



Πολυτεχνική Σχολή
Τμήμα Μηχανικών Η/Υ & Πληροφορικής

Computational Intelligence, Project 2nd:

Genetic Algorithms

Εκπονήθηκε από τον:

Καπογιάννη Κωσταντίνο Α.Μ. 1072521

GitHub Link:

<https://github.com/kwnkapog/Genetic-Algorithms-Project>

Πάτρα 2024

Εισαγωγή: Οι Γενετικοί Αλγόριθμοι είναι αλγόριθμοι Μηχανικής Μάθησης και χρησιμοποιούνται για την αναζήτηση βέλτιστων λύσεων, λαμβάνοντας υπόψη όλους τους περιορισμούς του προβλήματος. Είναι εμπνευσμένοι από την θεωρία του Κάρολου Δαρβίνου, περί της Εξέλιξης των Ειδών, κατά την οποία τα πιο καλά προσαρμοσμένα άτομα ενός πληθυσμού αναπαράγονται οδηγώντας στην εξέλιξη του συγκεκριμένου πληθυσμού, μέσω των διαδοχικών γενεών.

Τα βασικά τμήματα του βασικού Γενετικού Αλγορίθμου είναι 4 στο πλήθος και αναλύονται παρακάτω:

1. **Γενετικοί Τελεστές:** Χρησιμοποιούνται προκειμένου να αλλάξουν την σύνθεση του πληθυσμού της επόμενης γενιάς. Οι τελεστές αυτοί συναντώνται και στη φύση, όπως η επιβίωση, η διασταύρωση και η μετάλλαξη.
2. **Χρωμόσωμα/Άτομο:** Εμπνευσμένο από την Βιολογία, το χρωμόσωμα στους Γενετικούς Αλγόριθμο, είναι μία ακολουθία γονιδίων και αντιπροσωπεύει μία από τις υποψήφιες λύσεις της συνάρτησης την οποία βελτιστοποιούμε.
3. **Πληθυσμός:** Μία ακολουθία χρωμοσωμάτων τα οποία αλληλοεπιδρούν μεταξύ τους και παράγουν απογόνους η μεταλλάσσονται.
4. **Συνάρτηση Καταλληλότητας:** Η συνάρτηση η οποία ορίζει πόσο κατάλληλο είναι το χρωμόσωμα, προκειμένου να έχει μεγαλύτερη πιθανότητα να αναπαραχθεί.

Χονδρικά, ένας γενετικός αλγόριθμος αποτελείται από 5 κύρια βήματα, τα οποία φαίνονται και αναλύονται παρακάτω:



1. **Αρχικοποίηση:** Κατά τη φάση της αρχικοποίησης παράγεται ένα σύνολο ατόμων, το οποίο ονομάζεται αρχικός πληθυσμός. Όπως ειπώθηκε και παραπάνω, το κάθε άτομο του πληθυσμού αποτελείται από μία σειρά παραμέτρων, οι οποίες καλούνται γονίδια.
2. **Ανάθεση Καταλληλότητας:** Ουσιαστικά η τιμή της καταλληλότητας του κάθε ατόμου, καθορίζει την ικανότητα του να ανταγωνιστεί άλλα άτομα του πληθυσμού. Η τιμή αυτή καθορίζει πόσο πιθανό είναι το συγκεκριμένο άτομο να επιλεγεί κατά την διαδικασία της αναπαραγωγής.
3. **Επιλογή:** Σε αυτή τη φάση τα άτομα του πληθυσμού επιλέγονται για να παράγουν απογόνους, κατά ζευγάρια. Υπάρχουν τρεις κύριες μέθοδοι για την επιλογή ατόμων. Αυτές είναι:
 - a. Rank based Selection.
 - b. Tournament Selection.
 - c. Roulette Wheel Selection.
4. **Αναπαραγωγή:** Κατά την φάση αυτή τα άτομα που έχουν επιλεγεί, αναπαράγονται ανά δύο. Η φάση της αναπαραγωγής, περιλαμβάνει δύο γενετικούς τελεστές, οι οποίοι είναι:
 - a. **Διασταύρωση:** Σε αυτή τη φάση, επιλέγεται τυχαία ένα σημείο μεταξύ των γονιδίων. Μετέπειτα τα γονίδια μεταξύ των δύο ανταλλάσσονται στο σημείο που καθορίστηκε. Υπάρχουν διάφορα είδη διασταύρωσης, όπως η διασταύρωση μονού ή διπλού σημείου. Μετά την ολοκλήρωση αυτού του τελεστή, τα παιδιά που προέκυψαν, έχουν κληρονομήσει γονίδια και από τους δύο γονείς.
 - b. **Μετάλλαξη:** Σε αυτόν τον τελεστή, τυχαία γονίδια από κάποια άτομα του πληθυσμού αλλάζουν, ώστε να διατηρηθεί η ποικιλομορφία του πληθυσμού. Ο συγκεκριμένος τελεστής συμβάλλει ώστε να αποφευχθεί η πρόωρη σύγκλιση του αλγορίθμου σε υποβέλτιστες λύσεις.
5. **Τερματισμός:** Μετά από κάποιες επαναλήψεις ο αλγόριθμος σταματάει έχοντας επιλεγεί ένα κατάλληλο κριτήριο τερματισμού, το οποίο συνήθως είναι η επίτευξη μίας τιμής-στόχου της συνάρτησης καταλληλότητας ή η μη ύπαρξη σημαντικών αλλαγών στο πληθυσμό από γενιά σε γενιά. Έτσι, ο αλγόριθμος εγγυάται ότι θα βρεθεί η βέλτιστη λύση για το συγκεκριμένο πρόβλημα εντός του τελικού πληθυσμού.

Οι Γενετικοί Αλγόριθμοι προσφέρουν πληθώρα πλεονεκτημάτων, μερικά από τα οποία είναι τα ακόλουθα:

1. Προσφέρουν λύσεις οι οποίες βελτιώνονται όσο περνάνε οι γενιές.
2. Δεν χρειάζεται πληροφορίες παραγώγου και μπορεί πολύ εύκολα να παραλληλοποιηθεί.
3. Μπορεί να χρησιμοποιηθεί για την επίλυση ενός μεγάλου εύρους προβλημάτων, όπως προβλήματα συνεχών και διακριτών μεταβλητών, προβλήματα πολλαπλών στόχων κ.α.
4. Έχει την ικανότητα να διευρύνει κατά πολύ τον χώρο αναζήτησης για την εύρεση των καλύτερων δυνατών λύσεων.

B1.

Αφού κατανοήθηκε η δομή του Γενετικού Αλγορίθμου, σειρά έχει ο προσδιορισμός των παραμέτρων του για την επίλυση του συγκεκριμένου προβλήματος που ζητείται:

α) Η κωδικοποίηση που επιλέγεται είναι το Bag of Words Μοντέλο με χρήση tf-idf. Η παρούσα τεχνική αναλύθηκε ενδελεχώς στο προηγούμενο project, συνεπώς δεν θα αναλυθεί περαιτέρω. Με την εφαρμογή αυτού του μοντέλου, παράγεται ένα λεξικό με 1678 tokens, δηλαδή με όλες τις διαφορετικές λέξεις που απαρτίζουν τις επιγραφές που βρίσκονται στην ίδια γεωγραφική περιοχή με την ημιτελή επιγραφή-στόχο. Επίσης, παράγεται ένα μητρώο, έτσι ώστε κάθε επιγραφή να απεικονίζεται ως ένα διάνυσμα 1678 θέσεων, όπως αναμένεται από το συγκεκριμένο μοντέλο. Όπως προτείνεται, επιλέγεται τα άτομα του πληθυσμού να είναι ζευγάρια ακέραιων τιμών, που θα είναι οι δείκτες των λέξεων στο λεξικό. Οι λέξεις αυτές θα συμπληρώνουν την ημιτελή επιγραφή και εκείνη θα συγκρίνεται με τρόπο που θα αναλυθεί παρακάτω. Συνεπώς κάθε άτομο του πληθυσμού δύναται να πάρει τιμές από 0-1677.

