***Υπολογιστική Νοημοσύνη ΙΙ***

1η εργασία μαθήματος

Μοντελοποίηση του ανοσοποιητικού συστήματος με χρήση γενετικών αλγορίθμων…

Δημιουργία ενός γενετικού αλγορίθμου και στη συνέχεια πειραματισμός με βάση κάποια αντισώματα που εξελίσσονται ανά γενιές και δρουν συμπληρωματικά εναντίον μιας ομάδας αντιγόνων.

2012

***ΚΑΜΖΟΛΑΣ ΙΩΑΝΝΗΣ 4456***

Τμήμα μηχανικών ηλεκτρονικών υπολογιστών και πληροφορικής

6/4/2012

Η γλώσσα που επέλεξα για την υλοποίηση του γενετικού αλγορίθμου είναι η Matlab. Παρόλο που ο κώδικας είναι αρκετά σχολιασμένος, καλό θα ήταν να διευκρινιστούν και εδώ κάποια θέματα. (Οι συμβάσεις της εκφώνησης της εργασίας έχουν ληφθεί όλες υπόψη)

**ΣΧΕΔΙΑΣΜΟΣ ΚΑΙ ΥΛΟΠΟΙΗΣΗ Γ.Α:**

* Καταρχάς, ο χρήστης δημιουργεί μια συγκεκριμένη ομάδα αντιγόνων μήκους L. Αυτά τα αντιγόνα είναι και η είσοδος του γενετικού αλγορίθμου(της συνάρτησης *g\_algo (antigen, Na, gen, p\_cross, p\_mut ))*, μαζί με το πλήθος τον αντισωμάτων (Na) που θέλουμε να δημιουργηθούν στην αρχική γενιά, τον αριθμό των γενεών (gen) για τον οποίο θέλουμε να λάβει χώρα η εξέλιξη του πληθυσμού μας, την πιθανότητα διασταύρωσης (p\_cross) και την πιθανότητα μετάλλαξης (p\_mut). Το μήκος του κάθε αντισώματος θεωρείται δεδομένο, γι αυτό και δεν ζητείται από την συγκεκριμένη συνάρτηση, αφού είναι όσο και το μήκος των αντιγόνων (L)!
* Κατά τη διαδικασία της απλής αναλογικής επιλογής, μιας και ο γενετικός αλγόριθμος θέλουμε να είναι σταθερού πληθυσμού, επιλέγω τόσα άτομα όσος και ο αρχικός πληθυσμός (δηλαδή Na). Έτσι θα προκύψουν Na /2 ζευγάρια από τα οποία δημιουργούνται τα Na  άτομα της επόμενης γενιάς.
* Κατά την διαδικασία της διασταύρωσης, όταν η διασταύρωση δεν θα πραγματοποιηθεί (δηλαδή ο ψευδοτυχαίος αριθμός είναι μεγαλύτερος της πιθανότητας διασταύρωσης ) τα δύο επιλεγμένα άτομα, θεωρώ πως περνούν στην επόμενη γενιά ως έχουν.

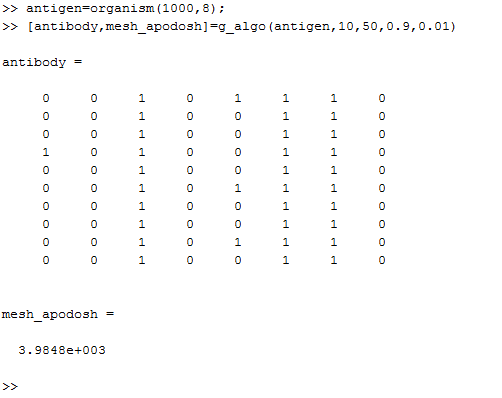
**ΕΚΤΕΛΕΣΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ:**

