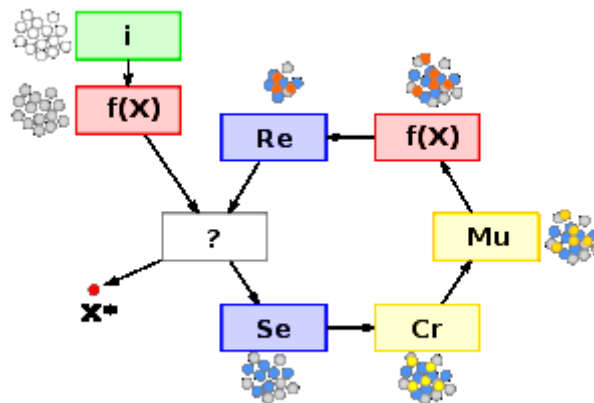




LIC. EN SISTEMAS DE INFORMACIÓN INTELIGENCIA ARTIFICIAL.

PARTE 1 - ALGORITMOS GENÉTICOS.



BREVE COMENTARIO.

Los algoritmos genéticos (AG) se inspiran en la *evolución biológica* y su base *genético-molecular*, fueron propuestos por John Henry Holland en 1970, (Michigan, US). El Dr. John Henry Holland fue un pionero en sistemas complejos y ciencia no lineal. También es conocido como el padre del Algoritmo genético. Profesor de Filosofía, de Ingeniería Eléctrica y de Ciencias de la computación en la Universidad de Michigan. Autor de numerosos libros sobre Sistemas complejos adaptativos.

Dado un problema específico a resolver, la entrada del AG es un conjunto de soluciones potenciales a ese problema, codificadas de alguna manera, y una métrica llamada función de aptitud, o fitness, que permite evaluar cuantitativamente a cada solución candidata. Estas candidatas pueden ser soluciones que ya se sabe que funcionan, con el objetivo de que el AG las mejore, pero se suelen generar aleatoriamente.

Los algoritmos genéticos han demostrado ser una estrategia enormemente poderosa y exitosa para resolver problemas, demostrando de manera espectacular el poder de los principios evolutivos. Se han utilizado algoritmos genéticos en una amplia variedad de campos para desarrollar soluciones a problemas tan o más difíciles que los abordados por los diseñadores humanos.

PROCESO USUAL AG

1. **Población:** el proceso comienza con un conjunto de individuos que se llama Población, cada individuo es una solución al problema que desea resolver.
2. **Función de aptitud:** La función de aptitud determina qué tan bueno es un individuo. Le da un puntaje de aptitud a cada individuo.
3. **Selección:** la idea de la fase de selección es seleccionar a las personas más aptas y dejarles pasar sus genes a la próxima generación. De la bibliografía se conocen varios métodos de selección los cuales puede investigar: selección elíptica, selección proporcional o de aptitud, selección por rueda de ruleta, selección escalada, selección por torneo, selección por rango, selección generacional, selección jerárquica.
4. **Cruzamiento:** el cruzamiento es la fase más importante en un algoritmo genético. Para cada par de padres que surgirán, se elige al azar un punto de cruce dentro de los genes.
5. **Mutación:** en ciertos descendientes nuevos, algunos de sus genes pueden ser sometidos a una mutación.

Nota selección por ruleta:

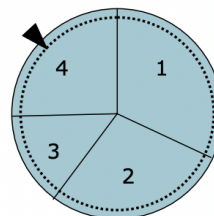
Propuesto por DeJong, es posiblemente el método más utilizado desde los orígenes de los Algoritmos Genéticos [Blickle and Thiele, 1995].

A cada uno de los individuos de la población se le asigna una parte proporcional a su ajuste de una ruleta, de tal forma que la suma de todos los porcentajes sea la unidad. Los mejores individuos recibirán una porción de la ruleta mayor que la recibida por los peores. Generalmente la población está ordenada en base al ajuste por lo que las porciones más grandes se encuentran al inicio de la ruleta. Para seleccionar un individuo basta con generar un número aleatorio del intervalo [0..1] y devolver el individuo situado en esa posición de la ruleta. Esta posición se suele obtener recorriendo los individuos de la población y acumulando sus proporciones de ruleta hasta que la suma exceda el valor obtenido.

Rueda de ruleta

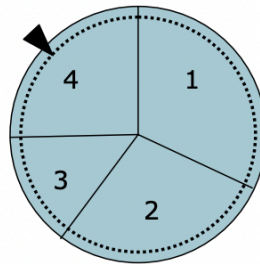
- Se asigna a cada individuo la probabilidad:
- Si algún individuo domina la población, se escala o normaliza
- Se elijen parejas aleatorias de individuos de acuerdo a su probabilidad
- Inconveniente: los individuos con más aptitud tiende a dominar la población en pocas generaciones

$$\Pr(x) = \frac{\text{aptitud}(x)}{\sum_{y \in \text{población}} \text{aptitud}(y)}$$



Rueda

Valor binario	Descodificación	Valor real	Aptitud	Probabilidad	Probabilidad acumulada
10010100	148	0,161	0.974	0.299	0.299
10010001	145	0,137	0.981	0.301	0.600
00101001	41	-0,678	0.540	0.166	0.766
01000101	65	-0,490	0.760	0.233	1.000



ACTIVIDADES - PROGRAMACIÓN GENÉTICA CON PYTHON.