β) Με την επιλεγμένη κωδικοποίηση, δεν προκύπτουν πλεονάζουσες τιμές γιατί οι επιτρεπόμενες τιμές που είναι δυνατόν να λάβουν τα γονίδια του κάθε ατόμου έχουν συγκεκριμένο εύρος από 0-1677. Ωστόσο, σε περίπτωση που υπάρξει άτομο του οποίου ένα γονίδιο να είναι εκτός του επιτρεπτού εύρους, η συνάρτηση καταλληλότητας μπορεί να σχεδιαστεί με τέτοιο τρόπο, ώστε να αποφευχθεί το συγκεκριμένο άτομο από το να αναπαραχθεί. Έτσι διασφαλίζεται ότι άτομα με εκτός εύρους τιμές δεν είναι πιθανό να αναπαραχθούν και να διαδώσουν τέτοιες τιμές στις επόμενες γενιές. Ωστόσο η συγκεκριμένη βιβλιοθήκη που χρησιμοποιήθηκε για την υλοποίηση, μας δίνει την δυνατότητα να ορίσουμε επακριβώς το σύνολο των επιτρεπτών τιμών που επιτρέπονται να λάβουν τα γονίδια.

γ) Προκειμένου να επιλέξουμε την μέθοδο αρχικοποίησης του πληθυσμού που θα αποδώσει τα καλύτερα αποτελέσματα για το συγκεκριμένο πρόβλημα, πρέπει να κατανοηθεί ποια μέθοδος ταιριάζει σε κάθε πρόβλημα. Γενικά, έχουν προταθεί αρκετοί εναλλακτικοί τρόποι για την αρχικοποίηση του πληθυσμού ενός γενετικού αλγορίθμου. Οι μέθοδοι αρχικοποίησης του πληθυσμού, γενικώς μπορούν να κατηγοριοποιηθούν χρησιμοποιώντας τρία βασικά χαρακτηριστικά τους: *τυχειότητα*, *συνθετικότητα*, *γενίκευση*. Παρακάτω φαίνεται γραφικά η κατηγοριοποίηση τους:

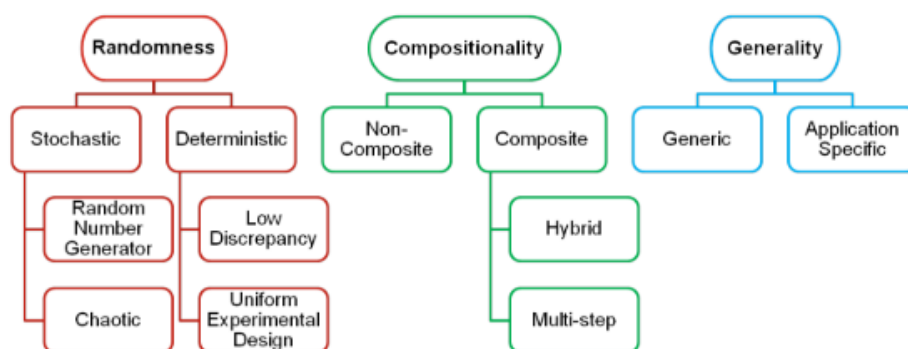


Fig. 1. Three categorizations of population initialization techniques, based on randomness, compositionality and generality.

Οι κυριότερες μέθοδοι που χρησιμοποιούνται για την αρχικοποίηση του πληθυσμού με βάση την πρώτη κατηγορία είναι οι εξής:

1. Στοχαστικές Μέθοδοι:

- a. **Γεννήτρια Τυχαίων Αριθμών (Random Number Generator):** Τα σημεία ή οι αριθμοί, δηλαδή τα άτομα στην δική μας περίπτωση, επιλέγονται τυχαία από ένα μεγαλύτερο σύνολο αριθμών, ισοπίθانا για κάθε σημείο του πληθυσμού. Το πρόβλημα που έγκειται στην συγκεκριμένη κωδικοποίηση είναι ότι χρειάζεται μεγάλο πλήθος δειγμάτων, ώστε να είναι αντιπροσωπευτικό, καθώς τα σημεία που παράγονται μπορεί να είναι κοντά μεταξύ τους ή να ταυτίζονται.
 - b. **Latin Hypercube Sampling:** Σε αυτή τη στοχαστική μέθοδο, προκειμένου να παραχθούν n διαφορετικά σημεία ή χρωμοσώματα, το σύνολο των τιμών, στην περίπτωσή μας το $[0,1677]$, χωρίζεται σε n διαφορετικά υποδιαστήματα, επιλέγεται ένα από αυτά τυχαία και παράγεται μία τιμή από το συγκεκριμένο υποδιάστημα. Έτσι, εξασφαλίζεται ότι θα παραχθούν τυχαίες τιμές που καλύπτουν όλο το διάστημα.
 - c. **Poisson Disk Sampling:** Παράγεται πάλι τυχαία ένα πλήθος τιμών, ωστόσο η συγκεκριμένη μέθοδος απορρίπτει τιμές οι οποίες βρίσκονται αρκετά κοντά σε κάποια ήδη υπάρχουσα τιμή ή ο υπολογισμός της απόστασης τους από τις ήδη υπάρχουσες τιμές είναι υπολογιστικά ακριβός. Η μέθοδος αυτή ωστόσο είναι υπολογιστικά ακριβή και δεν χρησιμοποιείται ευρέως.
 - d. **Chaotic Maps:** Η συγκεκριμένη μέθοδος χαρακτηρίζεται από την ικανότητά της να επισκέπτεται όλες τις τιμές μίας περιοχής (εργοδικότητα), από την τυχαιότητα των τιμών που προκύπτουν και από την κανονικότητά της. Ωστόσο η συγκεκριμένη μέθοδος έχει κάποια προβλήματα, όπως η αστάθεια της και η κακή της κλιμάκωση.
2. **Ντετερμινιστικές Μέθοδοι:**
- a. **Ακολουθίες Άρρητων Κλασμάτων:** Η συγκεκριμένη μέθοδος χρησιμοποιεί έναν άρρητο αριθμό για να παράγει μία ακολουθία δεκαδικών αριθμών. Γνωστά παραδείγματα της μεθόδου αυτής είναι οι ακολουθίες Richtmyer, Kronecker και Ramshaw.
 - b. **Ακολουθίες Πρώτων Αριθμών:** Η συγκεκριμένη μέθοδος παράγει ακολουθίες όπως η VDC η οποία χωρίζει το χώρο σε $b-1$ διαστήματα, όπου κάθε διάστημα χωρίζεται σε επίσης $b-1$ υποδιαστήματα κ.ο.κ. Υπάρχουν και άλλες ακολουθίες για περισσότερες διαστάσεις, όπως οι Halton και Hammersley.

Θεωρείται πως για την απλότητα του συγκεκριμένου προβλήματος και για τον μικρό αριθμό αρχικού πληθυσμού που θα αρχικοποιηθεί, οι δύο πρώτες στοχαστικές μέθοδοι που παρουσιάστηκαν είναι αρκετές για την παραγωγή ενός ικανοποιητικού αποτελέσματος. Έτσι, προτιμάται η μέθοδος *Latin Hypercube Sampling*, λόγω της θεωρητικά καλύτερης προσέγγισής της στο πρόβλημα της αρχικοποίησης ενός αντιπροσωπευτικού αρχικού πληθυσμού.

