Sinhgad Technical Education Society's

SINHGAD INSTITUTE OF TECHNOLOGY

(Affiliated to Savitribai Phule Pune University, Pune and Approved by, AICTE, New Delhi.)

Gat No. 309/310 , Kusgaon (Bk), off Mumbai –Pune, Expressway.

Lonavala, Pune, 410401, Website : www.sinhgad.edu

Department of Computer Engineering

BE Mini Project Presentation
SEM -1
2021-22

Group Members:
PUSHKARAJ CHAVAN
(BE C 58)

Subject:
AIML

Project Title:
Genetic Algorithm for optimization on a breast cancer dataset

INTRODUCTION

In this project we are using breast cancer dataset from UCI ML repository and apply the Genetic Algorithm on dataset to optimize result than traditional method.

DATASET

Name of Dataset: Breast Cancer Wisconsin Dataset

Source: UCI ML Repository

No. of Attributes: 31 (30 features and 1 result i.e. malignant or benign)

No. records: 569

Null values: 0

GENETIC ALGORITHM

Genetic Algorithm inspired by the Human genetic process of passing genes from one generation to another.

1. Firstly, we defined our initial population as our countrymen.
2. We defined a function to classify whether is a person is good or bad.
3. Then we selected good people for mating to produce their off-springs.
4. And finally, these off-springs replace the bad people from the population and this process repeats.

This is how genetic algorithm actually works, which basically tries to mimic the human evolution to some extent.

GENETIC ALGORITHM
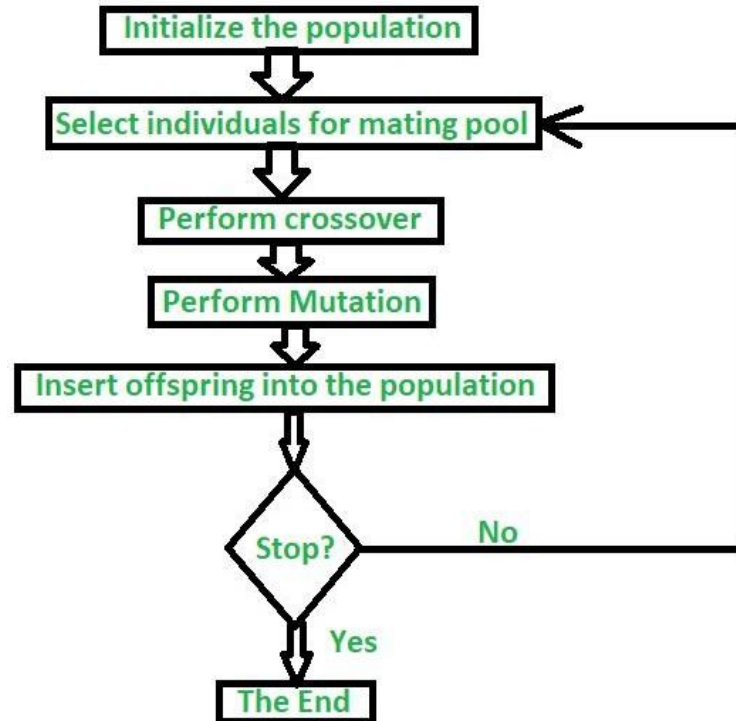
GENETIC ALGORITHM

- **Initial Population**– Population will contain individuals, each having their own set of chromosomes. Chromosomes are binary strings so in this project we are creating function to generate population. Function takes size of population, in this case 200 and number features in dataset which generate boolean type chromosomes for population and store it. As Chromosomes has 1 and 0 boolean values, 1 is acceptable and 0 is non-acceptable feature.
- **Fitness function**– Finds the fitness value of the each of the chromosomes. Here each chromosome value is used as training set to generate ML model and tested it. For accepting train data set, value 1 in chromosome for particular feature is considered and with 0 neglected.
- **Selection**– Select the best fitted chromosomes as parents to pass the genes for the next generation and create a new population. Here number of parents is 100 so 100 parents are created to generate new off-springs

GENETIC ALGORITHM

- **Cross-over**– Create new set of chromosome by combining the parents and add them to new population set
- **Mutation**– Perfrom mutation which alters one or more gene values in a chromosome in the new population set generated. Mutation helps in getting more diverse opportunity. Obtained population will be used in the next generation

Repeat step 2-5 again for each generation

MACHINE LEARNING MODEL

In this project we are using Decision Tree Classifier for Training and Testing of model.

Decision Tree consists of :

1. **Nodes** : Test for the value of a certain attribute.

2. **Edges/ Branch** : Correspond to the outcome of a test and connect to the next node or leaf.

3. **Leaf nodes** : Terminal nodes that predict the outcome

MACHINE LEARNING MODEL

How Decision Tree works?

1. **Step-1:** Begin the tree with the root node, says S, which contains the complete dataset.
2. **Step-2:** Find the best attribute in the dataset using **Attribute Selection Measure (ASM).**
3. **Step-3:** Divide the S into subsets that contains possible values for the best attributes.
4. **Step-4:** Generate the decision tree node, which contains the best attribute.
5. **Step-5:** Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

CODE

https://colab.research.google.com/drive/1fvW-8dqfyo-Qw0ry1EpiHXBs_q_uUGZC?usp=sharing