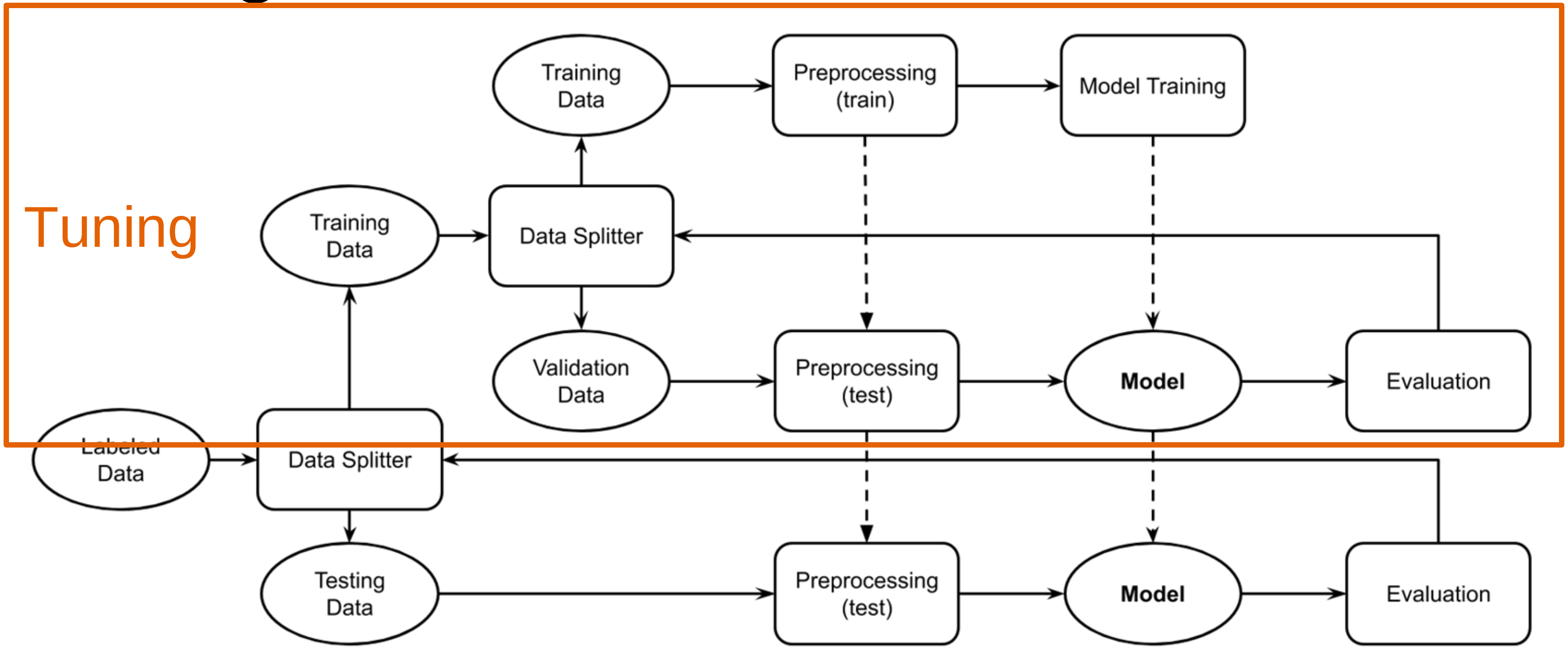
Foundations of Data Science & Analytics: Hyperparameter Tuning

Ezgi Siir Kibris

Introduction to Data Mining, 2nd Edition
by
Tan, Steinbach, Karpatne, Kumar

Tuning


Tuning
Training
Data
Preprocessing
(train)
Model Training
Training
Data
Data Splitter
Validation
Data
Preprocessing
(test)
Model
Evaluation
Labeled
Data
Data Splitter
Testing
Data
Preprocessing
(test)
Model
Evaluation

Tuning

A black box **optimization** process:

- Search within a **decision** space
    - E.g. criterion and max_depth of a decision tree learner
- To find the set of decisions that generate optimal (or suboptimal) **objectives**
    - E.g. f1 score and runtime

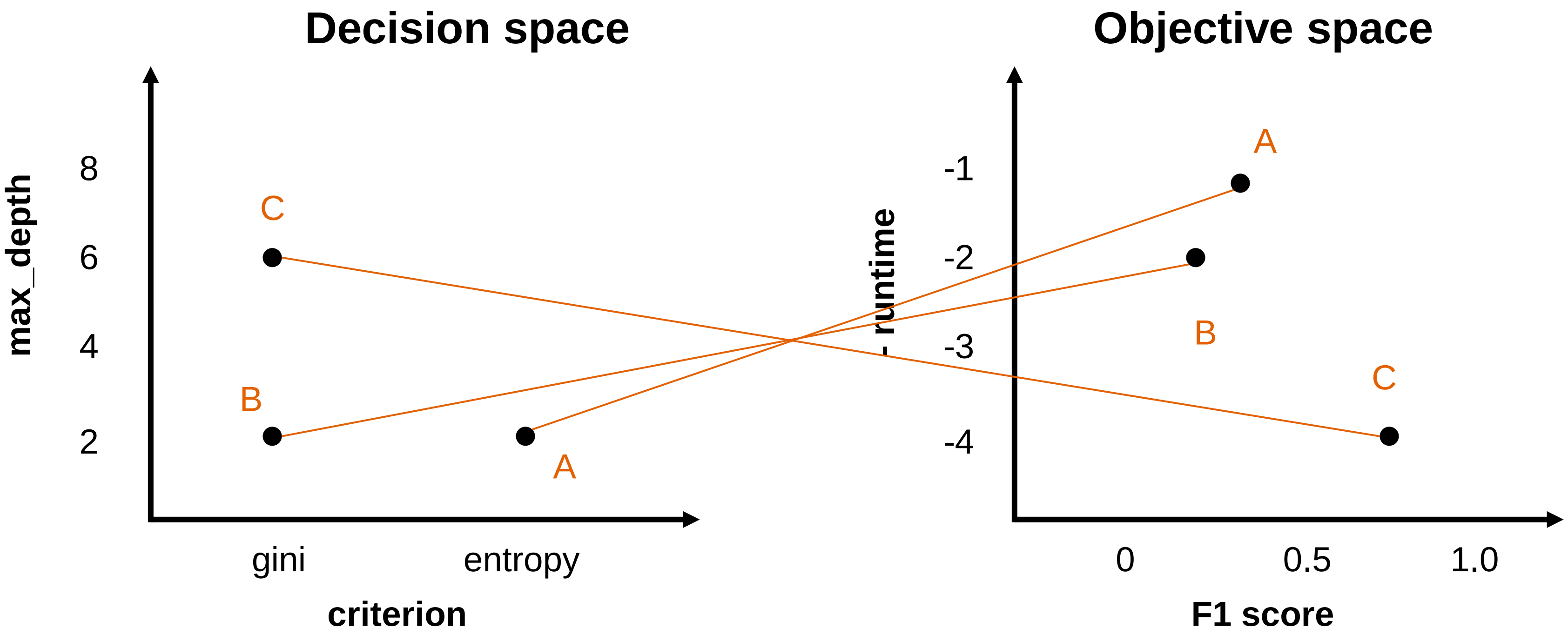
Tuning

A **black box** optimization process:

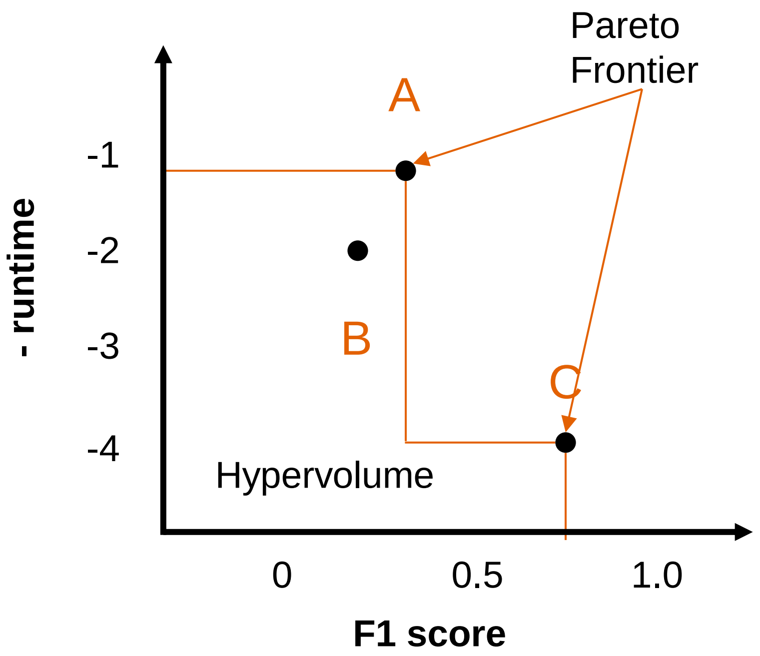
- Optimize $\max_{x} y = f(x)$

where x = decisions (hyperparameters)
y = objectives (performance metrics)

Tuning


Decision space
max_depth
8
6
4
2
C
B
A
gini
entropy
criterion
Objective space
- runtime
-1
-2
-3
-4
A
B
C
0
0.5
1.0
F1 score

Which one is better?