δ) Υπάρχουν πολλές μέθοδοι για τον έλεγχο ομοιότητας ή απόστασης διανυσμάτων, οι πιο συνηθισμένες εκ των οποίων παρουσιάζονται παρακάτω:

1. **Ευκλείδεια Απόσταση:** Είναι από τις πιο κλασσικές μεθόδους για τον υπολογισμό απόστασης διανυσμάτων σε έναν διανυσματικό χώρο. Η Ευκλείδεια απόσταση δύο διανυσμάτων ορίζεται ως η δεύτερη νόρμα της διαφοράς των δύο διανυσμάτων. Η συγκεκριμένη μέθοδος έχει το χαρακτηριστικό ότι η τιμή της απόστασης αυξάνεται τετραγωνικά όσο πιο μακριά είναι τα διανύσματα. Ωστόσο, παρότι την θεμελιώδη της σημασία, σε διανυσματικούς χώρους μεγάλης διαστατικότητας, όπως η συγκεκριμένη εργασία, τείνει να μην κλιμακώνεται καλά.

2. **Απόσταση Manhattan:** Για την απόσταση Manhattan ισχύουν τα ίδια με την Ευκλείδεια απόσταση, μόνο που η συγκεκριμένη απόσταση ορίζεται ως η πρώτη νόρμα της διαφοράς των διανυσμάτων. Κλιμακώνεται καλύτερα από την ευκλείδεια, αλλά όπως εκείνη χρειάζεται τα διανύσματα να είναι των ίδιων διαστάσεων για να υπολογιστούν.
3. **Απόσταση Hamming:** Η απόσταση Hamming είναι μία μέθοδος που χρησιμοποιείται για δυαδικά δεδομένα ή για δεδομένα που παίρνουν διακριτές τιμές, όπως προτάσεις φυσικής γλώσσας. Ωστόσο, έχοντας χρησιμοποιήσει Tf-idf αναπαράσταση για τις επιγραφές, οι τιμές των διανυσμάτων είναι πραγματικές, συνεπώς η απόσταση Hamming είναι ακατάλληλη μέθοδος για την σύγκρισή τους.
4. **Απόσταση Συνημίτονου:** Η συγκεκριμένη μέθοδος βρίσκει το πόσο μεγάλη ή μικρή είναι η απόσταση μεταξύ δύο διανυσμάτων, μετρώντας την γωνία μεταξύ τους. Τα πλεονεκτήματα της συγκεκριμένης μεθόδου είναι ότι εφαρμόζεται ανεξαρτήτως του μήκους των διανυσμάτων και δουλεύει καλά για αραιά διανύσματα, καθώς υπολογίζει μόνο τα μη μηδενικά στοιχεία των διανυσμάτων. Κυμαίνεται στο διάστημα $[0,1]$, για την περίπτωση μας, με 0 να υποδηλώνει την πλήρη ανεξαρτησία των διανυσμάτων μεταξύ τους (ορθογωνιότητα), ενώ με 1 να δηλώνεται η ταύτιση των διανυσμάτων. Για την συγκεκριμένη εργασία, λόγω της κωδικοποίησης των δεδομένων και της αραιότητας που μια τέτοια κωδικοποίηση συνεπάγεται, είναι η καταλληλότερη μέθοδος για τον υπολογισμό ομοιότητας τους.
5. **Συσχέτιση Pearson:** Η συγκεκριμένη μετρική μας δίνει την τιμή της συσχέτισης δύο διανυσμάτων, είναι δηλαδή χρήσιμη στην ανακάλυψη γραμμικών εξαρτήσεων μεταξύ των διανυσμάτων και άρα των επιγραφών στην συγκεκριμένη περίπτωση. Οι τιμές της συσχέτισης Pearson έχουν εύρος από -1, δηλαδή πλήρη αρνητική συσχέτιση, έως 1, δηλαδή πλήρη θετική συσχέτιση, με τιμή 0 να δείχνει καμία συσχέτιση. Ωστόσο η εφαρμογή της συγκεκριμένης τεχνικής δεν έχει νόημα, δεδομένου ότι οι tf-idf τιμές δείχνουν κάποια συσχέτιση μεταξύ των λέξεων αφού υπολογίζονται λαμβάνοντας υπόψη την εμφάνιση της κάθε λέξης στο σύνολο των επιγραφών.

Είναι προφανές ότι η ομοιότητα συνημίτονου είναι η καταλληλότερη επιλογή για το υπολογισμό ομοιότητας, καθώς τα διανύσματα με τα οποία δουλεύουμε, λόγω κωδικοποίησης, είναι αρκετά αραιά. Επίσης είναι η μόνη από τις μεθόδους κατά την οποία όσο αυξάνεται η ομοιότητα, τόσο αυξάνεται και η μετρική.

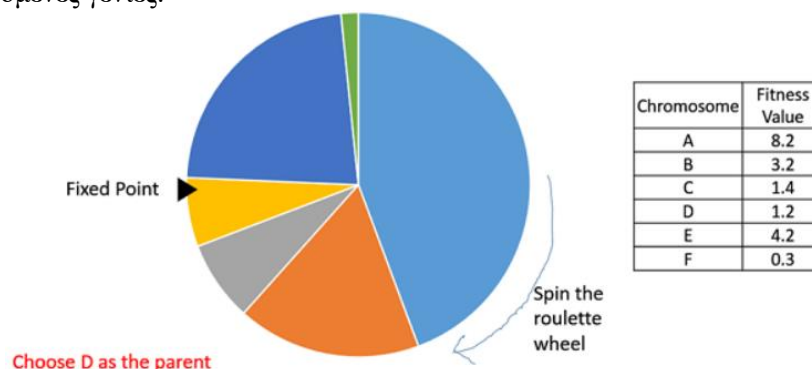
ε) Η συνάρτηση καταλληλότητας είναι βασικό κομμάτι της παραμετροποίησης του Γενετικού Αλγορίθμου, καθώς η κατασκευή της εξαρτάται άμεσα από το πρόβλημα για το οποίο αναζητάται λύση με τη χρήση του. Για το συγκεκριμένο project χρησιμοποιούμε την συνάρτηση `find_k_nearest()` που υλοποιήθηκε, σκοπός της οποίας είναι να βρει τις k κοντινότερες επιγραφές σε μία ζητούμενη επιγραφή. Έπειτα στην συνάρτηση καταλληλότητας, η ζητούμενη ημιτελής επιγραφή συμπληρώνεται με τις λέξεις, με δείκτες τις τιμές των γονιδίων του συγκεκριμένου χρωμοσώματος που αξιολογείται. Η συμπληρωμένη πλέον επιγραφή, συγκρίνεται εκ νέου με τις top-k επιγραφές που βρέθηκαν πριν. Η τιμή καταλληλότητας λοιπόν, του συγκεκριμένου χρωμοσώματος είναι η μέση τιμή των ομοιοτήτων της συμπληρωμένης επιγραφής με τις top-k επιγραφές που βρέθηκαν προηγουμένως. Ο λόγος που οι k κοντινότερες επιγραφές δεν επιλέγονται για την κάθε μία συμπληρωμένη επιγραφή ξεχωριστά, είναι για να υπάρχει μία κοινή συνισταμένη στον υπολογισμό καταλληλότητας των διάφορων χρωμοσωμάτων. Αξίζει να αναφερθεί, ότι στην συνάρτηση καταλληλότητας έχει υλοποιηθεί έλεγχος ώστε αν κάποια συμπληρωμένη επιγραφή έχει τιμή καταλληλότητας μικρότερη από την μέση τιμή των τιμών ομοιότητας των top_k επιγραφών με την μη-συμπληρωμένη, η τιμή καταλληλότητας του συγκεκριμένου χρωμοσώματος μειώνεται στο μισό της αρχικής του τιμής, προκειμένου να αποθαρρυνθεί ο

