

UNIVERSIDAD MAYOR DE SAN ANDRÉS
FACULTAD DE CIENCIAS PURAS Y NATURALES
CARRERA DE INFORMÁTICA



Examen Parcial - Pregunta 1

Nombres: EDUARDO MEDRANO AYARDE CI: 6989411

Materia: Inteligencia Artificial (INF-354)

Docente: M.Sc. Moises Martin Silva Choque

Fecha: 27 de julio de 2020

1. Use 3 columnas del dataset, generándose un algoritmo genético con el uso de DEAP

1.1. Descripción del Dataset

Para el caso de estudio con el DEAP, se hizo uso del dataset Haberman, del siguiente enlace <https://archive.ics.uci.edu/ml/datasets/Haberman%27s+Survival>, el cual describe, sobre la evolución y supervivencia de mujeres que hayan sido operadas, para tratar el cáncer de mama, las columnas del dataset son los siguientes:

- Edad: Edad de las mujeres, cuando fueron sometidas a la operación.
- Año de intervención quirúrgica: restando 2000-1900, la columna muestra el año en el que se realizaron la operación.
- Ganglios: Es un número entero, que nos señala, cuantos ganglios cancerosos fueron detectados.
- Supervivencia: 1 si sobrevivió más de 5 años después de la operación, 2 si murió entre los 5 años.

Para el uso del DEAP, se hará uso de las 3 primeras columnas.

El proceso de evolución genética, nos dirá cuantas generaciones se necesitaran, para que mujeres de 21 años ya presenten cáncer de mama. Para este propósito el algoritmo se desarrolla bajo las siguientes características:

- Se evaluarán grupos de 10 mujeres en 10 mujeres.
- La población será de 1000 mujeres.
- Será una función de minimización, esto por el hecho de que se desea ver a mujeres jóvenes y cuantas generaciones se tomara.

- En la función de evaluación se tiene los siguientes parámetros:
 - Se hará una sumatoria de todos los ganglios cancerosos encontrados en el grupo de 10 mujeres.
 - La sumatoria se realiza solo si, el grupo de estudio se encuentra en el rango de edad del dataset, el año de la cirugía fue entre 1960-1970 y por último la edad de la mujer fue menor a 35 años.
 - Al cumplir esto se pregunta si la sumatoria de ganglios cancerosos, es mayor a 10, de ser así se devuelve el valor máximo del grupo de estudio, caso contrario se devuelve a la más joven.
- El algoritmo de cruce es cxUniform.
- La población es aleatoria en un rango de edad de 21 a 80 años.

Todo el análisis anterior, está desarrollado bajo el siguiente código:

```
1  import array
2  import random
3  import numpy
4  import pandas as pd
5  from deap import algorithms
6  from deap import base
7  from deap import creator
8  from deap import tools
9
10 #leemos los datos provenientes del dataset
11
12 data=pd.read_csv("haberman.csv",sep=",")
13 creator.create("FitnessMax", base.Fitness, weights=(-1.0,))
14 creator.create("Individual", list, fitness=creator.FitnessMax)
15 toolbox = base.Toolbox()
16
```

```

17 #colocamos en el grupo de mujeres que se someten a una operación por cancer de ma
18 toolbox.register("attr_bool", random.randint, 21, 80)
19 toolbox.register("individual", tools.initRepeat, creator.Individual, toolbox.attr
20 toolbox.register("population", tools.initRepeat, list, toolbox.individual)
21
22 def evacancer(individual,edad,t_opera,ganglios_auxiliares):
23
24 #vemos cuantas mujeres sometidas a operación generaron ganglios cancerosos en el
25     aux=0
26     gnp=numpy.array(ganglios_auxiliares)
27     enp=numpy.array(edad)
28     indi=numpy.array(individual)
29     for a in range(0,len(individual)-1):
30         if individual[a] in edad and t_opera[a]>60 and edad[a]<35:
31             aux+=gnp[a]
32 #ya definido los ganglios generados entre todas las mujeres podemos decir que si
33 # sumados todos los ganglios cancerosos el numero es mayo o igual que 10
34 # el conjunto de mujeres que no estaba en el rango de edad para generar ganglios
35 # caso contrario no generara ningun ganglio
36 # y en un caso devolvemos la mujer de mas edad o la mas joven
37 if aux>=10:
38     return indi.max(),
39 else:
40     return indi.min(),
41 toolbox.register("evaluate", evacancer,edad=data["edad"],t_opera=data["year"],gan
42 toolbox.register("mate", tools.cxUniform, indpb=0.5)
43 toolbox.register("mutate", tools.mutFlipBit, indpb=0.05)
44 toolbox.register("select", tools.selTournament, tournsize=3)
45
46 def main():
47     pop = toolbox.population(n=1000)
48     hof = tools.HallOfFame(1)