Considere los contenidos vistos en clase de teoría y los códigos de los ejemplos (GA 1.ipynb, GA 2.ipynb) vistos en práctica.

Entrega el código python en formato Notebook (ipynb). Consideraciones:

- Comente cada parte del código
- Explique claramente cual es la función de aptitud en cada caso
- Documente el proceso de cruzamiento y mutación
- Separe el código en bloques de código legibles, en lo posible por etapa según **(PROCESO USUAL AG)**

ACTIVIDAD 1 - BÚSQUEDA DE LA CADENA DE TEXTO.

Encuentre una cadena de texto determinada utilizando un AG

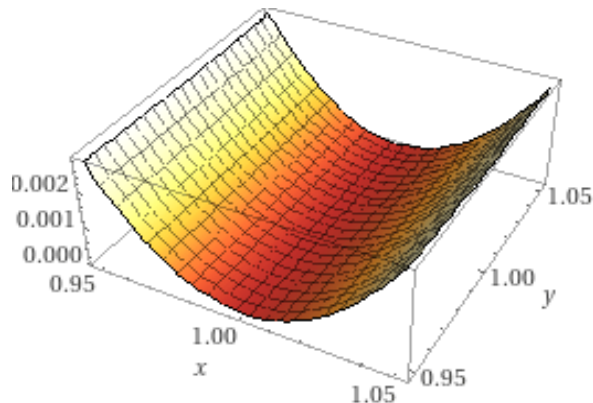
TARGET = "Los algoritmos genéticos sirven para optimizar". Considere la lista, donde cada elemento *i-ésimo* es un número diferente de población.

si **PS** = [100,1000,50,200,300,50,500], ¿Cuál tamaño de población se ajusta mejor al problema?

Realice un programa que automáticamente encuentre la solución en base a la lista definida.

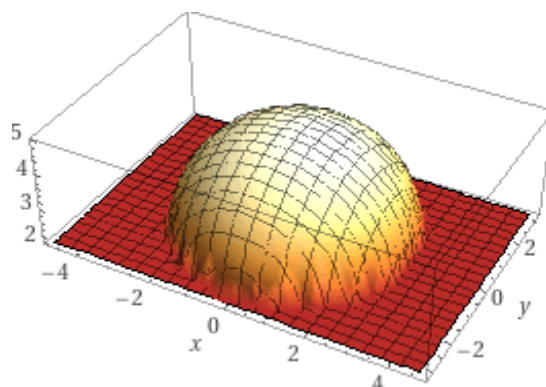
ACTIVIDAD 2 - OPTIMIZAR FUNCIONES EN EL ESPACIO.

Considere la siguiente función definida en el dominio del espacio \mathbb{R}^2 ,
 $f(x,y) = (x - 1)^2 + (x - y)^4$



Utilice un AG para aproximar el mínimo de la misma, utilizando el método de la selección por rueda de ruleta.

Considere la siguiente función definida en el dominio del espacio \mathbb{R}^2 ,
 $f(x,y) = 2 + \sqrt{9 - x^2 - y^2}$. Proponga un método diferente al ejercicio anterior y programe un AG para encontrar aproximar máximo para la parte real de la misma, osea $9 - x^2 - y^2 \geq 0$:



Considere la siguiente función definida en el dominio del espacio \mathbb{R}^3
 $f(x,y,z) = z - x^2 - 15y^2$, busque la mejor solución para el máximo con la siguiente restricción $-2 < x,y,z < 11$.

ACTIVIDAD 3 - SELECCIÓN DE VARIABLES CON AG.

En aprendizaje de máquina y estadística, selección de característica, también conocida como selección de variable, selección de atributo o selección de subconjunto variable, es el proceso de seleccionar un subconjunto de características pertinentes (variables, predictores) para su uso

en construcción de modelos. Las técnicas de selección de la característica son utilizadas por cuatro razones:

- Simplificación de modelos con el fin de hacerlas más sencillas de interpretar para los usuarios/investigadores
- Tiempo de entrenamiento más corto,
- Para evitar la maldición de la dimensión (curse of dimensionality).
- Reducir overfitting

Zachary Warnes escribió un [artículo](#) sobre cómo utilizar algoritmos genéticos para seleccionar variables. Estudiar el artículo y aplicar el algoritmo genético para seleccionar las mejores features del set de datos de [Wine](#), pruebe con diferentes números de generaciones y reporte los resultados.

Identifique en el algoritmo las siguientes etapas:

- Seleccionar
- Cruzar
- Mutar
- Reemplazar