Thuật toán di truyền là gì?

Một thuật toán di truyền thuộc về một lớp các thuật toán tiến hóa rộng rãi lấy cảm hứng từ quá trình tiến hóa sinh học. Tất cả chúng ta đều biết về sự tiến hóa sinh học [1] — đó là một chọn lọc bố mẹ, sinh sản và biến dị của đời con. Mục đích chính của quá trình tiến hóa là để sinh ra những thế hệ con cái tốt hơn về mặt sinh học so với bố mẹ của chúng. MỘT thuật toán di truyền chủ yếu dựa trên Thuyết tiến hóa của Darwin bằng cách chọn lọc tự nhiên, và nó cố gắng mô phỏng giống nhau. Trực giác cơ bản là lựa chọn những cá thể tốt nhất làm bố mẹ từ quần thể, yêu cầu họ kéo dài thế hệ của mình bằng cách sinh sản và sinh con trong quá trình sinh sản nơi các gen của cả bố và mẹ giao nhau ở đó xảy ra một lỗi được gọi là đột biến. Những đứa trẻ này một lần nữa được yêu cầu sinh sản con cái của họ, và quá trình này tiếp tục, dẫn đến các thế hệ khỏe mạnh hơn. Cái này lý thuyết đã truyền cảm hứng cho tính toán tiến hóa để giải quyết các vấn đề tối ưu hóa, lựa chọn tính năng, vấn đề ba lô cổ điển, v.v.

Hãy tìm hiểu ứng dụng của giải thuật di truyền với bài toán cái ba lô. Giả sử chúng ta đang đi tìm kho báu, và sau tất cả những nỗ lực và chăm chỉ, cuối cùng chúng ta tìm kho báu trong một hang động sâu bên dưới đầy đồ trang sức bằng vàng và kim cương. đầu tiên điều chúng tôi muốn làm là lấp đầy ba lô của mình với càng nhiều đồ trang trí càng tốt. Tuy nhiên, một số tham số phải được quan tâm trong bài toán của chúng ta, và ba lô có không gian hạn chế. Nó không thể mang trọng lượng hơn 35 kg. Tiếp theo, chúng ta phải chọn đồ trang trí một cách tối ưu sao cho chiếc ba lô không bị quá tải, tất cả các đồ trang trí chúng tôi chọn phải có giá trị cao, và một đồ trang trí không được làm hỏng cái khác trong ba lô - đây là nơi thuật toán di truyền đi vào hoạt động để tối ưu hóa vấn đề của chúng tôi bằng cách quan tâm đến tất cả các tham số Bây giờ chúng ta đã có một ý tưởng cơ bản về thuật toán di truyền. Hãy xem các bước liên quan và viết mã triển khai của chúng tôi bằng Python. Các bước trong giải thuật di truyền

1. Khởi tạo dân số

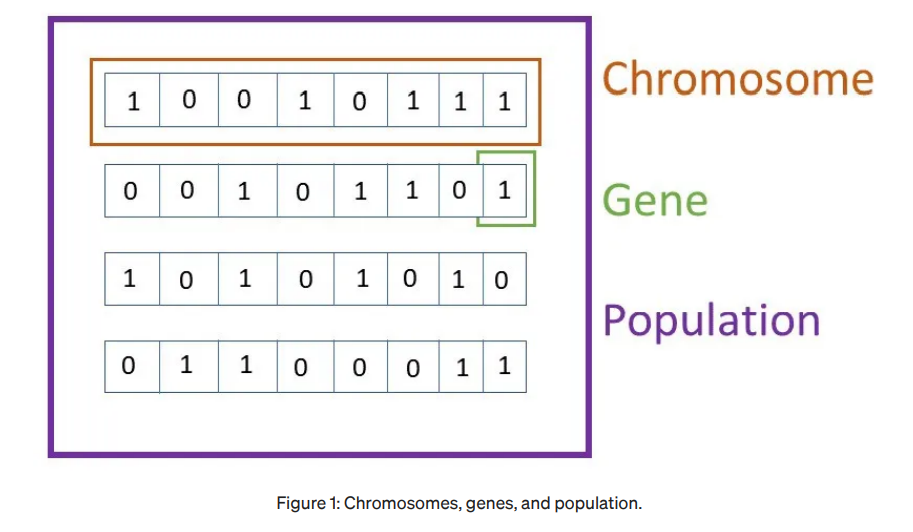
2. Chọn cha mẹ bằng cách đánh giá thể lực của họ