- Single objective:
  - No problem
- Multi-objective:
  - Binary domination
  - A binary dominates B if
    - A is not worse than B in any objective

  AND
    - A is better than B in at least one objective

How to find the optimal decisions for multiple objectives?

- Grid search
- Random search
- Evolutionary algorithm

Grid search

- Exhaustive search of the decision space
- Embarrassingly parallel
- Curse of dimensionality

- max_depth = {1, 2, 3, 4, 5, 6, 7, 8}
- criterion = {gini, entropy}
- 8 * 2 = 16 evaluations

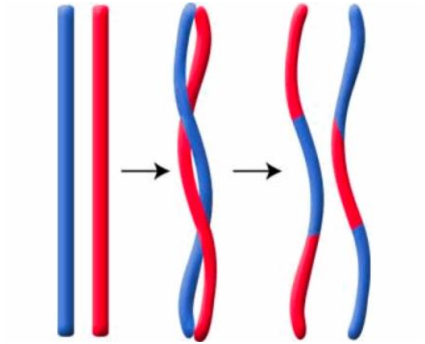
Random search

- Random sample from the decision space
- Embarrassingly parallel
- Outperforms grid search in problems of low intrinsic dimensionality (only a few decisions affect the objectives)

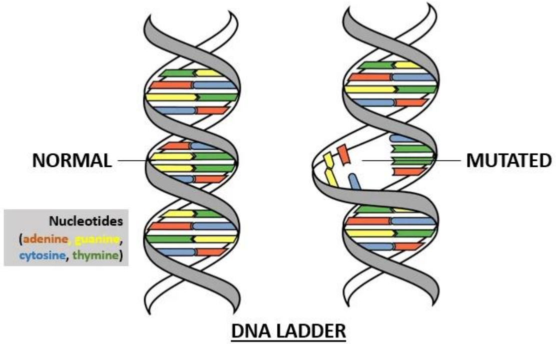
Evolutionary algorithm

- **Continuity Assumption:**
  - Two points are close in decision space -> also close in objective space
- **Exploitation:**
  - Evaluate points that are similar to the current best performing ones.
- **Exploration:**
  - Evaluate points in the most unseen regions (or just randomly)

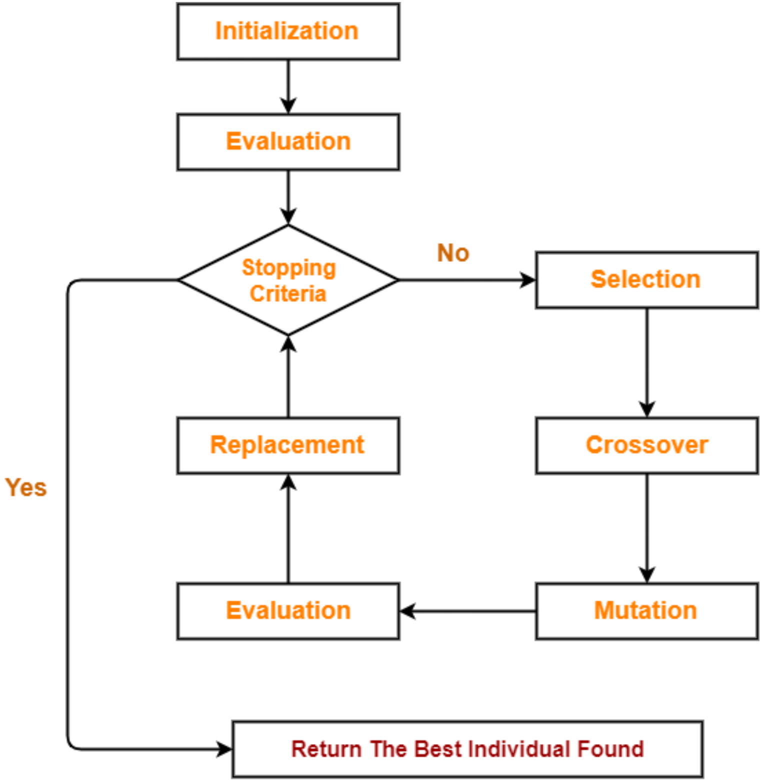
Genetic Algorithm (GA)