```

```

49     stats = tools.Statistics(lambda ind: ind.fitness.values)
50     stats.register("avg", numpy.mean)
51     stats.register("std", numpy.std)
52     stats.register("min", numpy.min)
53     stats.register("max", numpy.max)
54     pop, log = algorithms.eaSimple(pop, toolbox, cxpb=0.5, mutpb=0.2, ngen=200,
55     stats=stats, halloffame=hof, verbose=True)
56     return pop, log, hof
57 if __name__ == "__main__":
58     main()

```

De donde se obtiene los siguientes resultados:

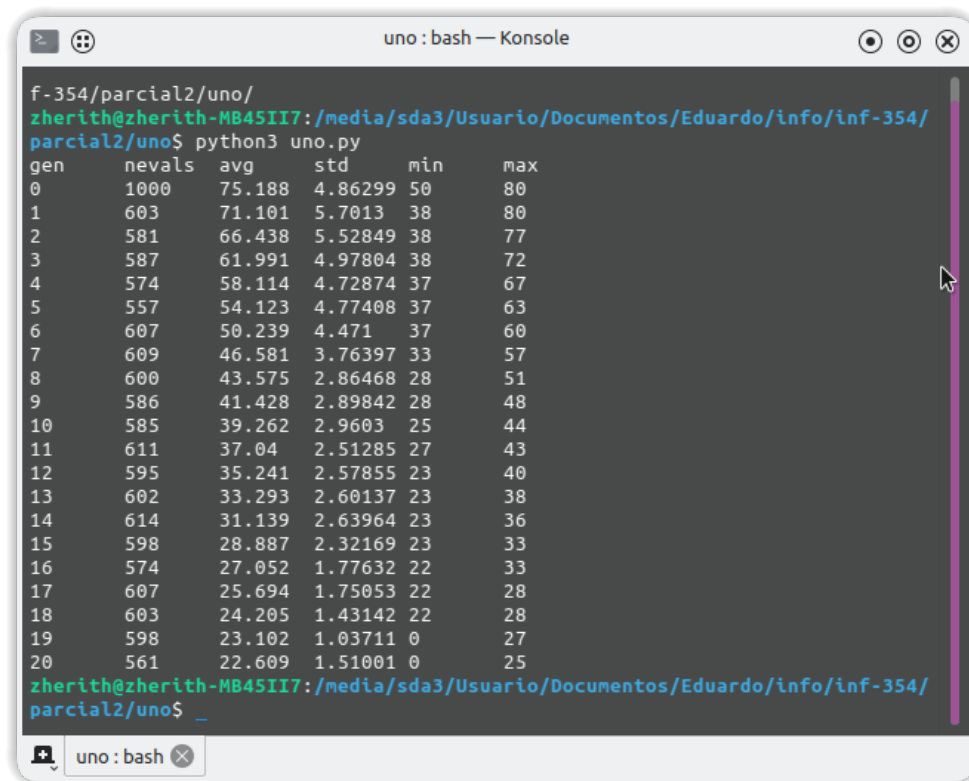


Figura 1: Resultados Obtenidos del Dataset Haberman

De donde se puede apreciar que, para la generación 18, mujeres entre el rango de edad de 22 a 28, pueden presentar ganglios cancerosos.