3. Lai bố mẹ để sinh sản

4. Đột biến thế hệ con cháu

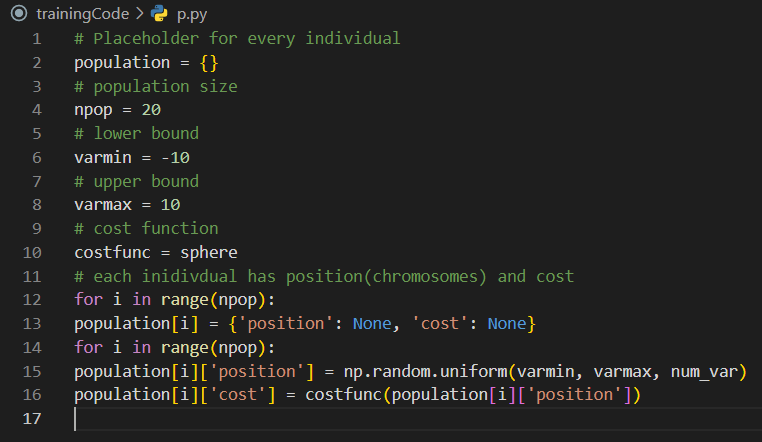
5. Đánh giá con lai

6. Hợp nhất quần thể con với quần thể chính và sắp xếp

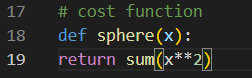
“Trong một giải thuật di truyền, một quần thể các giải pháp ứng cử viên (được gọi là các cá thể, sinh vật hoặc kiểu hình) cho một vấn đề tối ưu hóa được phát triển theo hướng tốt hơn các giải pháp. Mỗi giải pháp ứng cử viên có một tập hợp các thuộc tính (nhiễm sắc thể hoặc kiểu gen) có thể bị đột biến và thay đổi; theo truyền thống, các giải pháp được thể hiện trong nhị phân dưới dạng chuỗi 0 và 1, nhưng cũng có thể mã hóa khác [2] [3]” 

1. Khởi tạo

Thuật toán thường bắt đầu với dân số được tạo ngẫu nhiên. kích thước của dân số phụ thuộc vào bản chất của vấn đề. Chúng ta có thể sử dụng mã hóa 0 và 1. Tuy nhiên, trong hướng dẫn này, chúng tôi sẽ sử dụng các số được phân phối đồng đều để đại diện cho mỗi gen.



Chúng tôi tạo một từ điển để giữ dân số và mỗi cá nhân được liên kết với nhiễm sắc thể (vị trí) và chi phí. Vị trí được lấp đầy với tạo ngẫu nhiênm số (gen) phân bố đồng đều với giới hạn dưới -10 và giới hạn trên +10. Chi phí là chức năng chi phí mà chúng tôi đang cố gắng tối ưu hóa. Trong hướng dẫn này, chúng ta sẽ tối ưu hóa tổng bình phương của x, trong đó x là gen riêng lẻ của mỗi nhiễm sắc thể.