αλγόριθμος από το να το επιλέξει για αναπαραγωγή. Το συγκεκριμένο σενάριο, σημαίνει ότι η συμπληρωμένη επιγραφή που προκύπτει είναι χειρότερη από εκείνη που θα είχαμε αν την αφήναμε κενή. Θεωρητικά η μέγιστη και η ελάχιστη τιμή της καταλληλότητας είναι ίδιες με εκείνη της συνάρτησης της ομοιότητας συνημίτονου, δηλαδή το διάστημα $[0,1]$. Αυτό συμβαίνει γιατί και τα ίδια τα διανύσματα των λέξεων που προκύπτουν από την κωδικοποίηση, απαρτίζονται από τιμές που προκύπτουν από το διάστημα $[0,1]$. Άρα το χειρότερο σενάριο είναι η επιγραφή να μην ταιριάζει καθόλου με καμία από τις k επιγραφές, ενώ το καλύτερο θεωρητικά σενάριο είναι η συμπληρωμένη συνάρτηση να ταιριάζει απόλυτα και με τις k επιγραφές, το οποίο ωστόσο είναι αρκετά δύσκολο να συμβεί στην πράξη.

στ) Ένα από τα σημαντικότερα βήματα κατά την υλοποίηση ενός Γενετικού Αλγορίθμου είναι ο καθορισμός των γενετικών τελεστών, ανάλογα με το πρόβλημα το οποίο προσπαθούμε να επιλύσουμε.

ι) Ειδικά για τον γενετικό τελεστή της **επιλογής** των γονέων οι οποίοι θα καθορίσουν την επόμενη γενιά, οι κυριότερες επιλογές είναι τρεις:

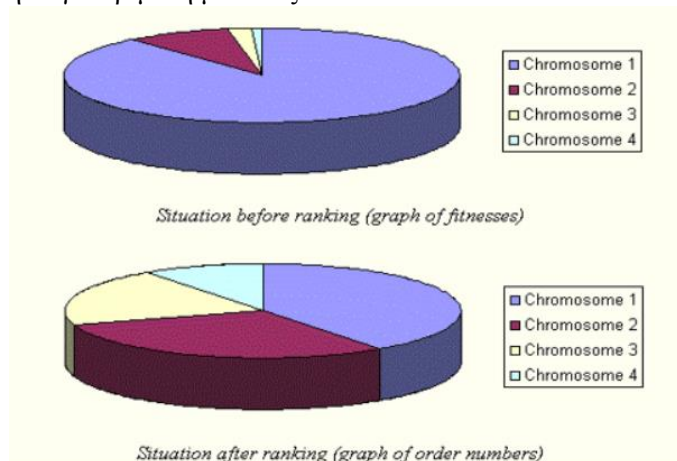
1. **Roulette Wheel Selection (RWS):** Ο συγκεκριμένος τελεστής κατασκευάζει μία ρουλέτα η οποία χωρίζεται σε τόσα σημεία, όσα και τα άτομα του πληθυσμού. Το κάθε άτομο καταλαμβάνει μέρος της ρουλέτας ανάλογο με το πόσο μεγάλη είναι η τιμή της καταλληλότητας του. Έπειτα η ρουλέτα γυρνάει και επιλέγεται το άτομο στο οποίο αντιστοιχεί το σημείο που σταματάει η «βελόνα». Με αυτόν τον τρόπο εξασφαλίζεται ότι τα άτομα που είναι πιο κατάλληλα για επιβίωση, έχουν μεγαλύτερη πιθανότητα να επιλεγούν για αναπαραγωγή και άρα να μεταδώσουν τα γονίδια τους στις επόμενες γενιές.



2. **Rank Selection:** Σε αυτή τη μέθοδο τα άτομα που θα αναπαραχθούν επιλέγονται με βάση ένα βαθμό που τους δίνεται ανάλογα με την τιμή καταλληλότητας που έχουν. Για παράδειγμα στον παρακάτω πίνακα:

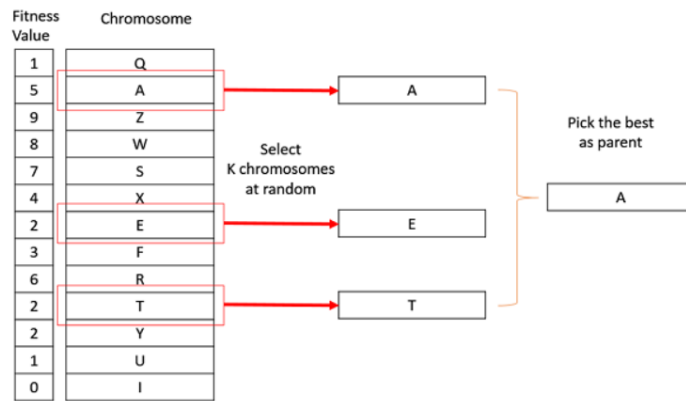
Chromosome	Fitness Value	Rank
A	8.1	1
B	8.0	4
C	8.05	2
D	7.95	6
E	8.02	3
F	7.99	5

Το χρωμόσωμα D έχει την χειρότερη τιμή καταλληλότητας και άρα του ανατίθεται χειρότερος βαθμός από το F που έχει την αμέσως χειρότερη κ.ο.κ. Η συγκεκριμένη μέθοδος δουλεύει καλά στην περίπτωση που υπάρχουν αρνητικές τιμές καταλληλότητας ή στο σενάριο που τα άτομα έχουν πολύ μακρινές τιμές καταλληλότητας. Παρακάτω φαίνεται γραφικά η αλλαγή των καταστάσεων όταν εφαρμόζεται η συγκεκριμένη μέθοδος:



Όπως παρατηρούμε τα χρωμοσώματα «απλώνουν» κάνοντας το πιο δίκαιο να επιλεγούν ιδιαίτερα όταν έχουν πολύ κοντινές τιμές. Ωστόσο κατ' αυτό το τρόπο οδηγούμαστε σε πολύ πιο αργή σύγκλιση του αλγορίθμου, καθώς τα καλύτερα χρωμοσώματα δεν διαφοροποιούνται τόσο από τα υπόλοιπα. Ακόμα, η συγκεκριμένη μέθοδος είναι αρκετά αργή, καθώς πρέπει πρώτα να ταξινομηθούν τα χρωμοσώματα πριν την ανάθεση των βαθμών τους.

