**Data Dictionary**

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Name** | **Data Type** | **Description/Process** | **Possible Values** |
| Program Duration | Integer | Minutes input by user for the length of time the program has to run | Default: 15 |
| Target Equation | EquationTree | The target for individuals’ fitness when being evaluated  (Not sure this is needed in the program as long as training data is provided as input) | ( (x^2) - 1 / 2) |
| Training Data | Nx2 Array | Collection of the training X and Y pairs for the target equation | (x1, y1) – (xN, yN) |
| Training Data Size | Integer | Number of X and Y pairs in the training data that will be used to iterate over the training data with each Individual equation |  |
| Population Size | Integer | Quantity of individual equations in the population | 100-1000 |
| Population | List | List of Individual Trees for fitness evaluation |  |
| Individual Tree(n) | EquationTree | Collection and sequence of Nodes making up a candidate equation | Example:  (x \* x + 20 -6 / 4/ 9 \* 114) / (2 \* x + 5 – 6 / 2 – x) |
| Tree Depth | Integer | Limit to number of levels deep a candidate tree can be | 1-30? |
| Individual ID | Integer | Unique identifier for each candidate in a population | 100-1000 |
| Node ID | Integer | Unique identifier of a specific tree node | 1-100? |
| Tree Operator Node | TreeNode | Operator or Operand node of the candidate equation tree | + - \* / |
| Tree Operand Node | TreeNode | Operator or Operand node of the candidate equation tree | x, integers -100 – 100? |
| Tree Operand Min | Integer | User input for the lowest an operand value can be | -100 - 0 |
| Tree Operand Max | Integer | User input for the highest an operand value can be | 1 - 100 |
| Tree Node Count | Integer | Number of total nodes in an Individual’s tree used for iterating over nodes to locate crossover and mutation locations |  |
| Tree Y Result(n) | Integer | Result of running the training Xs through the new equation tree individual | Any. Rounding will be fine given the wide range of training data Xs we will use |
| Tree Y Delta (n) | Integer | Absolute value of the difference between the Tree Y Result (n) and the Training Data Y (n) | Any positive integer |
| Individual Fitness Value | Integer | Sum of the Tree Y Deltas from 1 to quantity of training Ys | 0 to any with 0 being an exact match to the target equation result |
| Fitness Bar | Integer | User input percent of the population which will be kept after Natural Selection | 1 - 99 |
| PostSelection Population | List | Even numbered collection of Individual tree equations remaining after Natural Selection |  |
| Crossover TreeIDs A, B | Integer | Randomized IDs of two remaining trees in the PostSelection Population list; trees will be selected and paired for crossover with each other until non remail |  |
| Crossover Trees AMod(n), BMod(n) | EquationTree Array(Nx2) | Randomly selected Pairs of PostSelection trees which will be/have been targets for the crossover section of the other in the Crossover operation |  |
| Crossover NodeIDs A, B | Integer | Randomized Node ID of each of two Individuals’ trees that will be the crossover target in each |  |
| Crossover SubTrees A.B | EquationTree | Subset of each of the crossover individuals’ trees which will be placed at the Crossover Node ID in the other tree |  |
| Mutation NodeID(n) | Integer | Randomized NodeID generated for each tree in Crossover Trees population |  |
| Most Fit Equation | Integer | ID of the tree after crossover and mutation whose resulting fitness value is the lowest |  |
| Mutation Operator Node | Integer | Random replacement operator for a node in a crossover tree |  |
| Mutation Operand Node | Integer | Random replacement operand for a node in a crossover tree |  |

**Data Dictionary (alpha sort)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Name** | **Data Type** | **Description/Process** | **Possible Values** |
| Crossover NodeIDs A, B | Integer | Randomized Node ID of each of two Individuals’ trees that will be the crossover target in each |  |
| Crossover SubTrees A.B | EquationTree | Subset of each of the crossover individuals’ trees which will be placed at the Crossover Node ID in the other tree |  |
| Crossover TreeIDs A, B | Integer | Randomized IDs of two remaining trees in the PostSelection Population list; trees will be selected and paired for crossover with each other until non remail |  |
| Crossover Trees AMod(n), BMod(n) | EquationTree Array(Nx2) | Randomly selected Pairs of PostSelection trees which will be/have been targets for the crossover section of the other in the Crossover operation |  |
| Fitness Bar | Integer | User input percent of the population which will be kept after Natural Selection | 1 - 99 |
| Individual Fitness Value | Integer | Sum of the Tree Y Deltas from 1 to quantity of training Ys | 0 to any with 0 being an exact match to the target equation result |
| Individual ID | Integer | Unique identifier for each candidate in a population | 100-1000 |
| Individual Tree(n) | EquationTree | Collection and sequence of Nodes making up a candidate equation | Example:  (x \* x + 20 -6 / 4/ 9 \* 114) / (2 \* x + 5 – 6 / 2 – x) |
| Most Fit Equation | Integer | ID of the tree after crossover and mutation whose resulting fitness value is the lowest |  |
| Mutation NodeID(n) | Integer | Randomized NodeID generated for each tree in Crossover Trees population |  |
| Mutation Operand Node | Integer | Random replacement operand for a node in a crossover tree |  |
| Mutation Operator Node | Integer | Random replacement operator for a node in a crossover tree |  |
| Node ID | Integer | Unique identifier of a specific tree node | 1-100? |
| Population | List | List of Individual Trees for fitness evaluation |  |
| Population Size | Integer | Quantity of individual equations in the population | 100-1000 |
| PostSelection Population | List | Even numbered collection of Individual tree equations remaining after Natural Selection |  |
| Program Duration | Integer | Minutes input by user for the length of time the program has to run | Default: 15 |
| Target Equation | EquationTree | The target for individuals’ fitness when being evaluated  (Not sure this is needed in the program as long as training data is provided as input) | ( (x^2) - 1 / 2) |
| Training Data | Nx2 Array | Collection of the training X and Y pairs for the target equation | (x1, y1) – (xN, yN) |
| Training Data Size | Integer | Number of X and Y pairs in the training data that will be used to iterate over the training data with each Individual equation |  |
| Tree Depth | Integer | Limit to number of levels deep a candidate tree can be | 1-30? |
| Tree Node Count | Integer | Number of total nodes in an Individual’s tree used for iterating over nodes to locate crossover and mutation locations |  |
| Tree Operand Max | Integer | User input for the highest an operand value can be | 1 - 100 |
| Tree Operand Min | Integer | User input for the lowest an operand value can be | -100 - 0 |
| Tree Operand Node | TreeNode | Operator or Operand node of the candidate equation tree | x, integers -100 – 100? |
| Tree Operator Node | TreeNode | Operator or Operand node of the candidate equation tree | + - \* / |
| Tree Y Delta (n) | Integer | Absolute value of the difference between the Tree Y Result (n) and the Training Data Y (n) | Any positive integer |
| Tree Y Result(n) | Integer | Result of running the training Xs through the new equation tree individual | Any. Rounding will be fine given the wide range of training data Xs we will use |