Crossover



Mutation

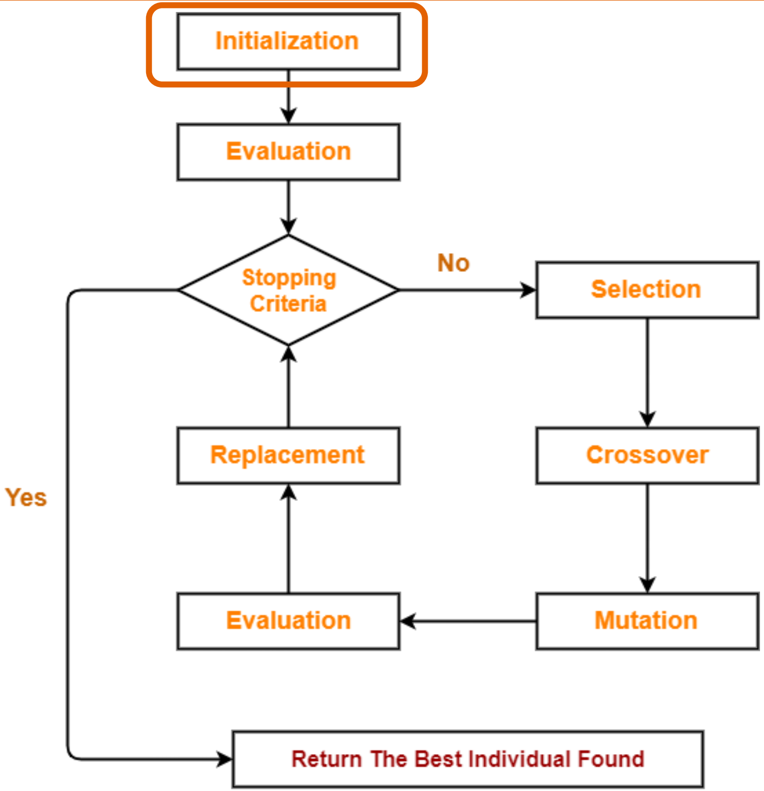


How Genetic Algorithm Works

Selection:
only the best ones survive

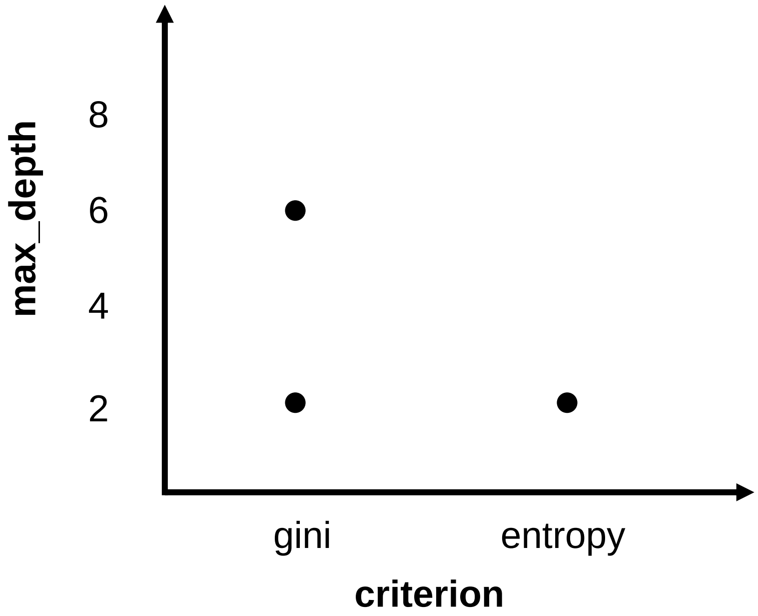
Initialization

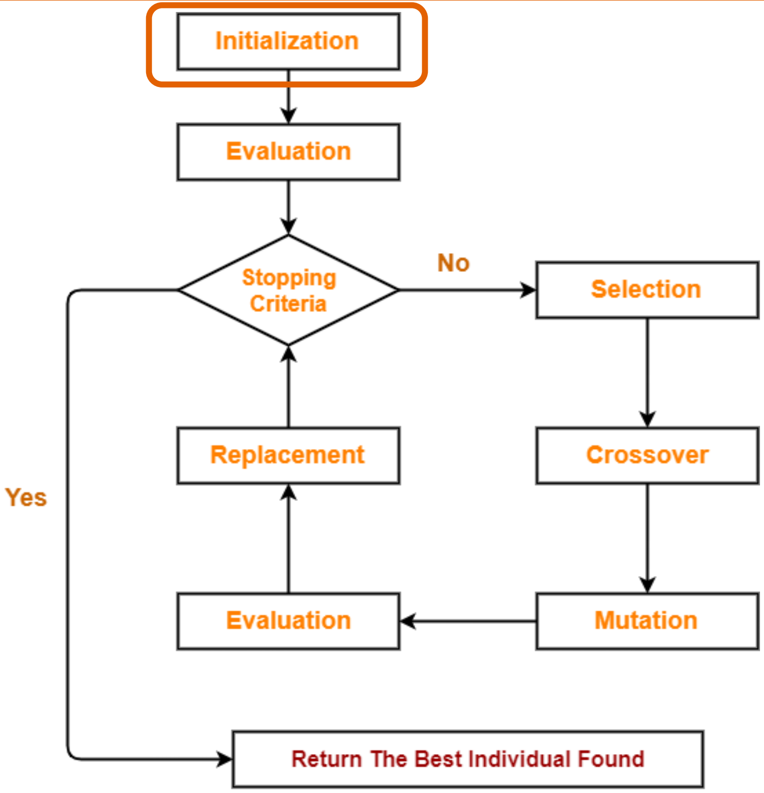
Random sample without replacement K points from the decision space to form **Generation 0.**



How Genetic Algorithm Works

Initialization

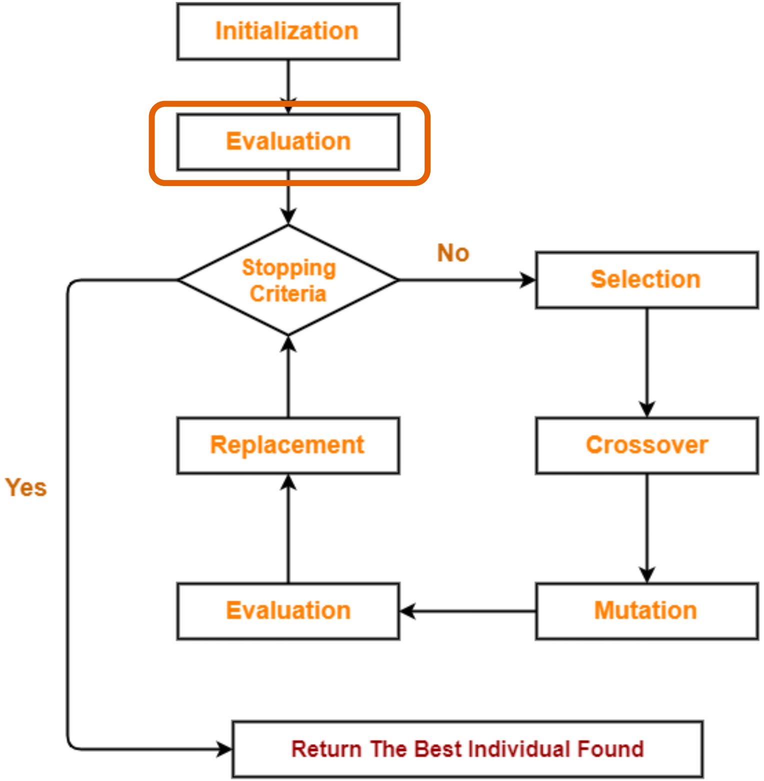




How Genetic Algorithm Works

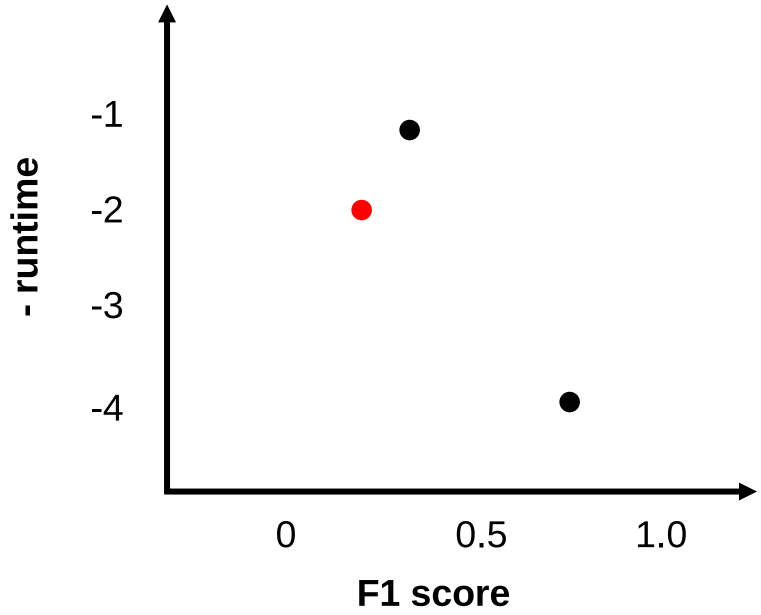
Evaluation

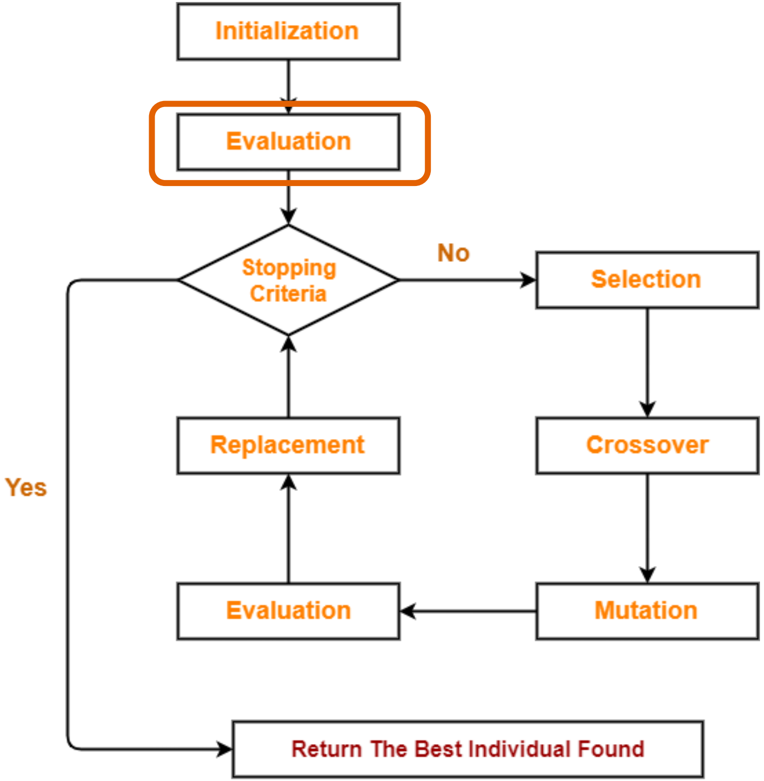
- Cross-validation:
  - Train model with each set of decisions as parameters.
  - Test on validation set to collect performance metrics
- Summarize objectives