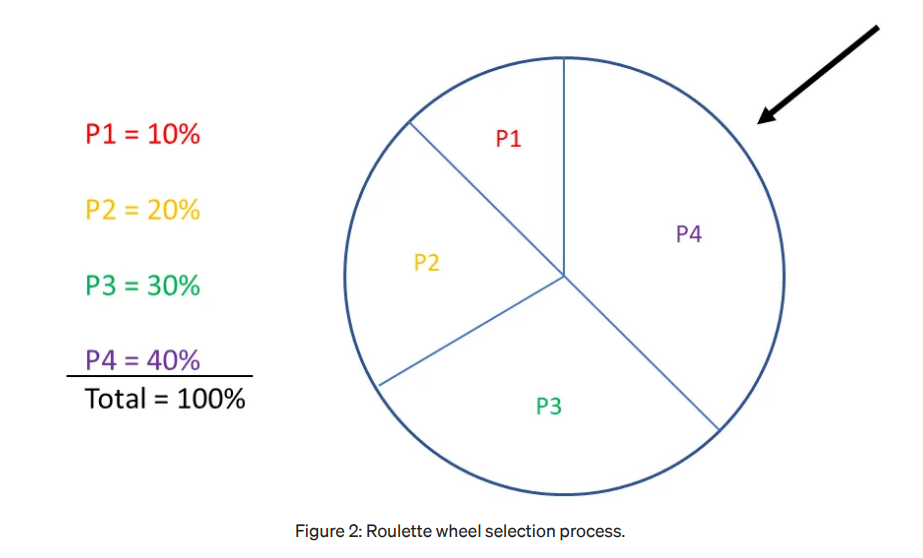
2. Lựa chọn phụ huynh

Trong mỗi thế hệ kế tiếp, một phần dân số hiện tại được chọn để tạo ra một thế hệ mới. Giải pháp cá nhân được lựa chọn thông qua một tập thể dục dựa trên quá trình [2]. Vì chúng ta đang ở thế hệ 0, chúng ta không có hậu duệ. Chúng tôi chọn cha mẹ từ dân số được tạo ngẫu nhiên của chúng tôi. Có ba phương pháp chính để xác định những cá thể phù hợp nhất và chọn lọc để nhân giống. Chọn ngẫu nhiên: Đây là cách chọn đơn giản và kém hiệu quả nhất. cha mẹ. Trong phương pháp này, chúng tôi xáo trộn dân số bằng cách thực hiện hoán vị và chọn 2 cá thể đầu tiên làm bố mẹ để lai tạo. Phương pháp này không phải là được khuyên dùng vì nó không tuân theo “Thuyết tiến hóa của Darwin bởi Chọn lọc tự nhiên,” trong đó các cá nhân được lựa chọn dựa trên thể lực của họ chứ không phải ngẫu nhiên.

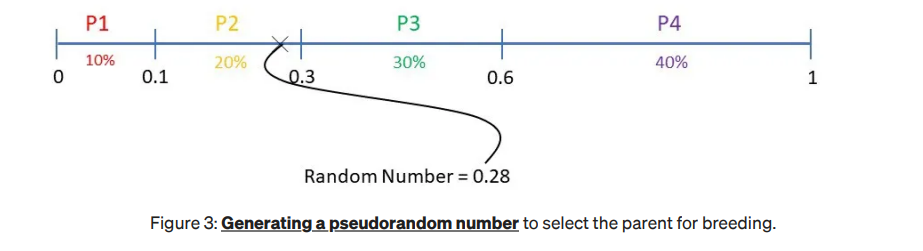
Lựa chọn giải đấu: Phương pháp này dựa trên xác suất lựa chọn của mỗi cá nhân. Chúng tôi tổ chức một số giải đấu giữa một nhóm được chọn ngẫu nhiên của các cá nhân, chọn một cá nhân từ mỗi nhóm là người chiến thắng và chạy lại giải đấu bằng cách nhóm những người chiến thắng từ lần lặp đầu tiên, hãy lặp lại quy trình cho đến khi chúng ta hội tụ được hai bố mẹ thắng cuộc để lai tạo. Thành viên tốt nhất của mỗi người nhóm trong mỗi lần lặp lại có xác suất được chọn cao nhất.