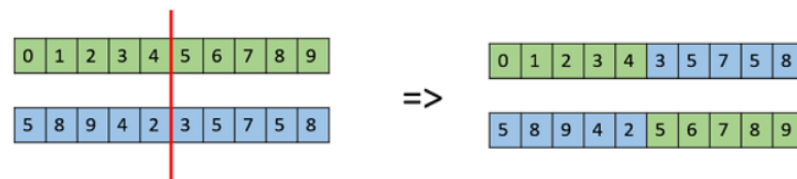
3. **Tournament Selection:** Η συγκεκριμένη μέθοδος επιλέγει τυχαία άτομα από το πληθυσμό και τρέχει τουρνουά μεταξύ τους. Το άτομο με την μεγαλύτερη τιμή καταλληλότητας κερδίζει το τουρνουά. Αναλόγως με το μέγεθος του τουρνουά επηρεάζει άμεσα το ποιες λύσεις επιλέγονται για αναπαραγωγή. Εύκολα καταλαβαίνουμε ότι για μεγάλο μέγεθος τουρνουά, χρωμοσώματα με γενικά μικρές τιμές καταλληλότητας έχουν μικρή πιθανότητα για να επιλεγούν για αναπαραγωγή, καθώς υπάρχει μεγάλη πιθανότητα να υπάρχει ένα χρωμόσωμα με μεγαλύτερη τιμή καταλληλότητας στο ίδιο τουρνουά. Οι δύο κυριότερες υποκατηγορίες του συγκεκριμένου τελεστή είναι τουρνουά με ή χωρίς αντικατάσταση. Η υποκατηγορία χωρίς επανεπιλογή εγγυάται ότι κατά την επιλογή N ατόμων από έναν πληθυσμό N στοιχείων, κάθε άτομο συμμετέχει σε k τουρνουά ακριβώς. Ανάλογα με τον αριθμό των επιλεγμένων στοιχείων, η επιλογή χωρίς αντικατάσταση δεν εγγυάται ότι κανένα άτομο δεν έχει επιλεγεί περισσότερες από μία φορές, απλώς εγγυάται ότι κάθε άτομο έχει ίσες πιθανότητες να συμμετάσχει στον ίδιο αριθμό τουρνουά. Μία γραφική απεικόνιση του συγκεκριμένου τελεστή φαίνεται παρακάτω:



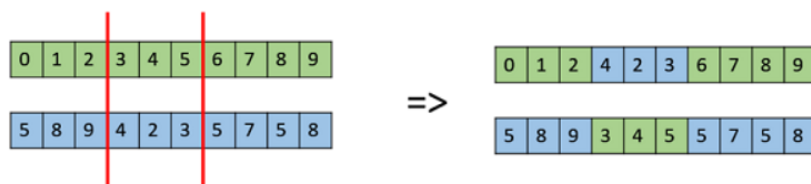
Ιδιαίτερα για την περίπτωση μας, οι γενετικοί τελεστές roulette wheel selection και tournament selection θεωρούνται ιδανικότερες λύσεις. Ωστόσο επειδή ο τελεστής του τουρνουά μπορεί να οδηγήσει σε πρόωρη σύγκλιση όταν το τουρνουά, και άρα και το selective pressure, είναι μεγάλο. Συνεπώς η ανακάλυψη του σωστού μεγέθους του τουρνουά για την παραγωγή των βέλτιστων αποτελεσμάτων προσθέτει περισσότερη πολυπλοκότητα, καθώς χρειάζονται περισσότερες εκτελέσεις του γενετικού αλγορίθμου, για μία μικρή σχετικά απόδοση. Συνεπώς από τους τελεστές επιλογής επιλέγεται η roulette wheel selection.

ii) Ειδικά για τον γενετικό τελεστή της διασταύρωσης, για τον τρόπο δηλαδή που τα επιλεγμένα χρωμοσώματα δημιουργούν νέα για την επόμενη γενιά, οι πιο συνηθισμένες επιλογές είναι επίσης τρεις:

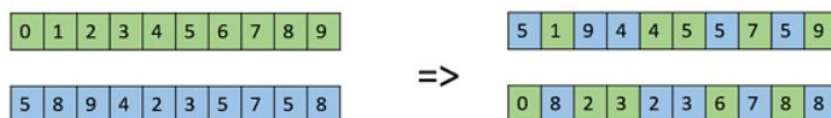
1. **Single Point Crossover:** Η συγκεκριμένη μέθοδος επιλέγει ένα σημείο μετά από το οποίο οι τιμές των γονιδίων των δύο χρωμοσωμάτων ανταλλάσσονται μεταξύ τους:



2. **Multi-Point Crossover:** Είναι ίδια μέθοδος με την παραπάνω, μόνο που επιλέγονται δύο σημεία μεταξύ των γονιδίων και ανταλλάσσονται οι τιμές των γονιδίων ανάμεσα στα δυο αυτά σημεία:



3. **Uniform Crossover:** Σε αυτή τη μέθοδο αντιμετωπίζεται κάθε γονίδιο ξεχωριστά. Για κάθε ένα γονίδιο των χρωμοσωμάτων αποφασίζεται αν θα υπάρξει ή όχι στο παιδί του με πιθανότητα $\frac{1}{2}$ και ανεξάρτητα για το κάθε γονίδιο:



Επειδή ο τρόπος που κωδικοποιήθηκαν τα χρωμοσώματα είναι ένα ζευγάρι ακέραιων τιμών, η διασταύρωση διπλού σημείου δεν είναι δυνατή, ενώ για αυτή την περίπτωση η διασταύρωση μονού σημείου με την ομοιόμορφη διασταύρωση είναι

ισοδύναμες υλοποιήσεις. Άρα επιλέγεται η διασταύρωση μονού σημείου ως τελεστής διασταύρωσης.

iii) Elitism: Η συγκεκριμένη μέθοδος εξασφαλίζει ότι το χρωμόσωμα με την καλύτερη τιμή καταλληλότητας θα περάσει ατόφιο στην επόμενη γενιά. Αυτό γίνεται προκειμένου περιοχές του χώρου αναζήτησης που επιστρέφουν καλά αποτελέσματα, να συνεχίσουν να εξερευνώνται για πιθανές τελικές λύσεις. Μία συνήθης πρακτική για την συγκεκριμένη μέθοδο είναι η δημιουργία ενός συνόλου ελιτισμού αποτελούμενου από το 10% των καλύτερων χρωμοσωμάτων του πληθυσμού. Από αυτό, το 50% περνάει αναλλοίωτο στην επόμενη γενιά, ενώ το χειρότερο 50% επιλέγεται για διασταύρωση με λιγότερο κατάλληλες λύσεις. Έτσι εγγυόμαστε καλύτερες λύσεις και ταχύτατη σύγκλιση του αλγορίθμου. Ωστόσο ακόμα και με απλό ελιτισμό, η σύγκλιση είναι ικανοποιητική.

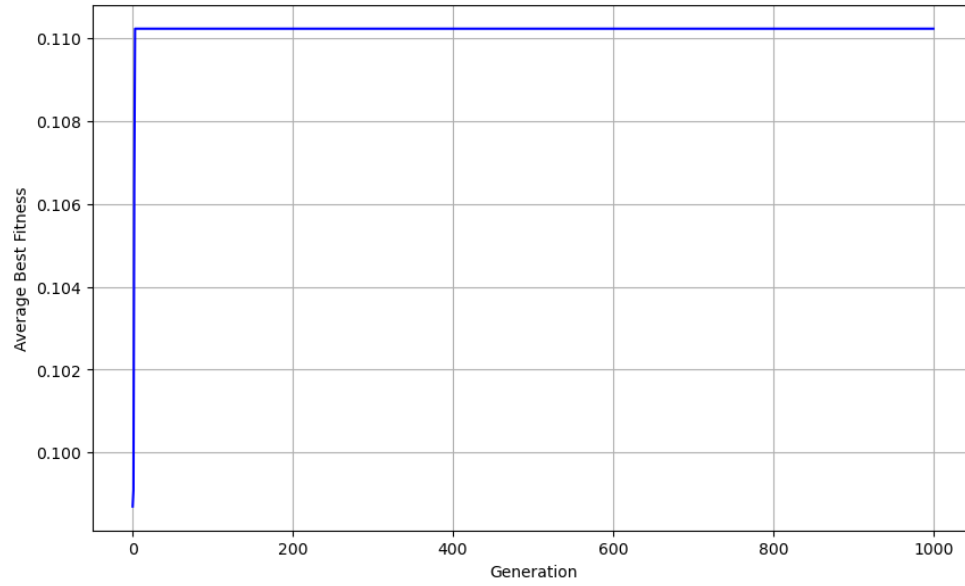
B3.