How Genetic Algorithm Works

Evaluation

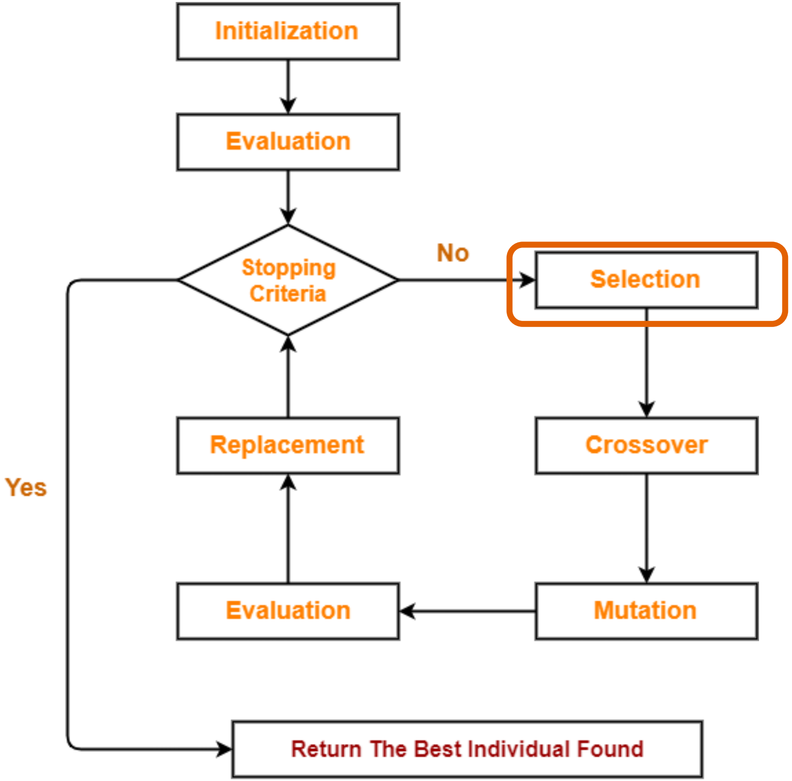




How Genetic Algorithm Works

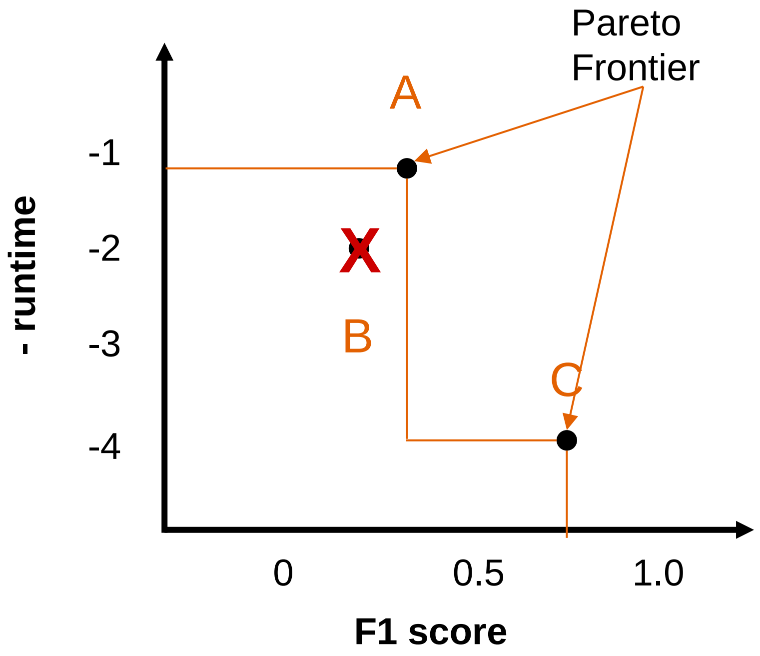
Selection

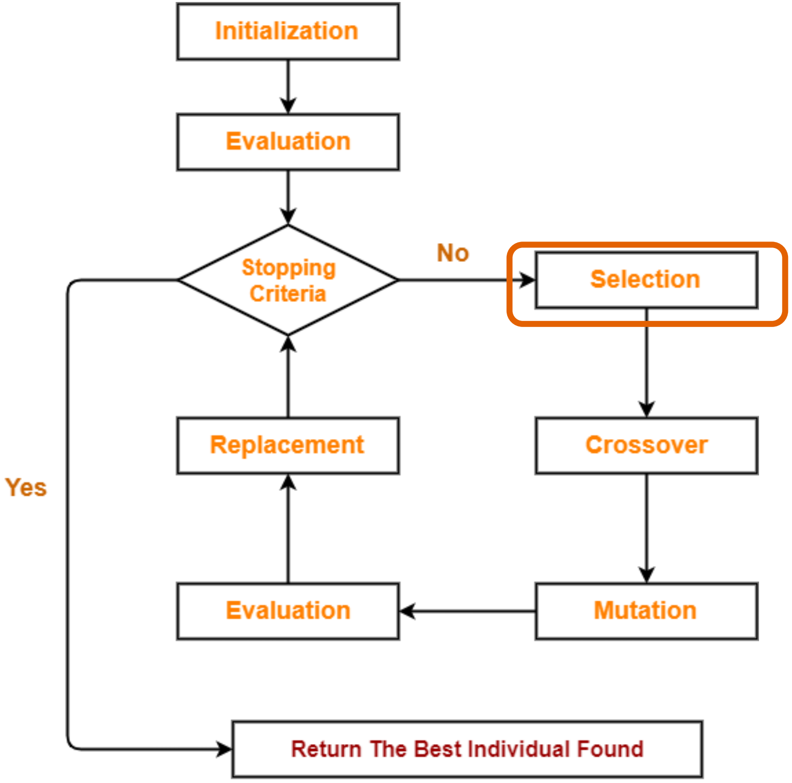
- Single objective:
  - Select M points that achieve best objective (e.g. M = 0.5 K).
- Multi-objective:
  - Select points that are NOT dominated (**Pareto Frontier**) by any point.