Lựa chọn bánh xe Roulette: Đây là một phương pháp được sử dụng rộng rãi và hiệu quả nhất cho tuyển chọn bố mẹ; do đó chúng tôi sẽ sử dụng nó ngày hôm nay trong thuật toán của chúng tôi. tất cả chúng ta đều biết cách hoạt động của bánh xe roulette trong sòng bạc, thả bóng, quay bánh xe và chờ đợi cho đến khi bánh xe dừng lại để xem quả bóng rơi vào chiếc bình nào. Hãy tìm hiểu sâu hơn về phần thực hiện.

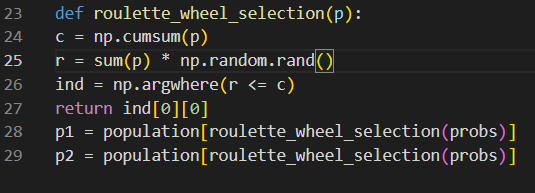
Sự khác biệt duy nhất giữa bánh xe roulette sòng bạc và bánh xe roulette phương pháp lựa chọn cha mẹ là trong bánh xe roulette của sòng bạc, mỗi bình có số tiền bằng nhau xác suất giữ quả cầu khi bánh xe ngừng quay. Tuy nhiên, ở đây chúng tôi xác định xác suất cho mỗi nồi (cá nhân của quần thể). xác suất của mỗi cá nhân gọi là thể lực của cá nhân.



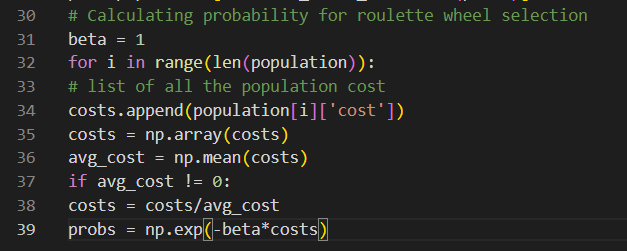
Chúng tôi có bốn cha mẹ P1, P2, P3 và P4, với xác suất được chọn cho giống lần lượt là 0,1; 0,2; 0,3; 0,4. Mũi tên được cố định tại một nơi, và bánh xe là xoay. Khi bánh xe ngừng quay, cha mẹ nơi mũi tên chỉ vào là được chọn để nhân giống—xác suất càng lớn thì diện tích trên bánh xe càng lớn, dẫn đến để xác suất được chọn cao hơn. Bây giờ, làm cách nào để triển khai bánh xe roulette theo chương trình? Chúng tôi mở bánh xe thành dòng đồng nhất và chia dòng cho số bố mẹ trong quần thể, và mỗi phụ huynh chiếm không gian trên dòng bằng xác suất của nó được chọn và mỗi điểm cắt là tổng xác suất tích lũy. Tạo một số ngẫu nhiên giữa 0 và 1 sẽ hoạt động giống như mũi tên chọn cha mẹ cho chăn nuôi. Ở đây, số ngẫu nhiên là 0,28; do đó người chiến thắng là P2.



Để làm cho nó đơn giản hơn nữa, chúng tôi tính toán tổng tích lũy của mỗi xác suất cha mẹ, nhân tổng của nó với một số được tạo ngẫu nhiên. Sau đó lấy chỉ số của cái đầu tiên cha mẹ có giá trị tích lũy lớn hơn số ngẫu nhiên. Ví dụ, P1 có giá trị tích lũy là 0,1, P2 có 0,3, P3 có 0,6 và P4 có 1. Nếu ngẫu nhiên số được tạo là 0,28, thì cha đầu tiên có giá trị tích lũy lớn hơn hơn 0,28 là P2 do đó là bố mẹ chiến thắng để nhân giống. Hàm argwhere() trả về một mảng Đúng và Sai dựa trên biểu thức được truyền dưới dạng tham số.