α) Ο πίνακας για τις διάφορες τιμές των παραμέτρων φαίνεται παρακάτω:

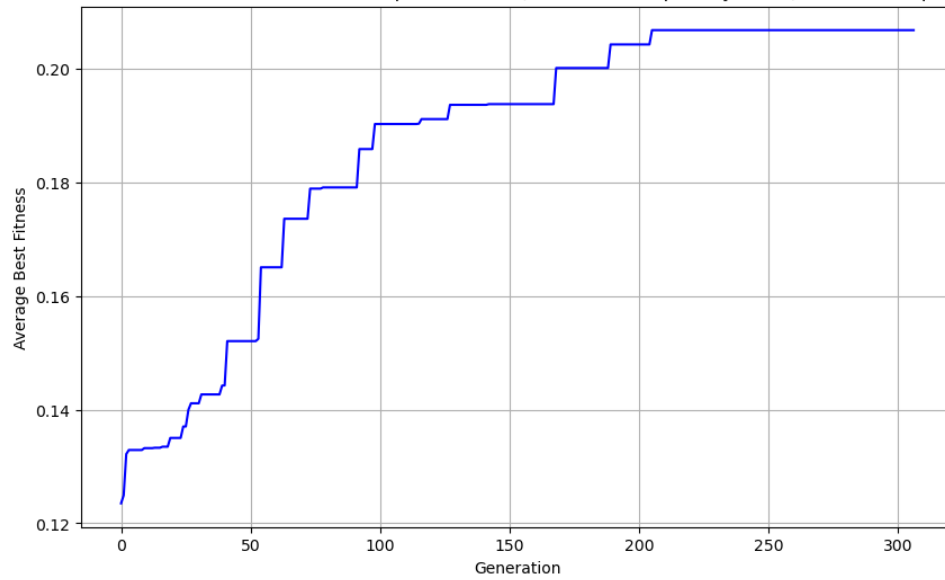
	Μέγεθος Πληθυσμού	Πιθανότητα Διασταύρωσης	Πιθανότητα Μετάλλαξης	Μέση Τιμή Βέλτιστου	Μέσος Αριθμός Γενεών
1	20	0,6	0,00	0.11023	1.3
2	20	0,6	0,01	0.20675	95.1
3	20	0,6	0,10	0.24202	145.2
4	20	0,9	0,01	0.17558	89.5
5	20	0,1	0,01	0.17962	112.5
6	200	0,6	0,00	0.22762	3.5
7	200	0,6	0,01	0.24229	59.1
8	200	0,6	0,10	0.28437	36.9
9	200	0,9	0,01	0.24495	92.4
10	200	0,1	0,01	0.26287	53.9

β) Οι γραφικές παραστάσεις των καμπυλών εξέλιξης για τη μέση απόδοση των καλύτερων ατόμων των πληθυσμών για κάθε τριάδα παραμέτρων, φαίνονται παρακάτω:

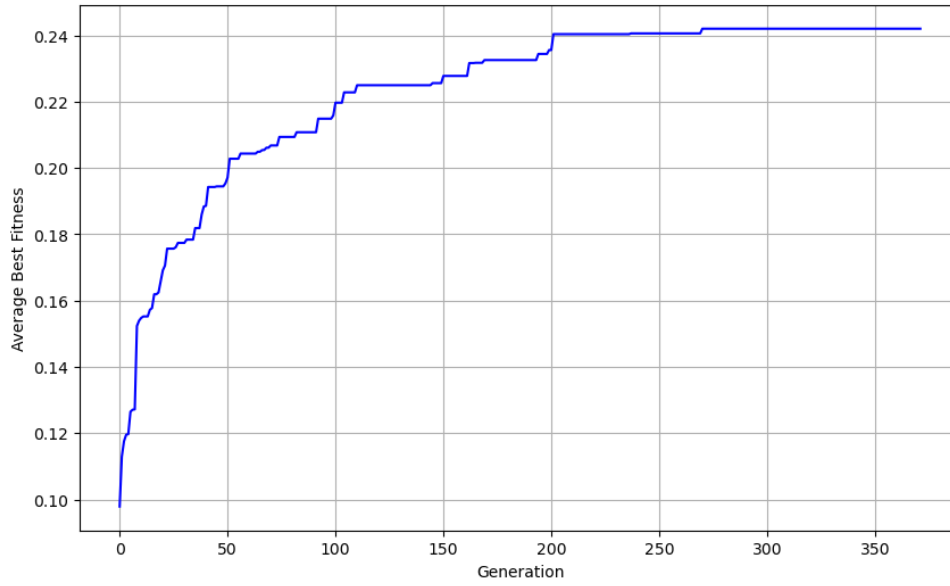
Evolution Curve of Best Solution for Initial Population = 20, Crossover Propability = 0.6, Mutation Propability = 0.0



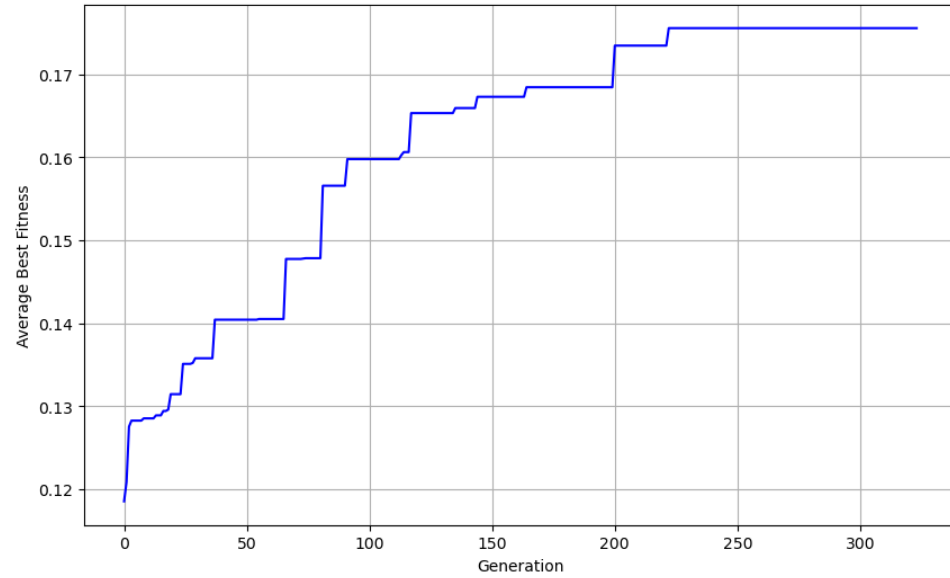
Evolution Curve of Best Solution for Initial Population = 20, Crossover Propability = 0.6, Mutation Propability = 0.01