How Genetic Algorithm Works

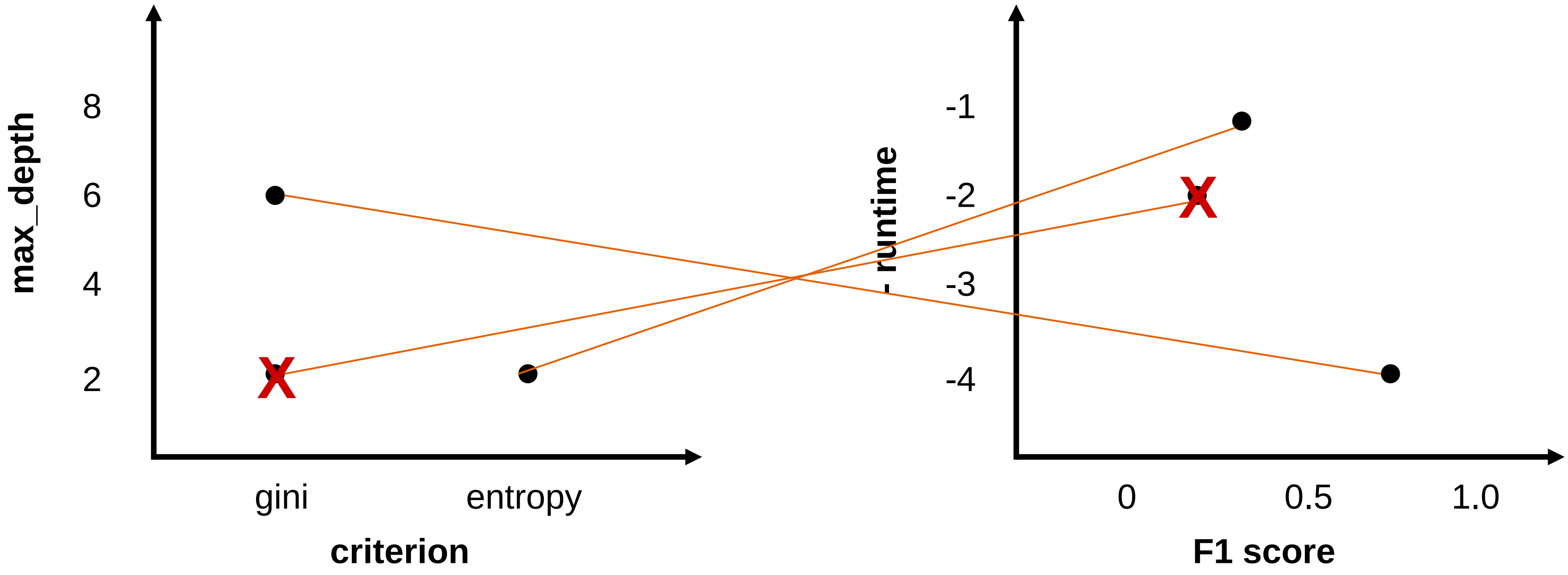
Selection



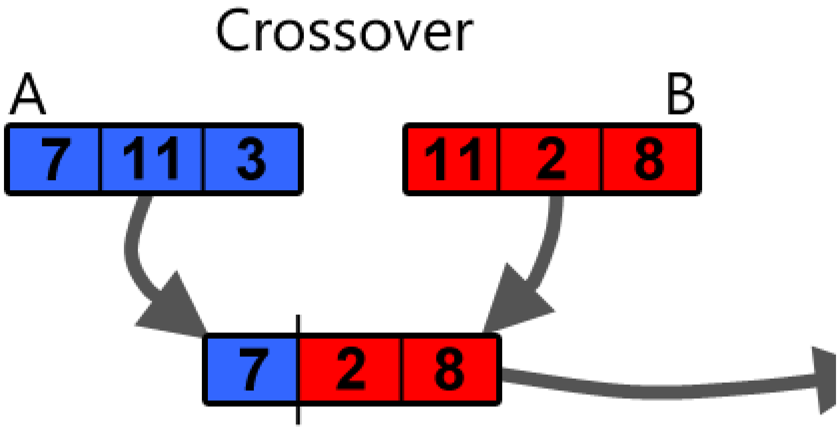


How Genetic Algorithm Works

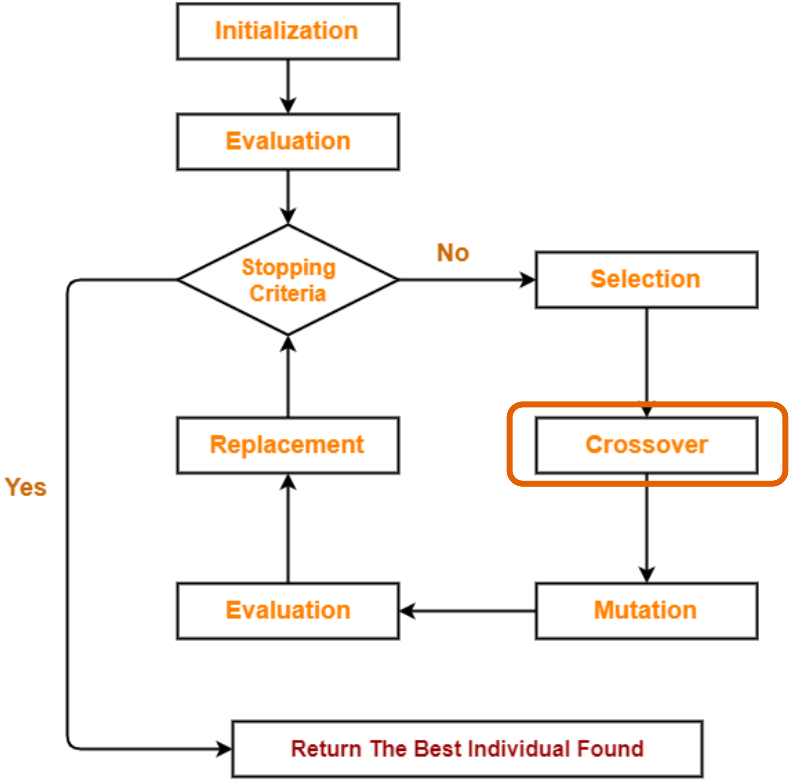
Selection


max_depth
8
6
4
2
gini
entropy
criterion
- runtime
-1
-2
-3
-4
0
0.5
1.0
F1 score
X
X

Crossover (exploitation)

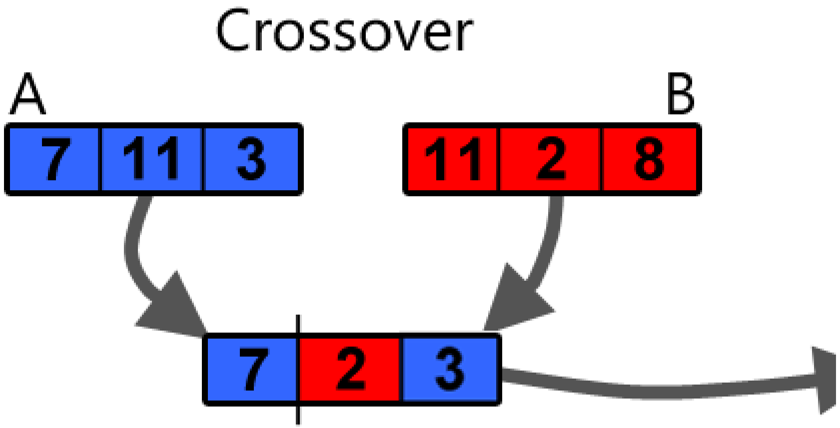


- **Single-point crossover**
  - random a number 0<k<n
  - 0 to k-1 bit from A, k to n bit from B
- **Uniform crossover**
  - each bit has equal probability from A or B

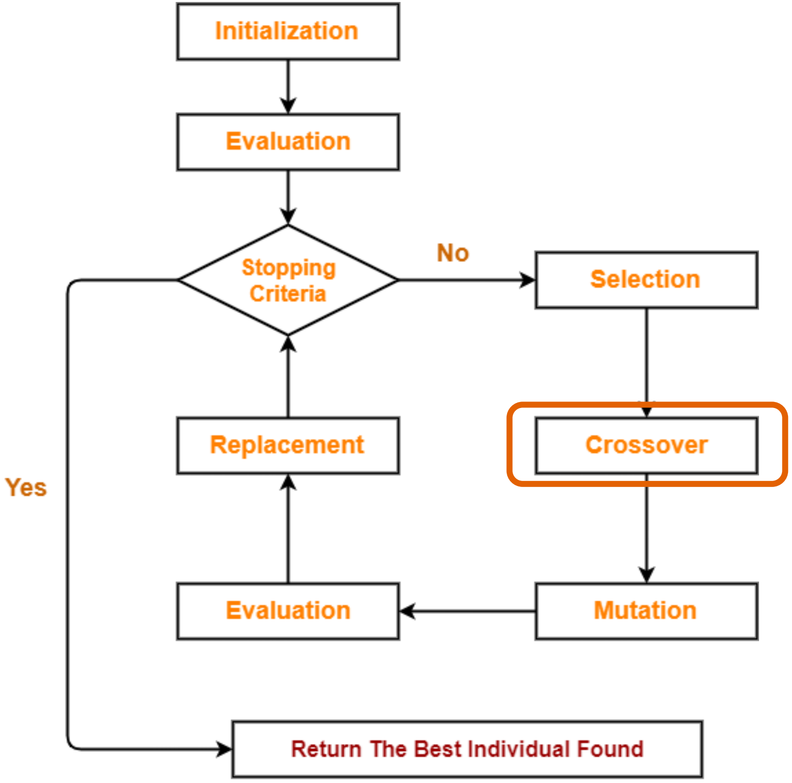


How Genetic Algorithm Works

Crossover (exploitation)