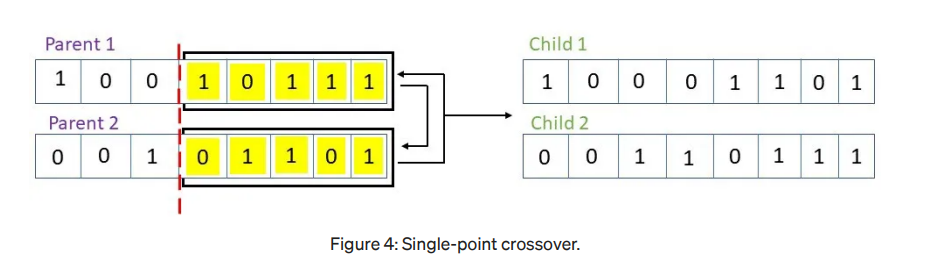
Chúng tôi tính toán xác suất của mỗi phụ huynh theo cấp số nhân của thời gian beta âm chi phí, trong đó beta là số nguyên được xác định trước và chi phí là chi phí của mỗi cấp độ gốc được chia bằng chi phí trung bình của tất cả các bậc cha mẹ trong dân số.



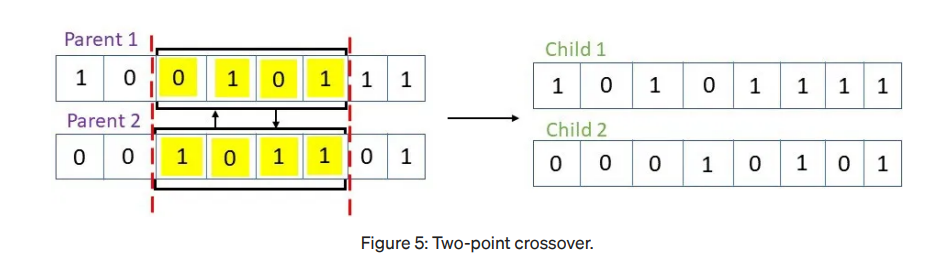
3. Chéo

Bây giờ chúng ta đã có hai bố mẹ để nhân giống, bước tiếp theo là thực hiện chéo/giao phối/sinh sản. Trao đổi chéo đề cập đến quá trình trong đó một số gen nhất định từ cả hai nhiễm sắc thể gốc được chồng lên nhau hoặc trộn lẫn hoặc đổi chỗ để tạo ra nhiễm sắc thể mới con đẻ. Vì thế hệ con là kết quả của sự trao đổi chéo của các nhiễm sắc thể bố mẹ, nên nó thừa hưởng cả đặc điểm của bố và mẹ. Có ba phương pháp để thực hiện chéo.

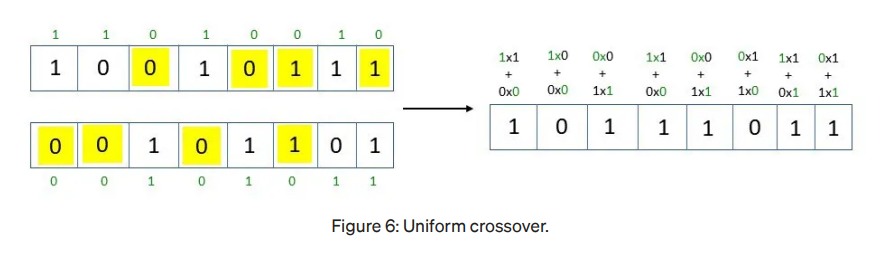
Trao đổi chéo đơn điểm: Trong phương pháp này, cả hai nhiễm sắc thể mẹ được cắt ở cùng một điểm ngẫu nhiên và các phần còn lại được hoán đổi để tạo ra hai điểm mới nhiễm sắc thể con cái. Các gen màu vàng đại diện cho phần cắt của nhiễm sắc thể.