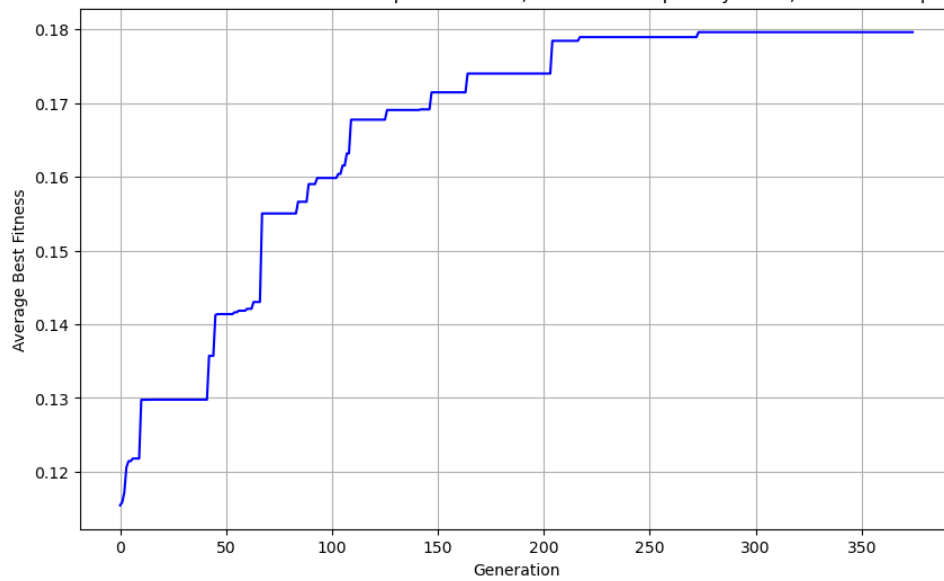
Evolution Curve of Best Solution for Initial Population = 20, Crossover Propability = 0.6, Mutation Propability = 0.1



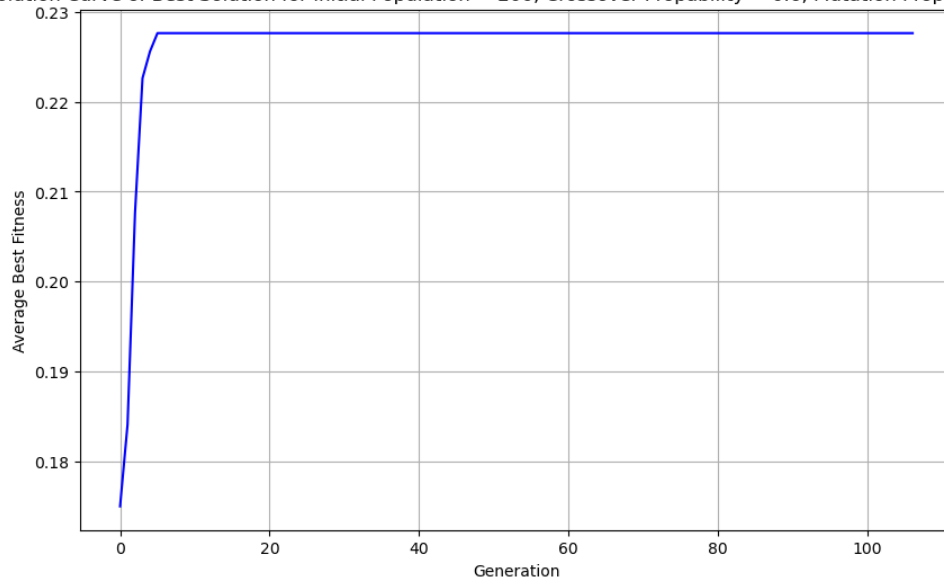
Evolution Curve of Best Solution for Initial Population = 20, Crossover Propability = 0.9, Mutation Propability = 0.01



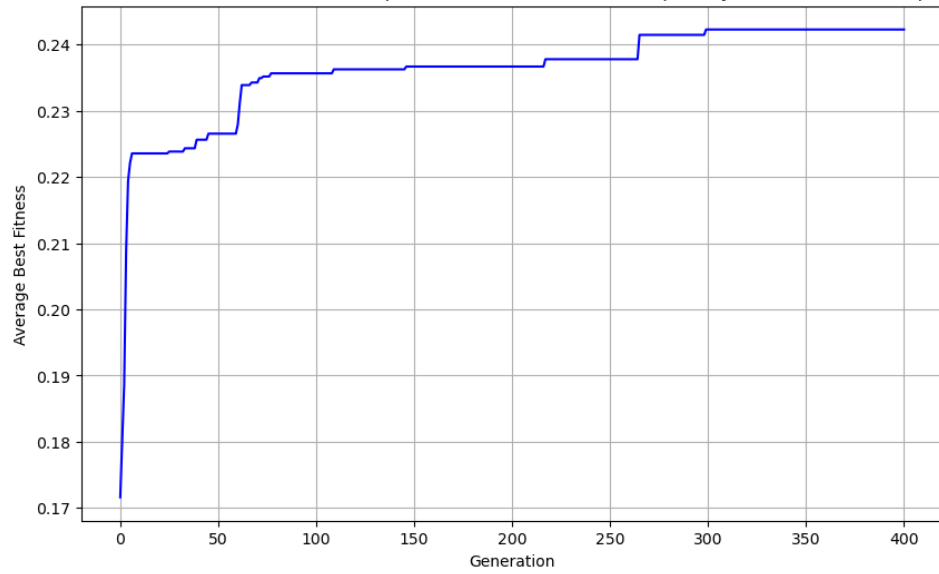
Evolution Curve of Best Solution for Initial Population = 20, Crossover Propability = 0.1, Mutation Propability = 0.01



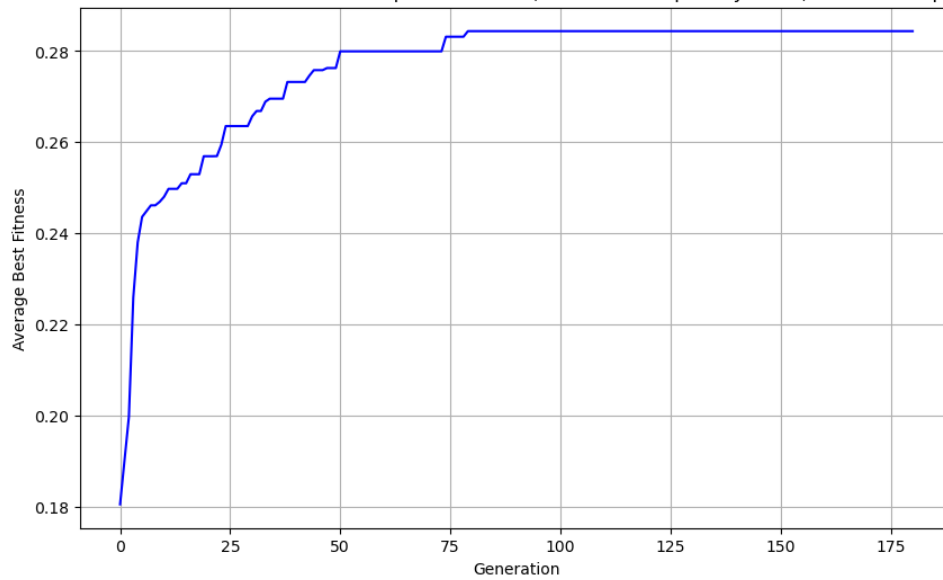
Evolution Curve of Best Solution for Initial Population = 200, Crossover Propability = 0.6, Mutation Propability = 0.0