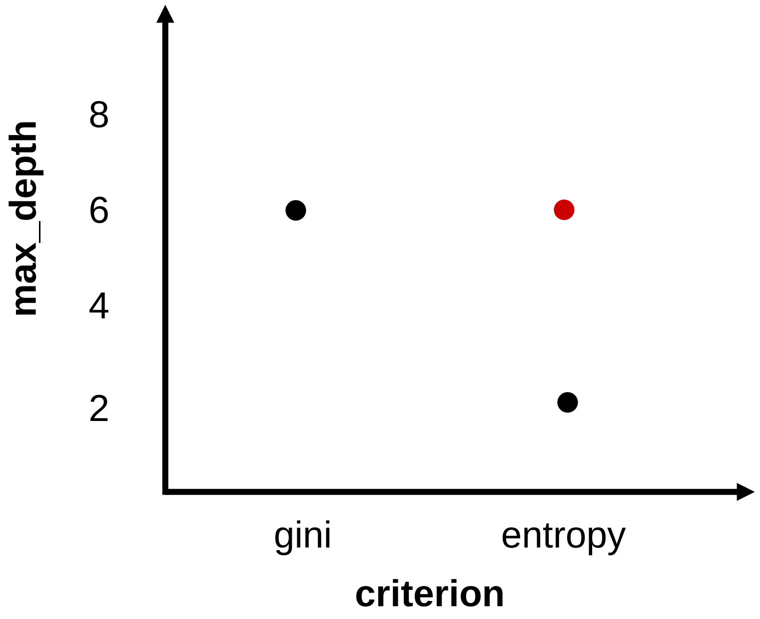


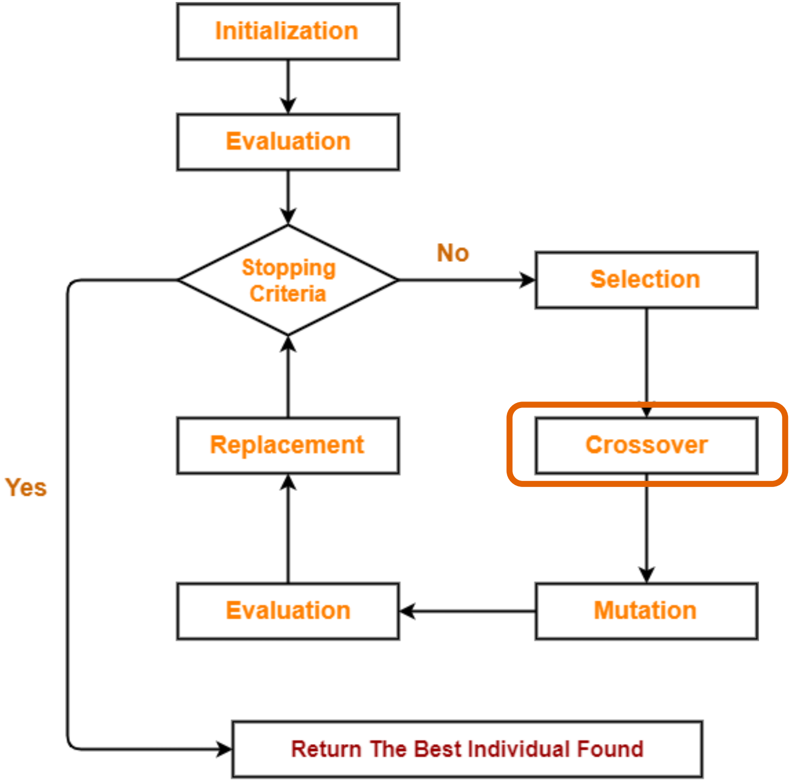
- Randomly select two points A and B
- Crossover and generate C
- Repeat until size of generation is K



How Genetic Algorithm Works

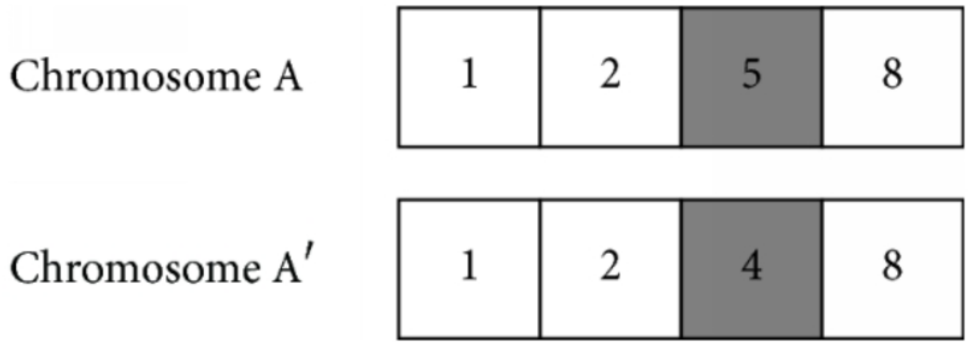
Crossover



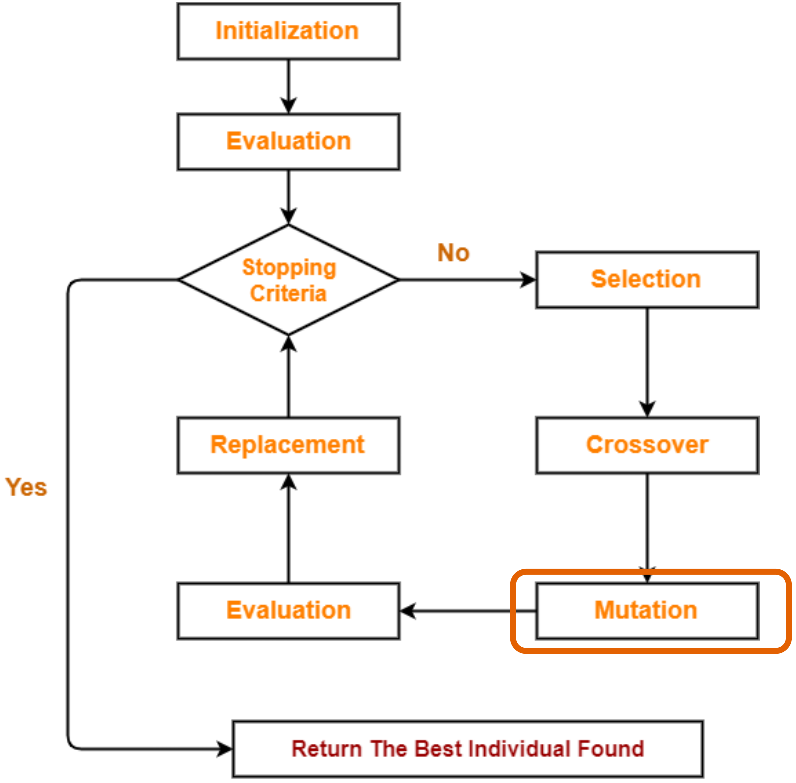


How Genetic Algorithm Works

Mutation (exploration)

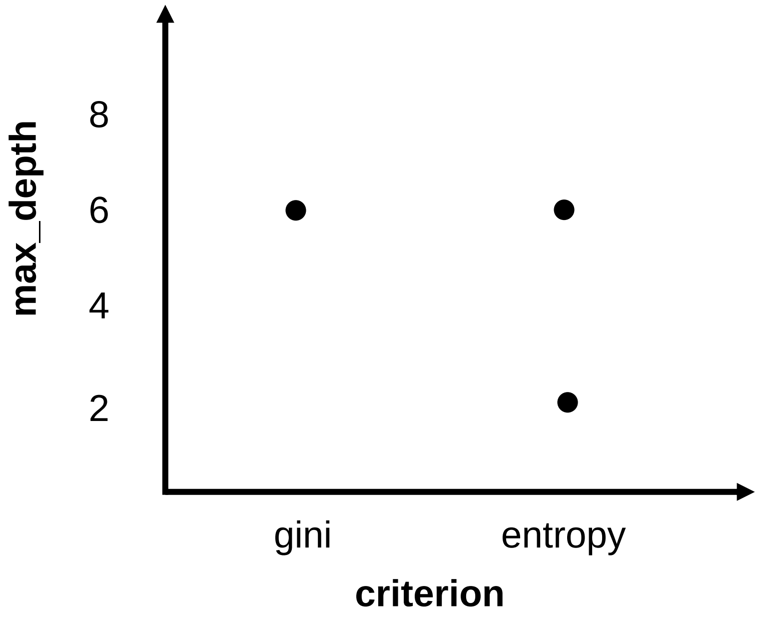


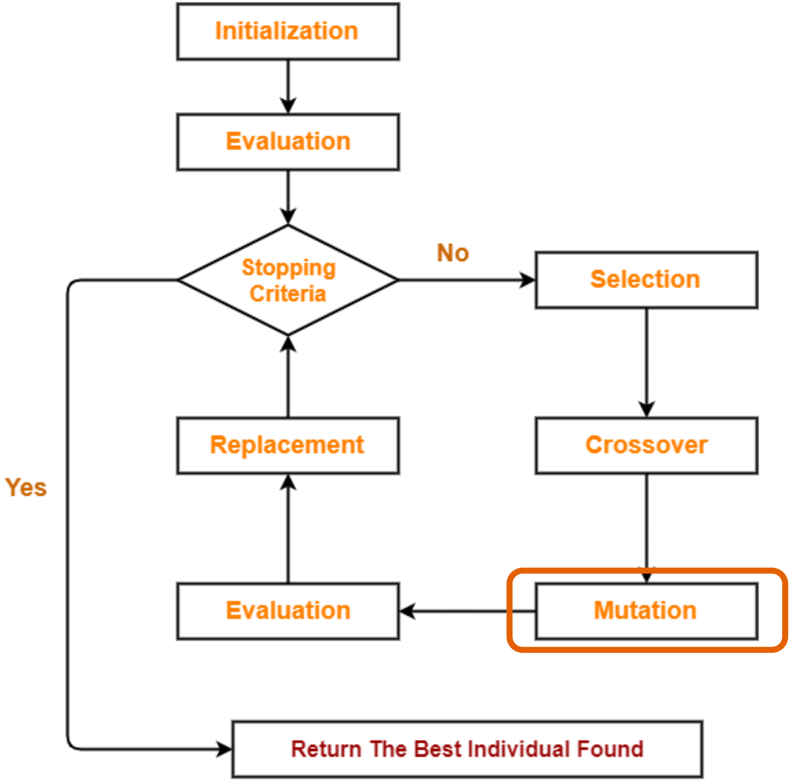
- Each bit has equal probability of being mutated
- To a random valid value except the original one