*Οι γενετικοί αλγόριθμοι από τη στιγμή που είναι στοχαστικοί αλγόριθμοι, δεν εξασφαλίζουν την ίδια απόδοση σε κάθε εκτέλεση τους. Γι αυτό το λόγο, για να έχω μια καλύτερη εικόνα από τα αποτελέσματα που θα πάρω , θα εκτελέσω για κάθε περίπτωση τον αλγόριθμο 10 φορές και θα πάρω τον μέσο όρο τους. Πιο αναλυτικά, θα παρουσιαστεί η 1η περίπτωση, ενώ πιο συνοπτικά οι υπόλοιπες, αφού ακολουθείται η ίδια ακριβώς διαδικασία!*

Καλώ τη συνάρτηση organism, για να δημιουργήσω 1000 αντιγόνα μήκους 8

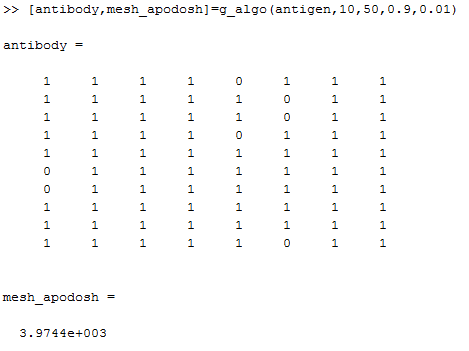
Ο Γ.Α. επιστρέφει τον πληθυσμό τον 10 αντισωμάτων της γενιάς 50, και την μέση απόδοση του πληθυσμού γι αυτή την εκτέλεση.

1η εκτέλεση:

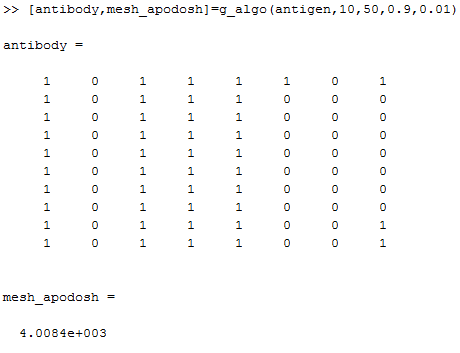


2η εκτέλεση:

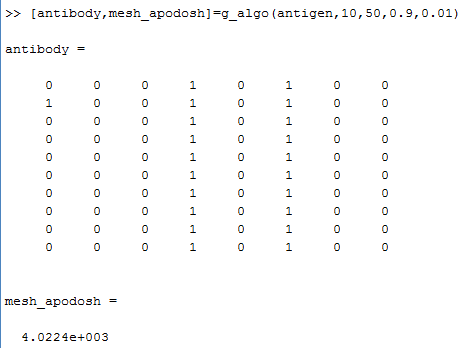
Ο πληθυσμός των αντιγόνων είναι ο ίδιος. Με την κλήση του γενετικού αλγορίθμου, δημιουργείται ο καινούργιος αρχικός πληθυσμός των αντισωμάτων.



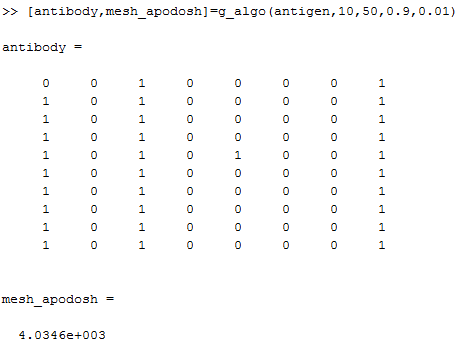
3η εκτέλεση:



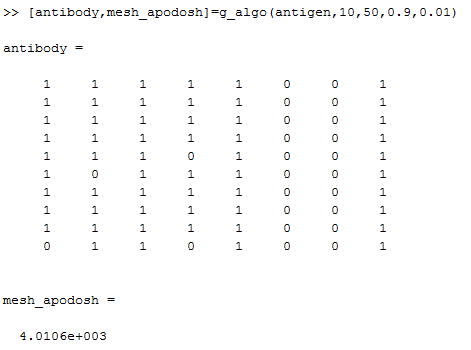
4η εκτέλεση:



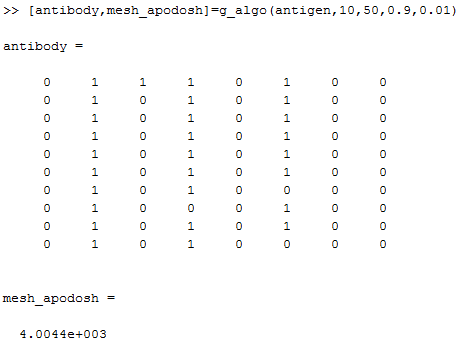
5η εκτέλεση:



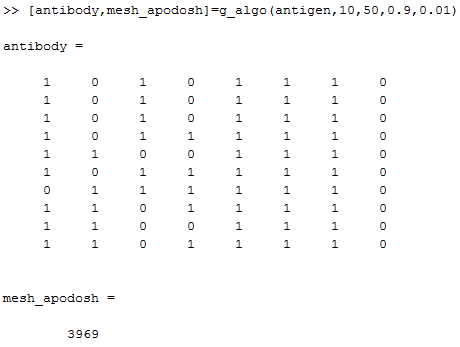
6η εκτέλεση:



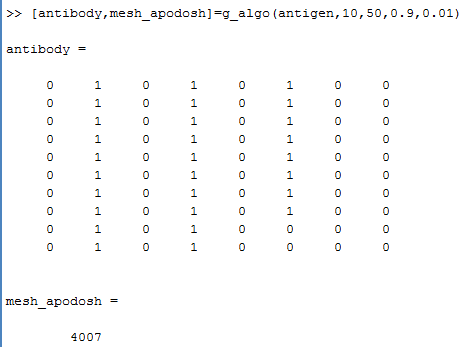
7η εκτέλεση:



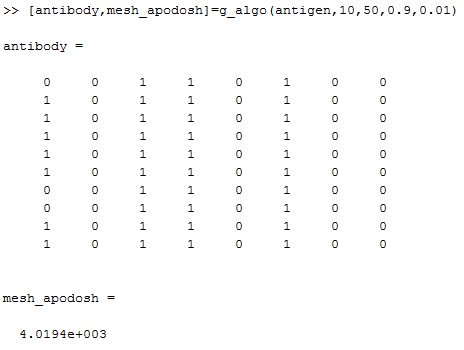
8η εκτέλεση:



9η εκτέλεση:



10η εκτέλεση:



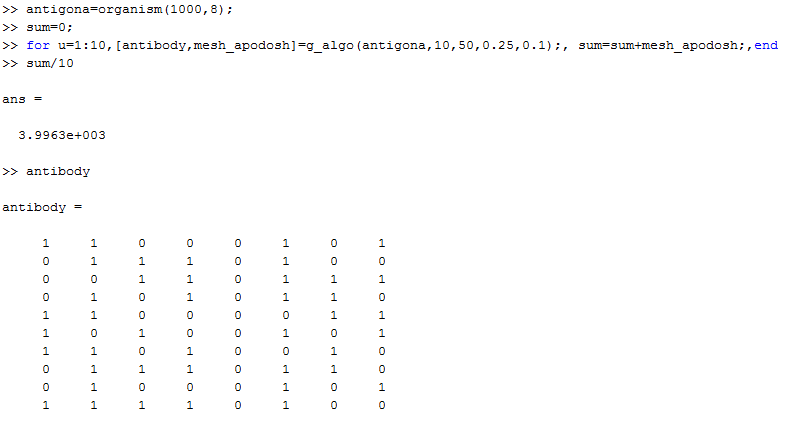
Ο μέσος όρος των 10 εκτελέσεων, δίνει σαν μέση απόδοση **4003.5**

Αυτό είναι λογικό, αφού με 1000 αντιγόνα μήκους 8, η μέγιστη τιμή αξιολόγησης θεωρητικά φτάνει μέχρι 8\*1000=8000