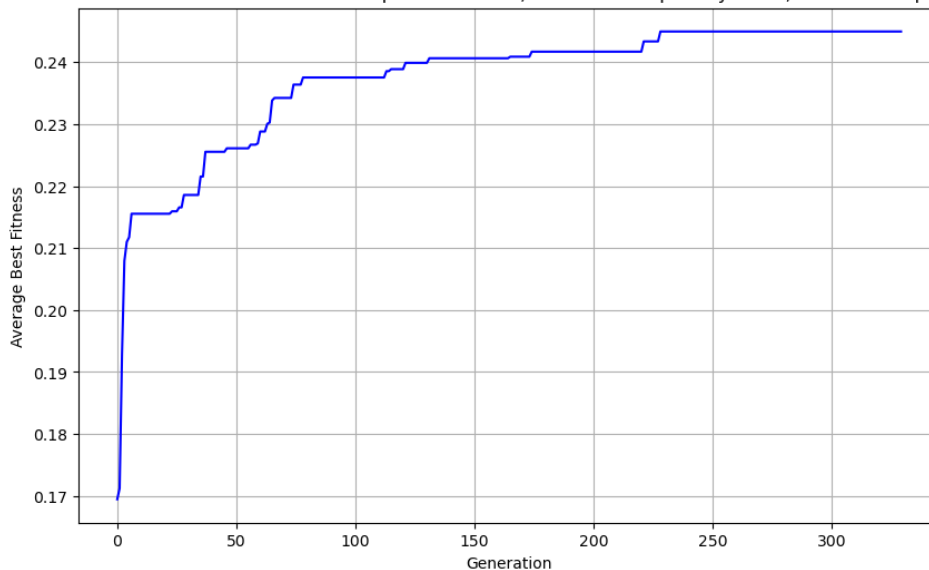
Evolution Curve of Best Solution for Initial Population = 200, Crossover Propability = 0.6, Mutation Propability = 0.01



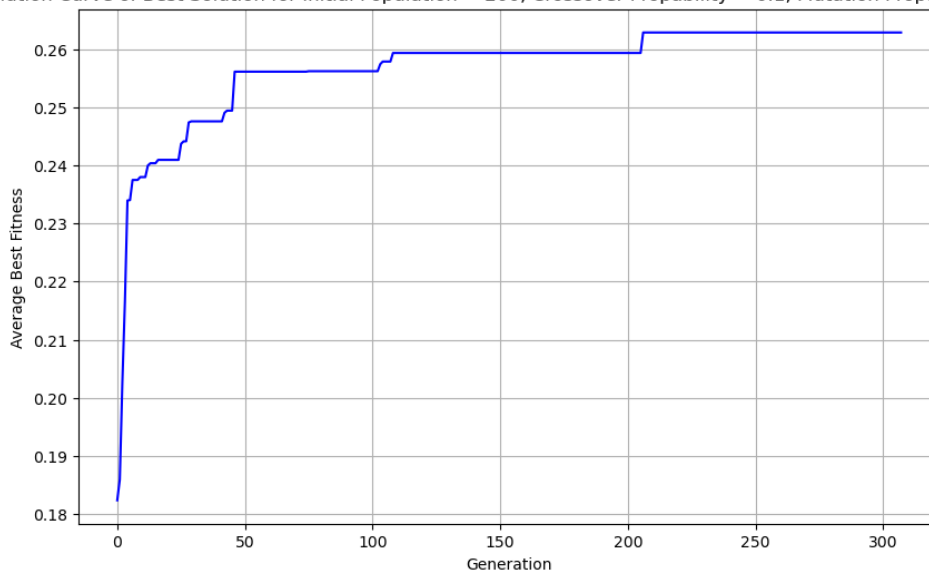
Evolution Curve of Best Solution for Initial Population = 200, Crossover Propability = 0.6, Mutation Propability = 0.1



Evolution Curve of Best Solution for Initial Population = 200, Crossover Propability = 0.9, Mutation Propability = 0.01



Evolution Curve of Best Solution for Initial Population = 200, Crossover Propability = 0.1, Mutation Propability = 0.01



Η καλύτερη λύση προκύπτει για 200 άτομα ανά πληθυσμό, πιθανότητα διασταύρωσης 0.6 και μετάλλαξης 0.1, όπως φαίνεται και από τον πίνακα. Η φράση που προκύπτει για την συγκεκριμένη λύση είναι η :

αθανατος αλεξανδρε ουδισ θαρσι.

γ) Γενικά, κάθε μία από τις παραμέτρους επηρεάζει με διαφορετικό τρόπο τα αποτελέσματα του Γενετικού Αλγορίθμου, δηλαδή την ταχύτητα σύγκλισης και την ποιότητα των λύσεων. Πιο συγκεκριμένα:

1. **Αριθμός Ατόμων:** Γενικά περισσότερα άτομα στο πληθυσμό ισοδυναμεί με περισσότερες υποψήφιες λύσεις και άρα εξερευνάται καλύτερα ο χώρος, αφού μεγαλώνει ο χώρος αναζήτησης, ενώ βοηθάει στο να ξεπεραστούν τοπικά βέλτιστα. Ωστόσο με περισσότερα άτομα στον εκάστοτε πληθυσμό, αυξάνεται κατά πολύ η πολυπλοκότητα, καθώς υπάρχουν περισσότερα χρωμοσώματα που πρέπει να αξιολογηθούν ανά γύρο.

2. **Πιθανότητα Διασταύρωσης:** Όσο μεγαλύτερη είναι η πιθανότητα διασταύρωσης, τόσο πιο διαφορετικές λύσεις παράγονται, οδηγώντας σε καλύτερη εξερεύνηση του χώρου αναζήτησης. Ωστόσο μία μεγάλη πιθανότητα διασταύρωσης μπορεί να καταστρέψει ενδεχομένως καλές λύσεις. Από την άλλη πλευρά, μικρότερες πιθανότητες διασταύρωσης διατηρούν λύσεις που ενδεχομένως να είναι καλές, ωστόσο δεν δίνουν την δυνατότητα για παραγωγή νέων λύσεων, με αποτέλεσμα την βραδύτερη σύγκλιση του αλγορίθμου.
3. **Πιθανότητα Μετάλλαξης:** Η μεταλλάξεις γενικά προσδίδουν ποικιλομορφία στον πληθυσμό, δίνοντας την δυνατότητα να εξερευνηθούν νέες περιοχές του χώρου αναζήτησης. Ωστόσο υψηλές τιμές μετάλλαξης αυξάνουν κατά πολύ την τυχειότητα του αλγορίθμου και είναι βέβαιο ότι θα καταστρέψουν καλές λύσεις, οδηγώντας σε μειωμένη απόδοση. Από την άλλη μειωμένη ή μηδενική πιθανότητα, στερεί από τον αλγόριθμο την ποικιλομορφία που χρειάζεται προκειμένου να εξερευνήσει νέες και ενδεχομένως καλύτερες λύσεις, οδηγώντας σε πρόωρη λύση με υπό-βέλτιστα αποτελέσματα.

Γενικά παρατηρώντας τα γραφήματα, καταλαβαίνουμε ότι προφανώς οι μεγαλύτεροι πληθυσμοί οδηγούν σε καθολικά καλύτερα αποτελέσματα. Προφανώς χρειάζεται μια μικρή αλλά αισθητή πιθανότητα μετάλλαξης για να μπορεί ο αλγόριθμος να παράγει λύσεις πιο μακριά στο χώρο αναζήτησης, ενώ μια μέτρια προς υψηλή πιθανότητα διασταύρωσης είναι καταλληλότερη για την εξερεύνηση καλύτερης λύσης χωρίς να καταστρέφει τις ήδη υπάρχουσες καλές λύσεις. Πράγματι η καλύτερη μέση λύση μας δίνεται για 200 άτομα πληθυσμού, πιθανότητα διασταύρωσης ίση με 0.6 και πιθανότητα μετάλλαξης ίση με 0.1.