How Genetic Algorithm Works

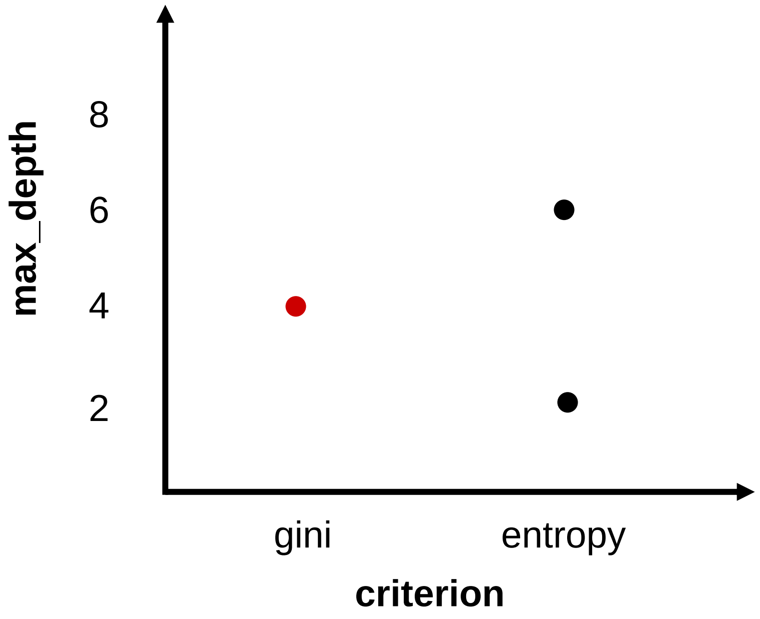
Mutation (exploration)

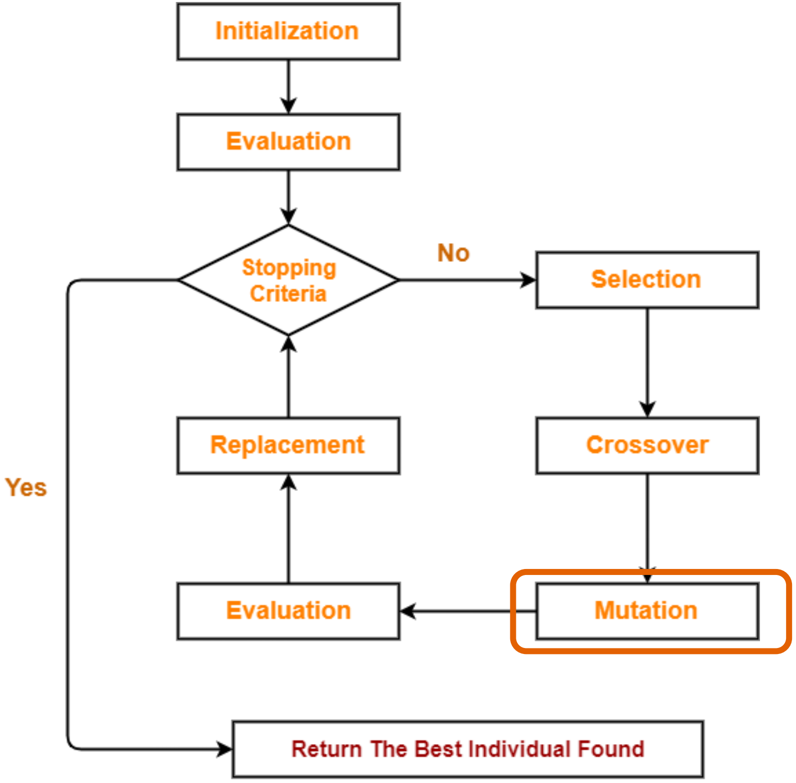




How Genetic Algorithm Works

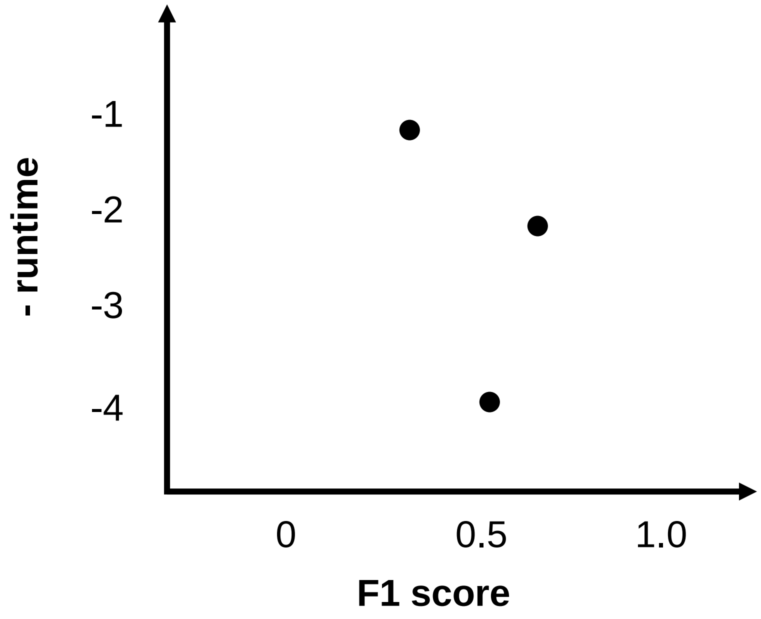
Mutation (exploration)