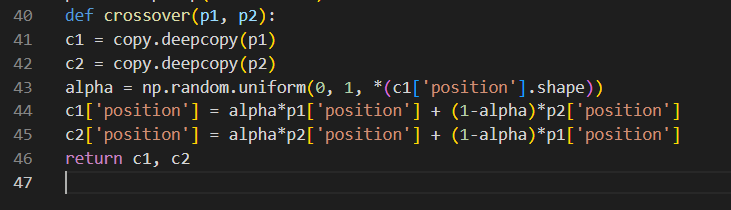
Giao nhau hai điểm: Một phương pháp tương tự như giao nhau một điểm, nhưng duy nhất sự khác biệt là các nhiễm sắc thể cha mẹ được cắt tại hai điểm ngẫu nhiên. Lại, phần bị cắt màu vàng của nhiễm sắc thể được hoán đổi cho nhau.



Trao đổi chéo đồng nhất: Trước tiên, chúng tôi chọn ngẫu nhiên những gen được cho là di truyền từ cả nhiễm sắc thể bố mẹ và gen không di truyền là được đánh dấu bằng màu vàng. Sau đó, chúng tôi mô hình hóa chúng thành các số 0 và 1, được viết bằng màu xanh lục. Gen cần di truyền được mã hóa là 1 và gen nên không được kế thừa được mã hóa thành 0. Chuỗi số 0 và 1 này sẽ được gọi là alpha kể từ bây giờ. Nhân giá trị gen với giá trị alpha tương ứng cho cả cha mẹ và sau đó thêm các kết quả để tạo ra một gen duy nhất của nhiễm sắc thể con cái. Hãy xem xét gen đầu tiên của mỗi nhiễm sắc thể cha mẹ. Đối với cha mẹ-1, giá trị gen là 1 và giá trị alpha tương ứng cũng là 1; do đó, 1x1=1. Đối với cha mẹ-2, giá trị gen là 0 và giá trị alpha tương ứng là cũng 0 do đó, 0x0=0. Gen đầu tiên của nhiễm sắc thể con cái là 1+0=1, và như vậy tiếp tục - theo cách này, chúng ta có con cái-1, để tái tạo con cái-2, chúng ta lấy bổ sung các giá trị của alpha và thực hiện quá trình tương tự.



Theo lập trình, chúng tôi sao chép cả cha mẹ vào biến con: c1, c2. Ngẫu nhiên tạo các giá trị alpha được phân phối đồng đều trong khoảng từ 0 đến 1, là giá trị gốc hình dạng (vị trí) của nhiễm sắc thể. Phần còn lại của quá trình vẫn giữ nguyên, ngoại trừ, trong theo lý thuyết, chúng tôi lấy phần bù của các giá trị alpha để tạo ra con cái-2, trong khi đó, trong chương trình, chúng tôi hoán đổi cha mẹ trong khi nhân với alpha, giống như lấy phần bù của các giá trị alpha.

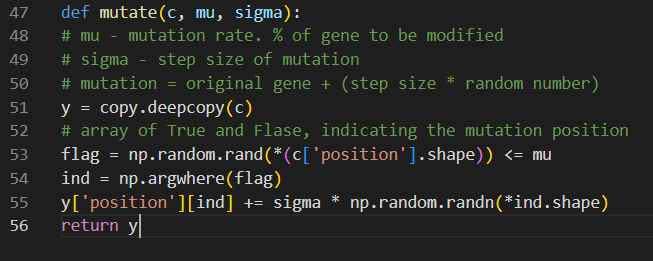


4. Đột biến

Đột biến là một quá trình tự nhiên xảy ra do lỗi sao chép hoặc sao chép gen. Trong khi thực hiện trao đổi chéo, chúng tôi đã sao chép các nhiễm sắc thể của bố mẹ bằng cách kết hợp các gen của cả bố và mẹ. Không có gì đảm bảo rằng việc sao chép các gen bố mẹ chính xác 100%. Luôn xảy ra lỗi, dẫn đến phạm vi của sự khám phá. Ví dụ, nếu cả bố và mẹ của bạn đều có mắt nâu và mắt xanh, đó có thể là do một đột biến xảy ra do lỗi trong khi sao chép gen của cha mẹ bạn và thế hệ tiếp theo của bạn có thể tiếp tục điều đó đặc trưng.

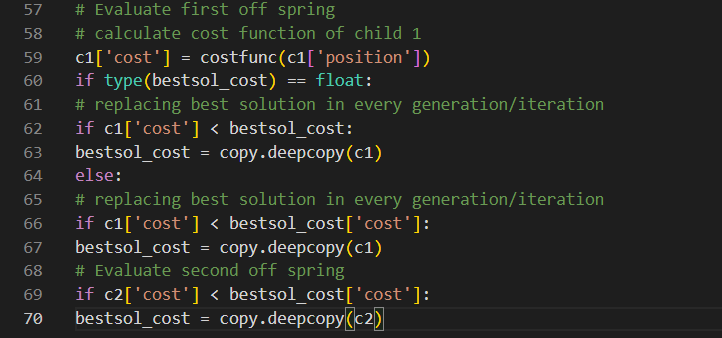
Việc gây đột biến nhiễm sắc thể trong giải thuật di truyền là cần thiết vì nó có thể dẫn đến trong các kết quả mang tính cách mạng sẽ giúp giải quyết vấn đề của chúng ta hiệu quả hơn. Vì vậy chúng tôi có ba tham số: nhiễm sắc thể con (c), tốc độ đột biến (mu) và bước kích thước (sigma). Tỷ lệ đột biến (mu) xác định tỷ lệ phần trăm của đứa trẻ nhiễm sắc thể trải qua đột biến.

Để xác định gen nào sẽ bị đột biến, chúng tôi tạo các số ngẫu nhiên và so sánh chúng đến tỷ lệ đột biến thì chúng ta tìm thấy các chỉ số của nhiễm sắc thể con (vị trí) có giá trị nhỏ hơn tỷ lệ đột biến bằng cách sử dụng hàm argwhere(). Thay thế những chỉ số có gen mới (đột biến) được tạo bằng cách nhân kích thước bước (sigma) với giá trị được tạo ngẫu nhiên và thêm nó vào gen ban đầu.



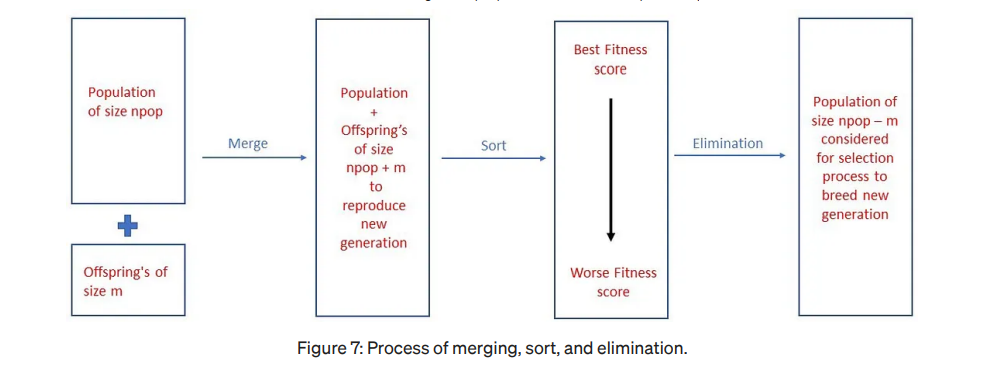
5. Đánh Giá Con Cái

Khi con cái trải qua đột biến, chúng ta cần đánh giá chúng với chi phí chức năng để xác định thể lực của họ. Ngoài ra, hãy thay thế giải pháp tốt nhất trong mọi thế hệ/lặp lại.