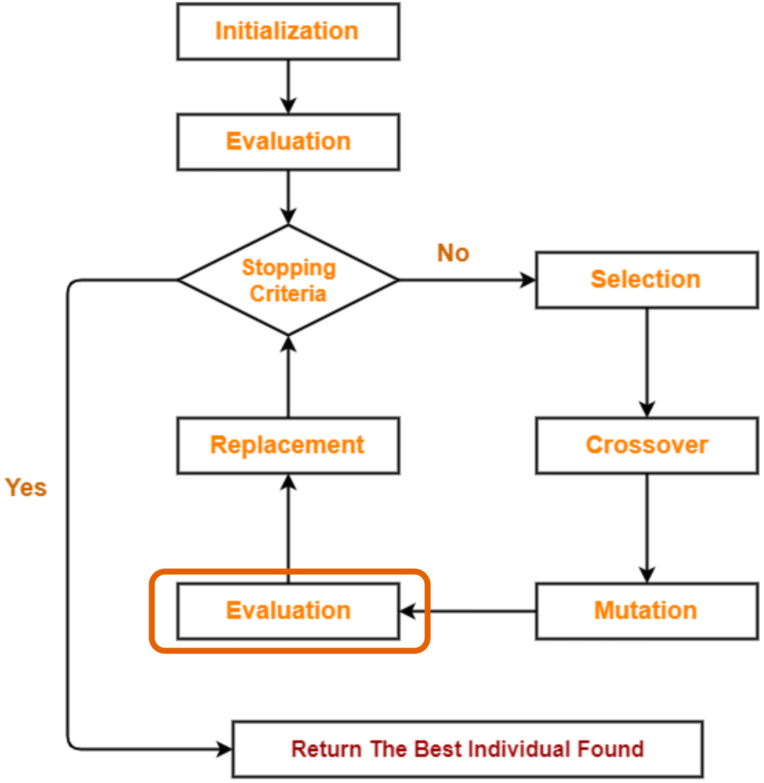




How Genetic Algorithm Works

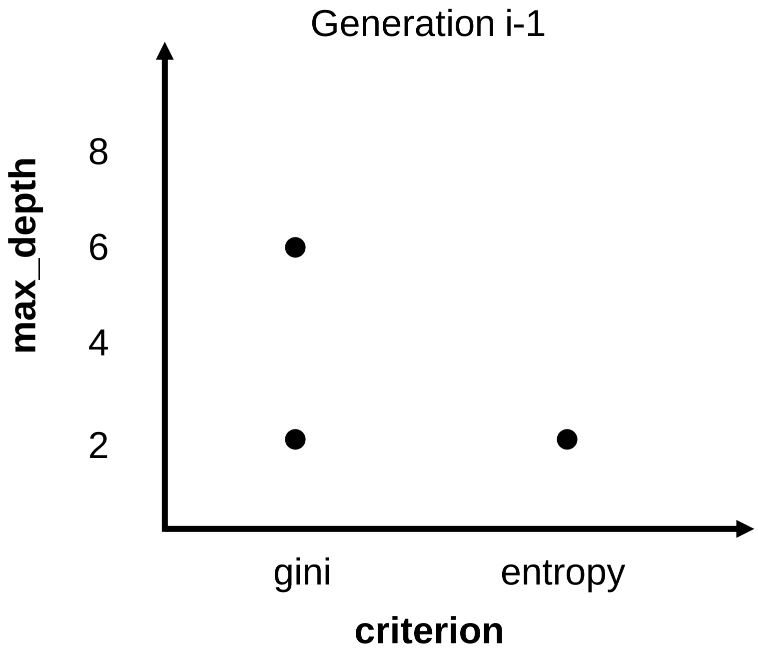
Evaluation

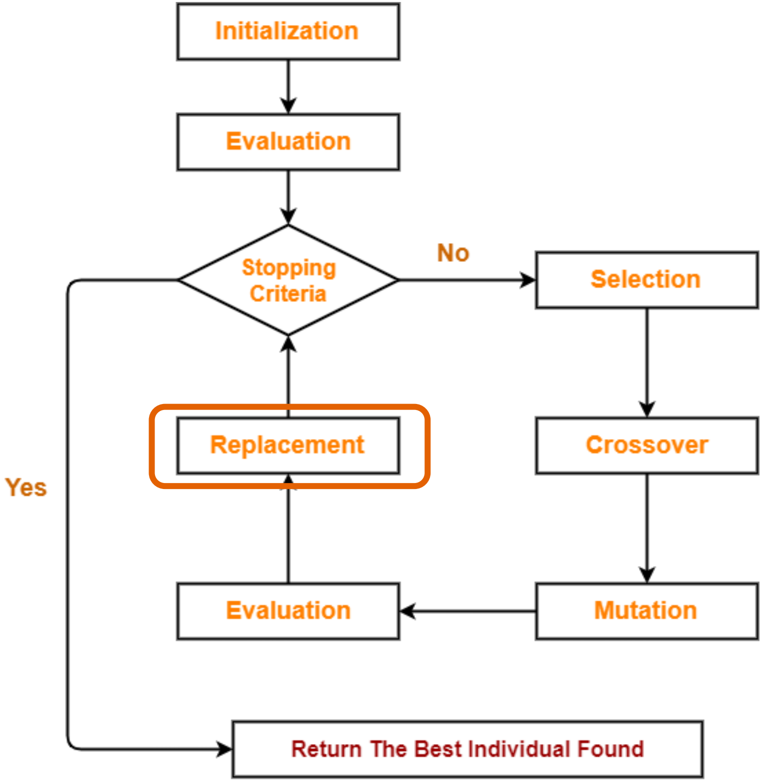




How Genetic Algorithm Works

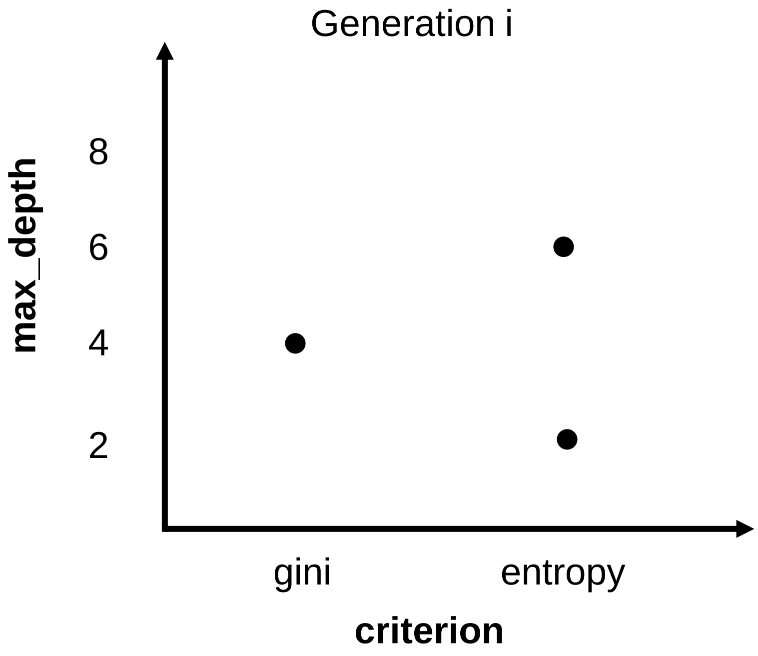
Replace

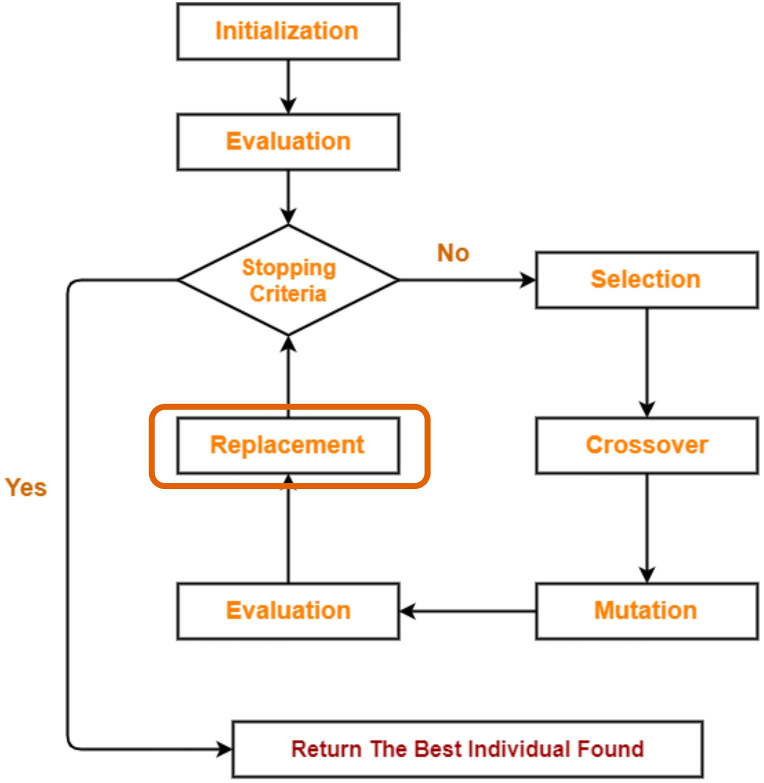




How Genetic Algorithm Works

Replace

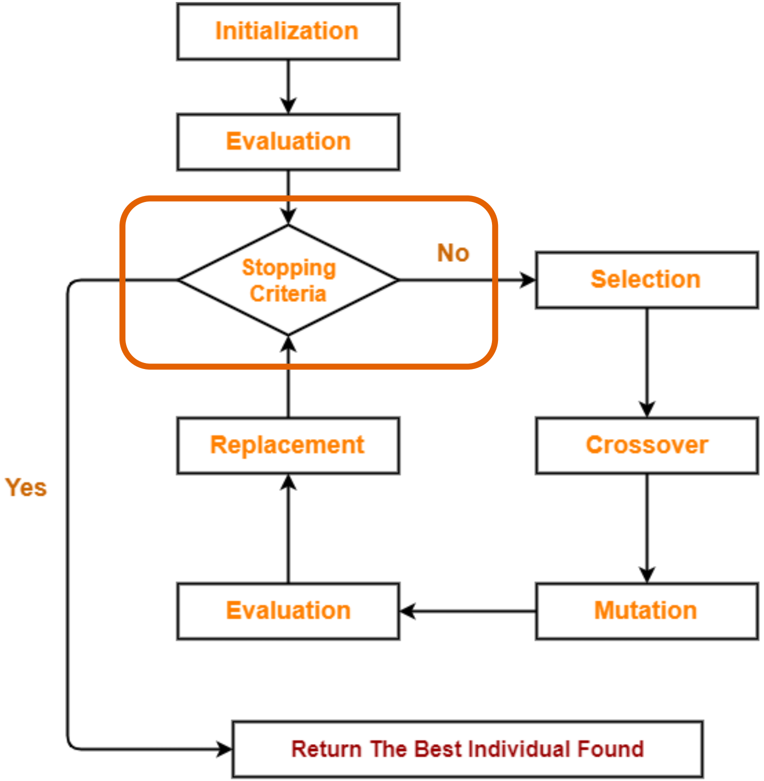




How Genetic Algorithm Works

Stop

- When generation i is the same as generation i-1
- Or when best point stays unchanged for generations



How Genetic Algorithm Works

Evolutionary algorithm

- **Advantage:**
  - More efficient (than random and grid search): find better solution with fewer evaluations when decision space is large.
- **Disadvantage:**
  - Does not guarantee global optimum.
  - Needs communication in parallelization.