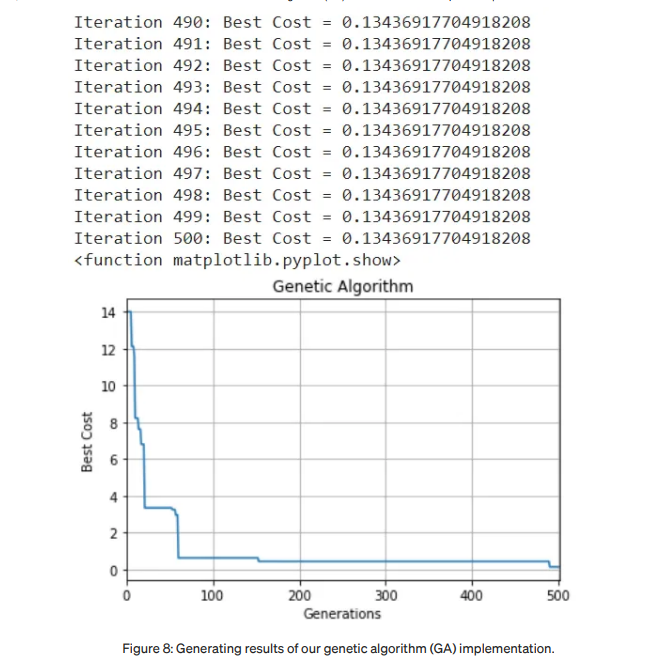
6. Hợp nhất Con cái với Dân số Chính và Sắp xếp

Việc hợp nhất các con lai là rất quan trọng để chúng được coi là bố mẹ để sinh sản thế hệ kế tiếp. Sau khi sắp xếp quần thể mới, chúng ta có những cá thể tốt hơn ở đứng đầu. Vì kích thước dân số vẫn giữ nguyên như lần lặp đầu tiên (npop), nên số cá thể ở đáy của quần thể đã sắp xếp bằng số cá thể các kết quả mới được tạo ra trong lần lặp trước đó bị loại khỏi lựa chọn quá trình để tạo ra những thế hệ con cháu mới, và quá trình này tiếp tục - đây là cách quá trình quá trình đào thải diễn ra.



Kết quả

Số lần lặp để chạy phụ thuộc vào bản chất của vấn đề. trong này hướng dẫn, chúng tôi chạy 500 lần lặp lại.



Chúng ta có thể thấy chi phí giảm như thế nào sau mỗi lần lặp lại và ở mức xấp xỉ 490 lặp lại, chi phí giảm xuống 0,134 và giữ nguyên trong suốt phần còn lại của 10 lần lặp. Do đó, cho chúng tôi giải pháp tối ưu của chúng tôi. TUYÊN BỐ MIỄN TRỪ TRÁCH NHIỆM: Các quan điểm thể hiện trong bài viết này là của (các) tác giả và không đại diện cho quan điểm của Đại học Carnegie Mellon cũng như các công ty khác (trực tiếp hoặc gián tiếp) liên quan đến (các) tác giả. Những bài viết này không có ý định là cuối cùng sản phẩm, nhưng đúng hơn là sự phản ánh của suy nghĩ hiện tại, cùng với việc là chất xúc tác cho thảo luận và cải tiến.

Tất cả các hình ảnh là từ (các) tác giả trừ khi có quy định khác.

Xuất bản qua Hướng tới AI

Tài nguyên

Hướng dẫn đồng hành

Kho lưu trữ Github.

Triển khai Google Colab.

Học thêm

Yarpiz trên Youtube

Người giới thiệu

[1] “Tiến hóa”. 2021. En.Wikipedia.Org. https://en.wikipedia.org/wiki/Evolution.

[2] “Thuật toán di truyền”. 2011. En.Wikipedia.Org.

https://en.wikipedia.org/wiki/Genetic\_algorithm.

[3] Hướng dẫn thuật toán di truyền, Darrel Whitley, 2021. Cobweb.Cs.Uga.Edu.

http://cobweb.cs.uga.edu/~potter/CompIntell/ga\_tutorial.pdf.

Theo

Đổi mới Trí tuệ nhân tạo Khoa học Công nghệ Giáo dục