***0<=4003.5<=8000***

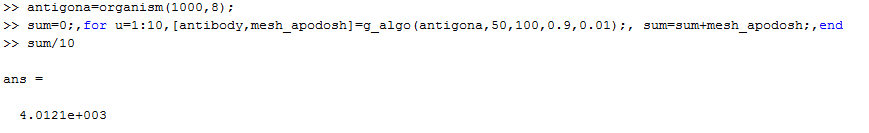
1. Στη δεύτερη περίπτωση , αλλάζει μόνο η πιθανότητα διασταύρωσης και η πιθανότητα μετάλλαξης.

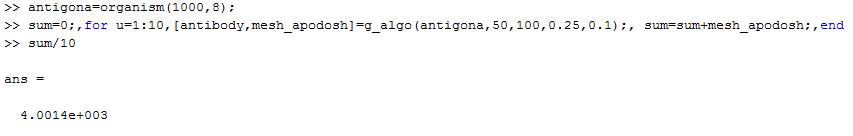
Δεν θα παρουσιάσω πάλι 10 στιγμιότυπα αλλά μόνο ένα χρησιμοποιώντας τις ακόλουθες εντολές στη Matlab:



Η μέση απόδοση σε αυτή την περίπτωση είναι **3996.3,** δηλαδή και πάλι στα αναμενόμενα αποτελέσματα, και πολύ κοντά σε αυτά της προηγούμενης περίπτωσης.

1. Τώρα αλλάζει ο αριθμός των γενεών σε 100, και ο αριθμός των αντισωμάτων σε 50. Οι πιθανότητες διασταύρωσης και μετάλλαξης είναι όπως και στην πρώτη περίπτωση.

 Η μέση απόδοση είναι τώρα περίπου **4012**. Τη συγκρίνουμε με την πρώτη περίπτωση με την οποία διαφέρουν μόνο στον αριθμό των γενιών της εξέλιξης και στον αριθμό των αντισωμάτων. Βλέπουμε μια μικρή βελτίωση όσον αφορά την απόδοση, αφού διπλασιάσαμε τον αριθμό των γενεών και η αξιολόγηση ήταν κατά 9 μονάδες καλύτερη σε σχέση με την πρώτη περίπτωση και κατά 16 σε σχέση με την δεύτερη!



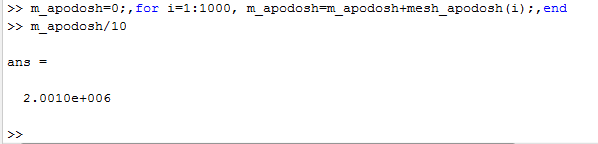
Σε αυτή την περίπτωση, η μέση απόδοση είναι περίπου **4001**. Η πιθανότητα διασταύρωσης που δίνουμε είναι μόλις 0,25 ενώ η πιθανότητα μετάλλαξης είναι 10πλάσια (0.1 αντί για 0.01) . Φαίνεται να είναι λίγο χειρότερα τα αποτελέσματα σε σχέση με πριν. Είναι όμως καλύτερα από αυτά της 2ης περίπτωσης όπου οι γενιές ήταν οι μισές.

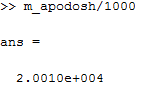
1. Στις δύο τελευταίες περιπτώσεις ακολουθούμε το ίδιο μοτίβο, όπου αρχικά αυξάνουμε τον αριθμό των γενεών σε 1000 και παίρνουμε περιπτώσεις για πιθανότητες διασταύρωσης-μετάλλαξης αντίστοιχα 0.9-0.01 και 0.25-0.1



*Το πρόβλημα είναι ότι για 10000 αντιγόνα που ζητείται, ο αλγόριθμος γινόταν απίστευτα αργός, έτσι τις μετρήσεις τις πήρα για 5000 αντιγόνα, θεωρώντας πως και για 10000 τα αποτελέσματα θα είναι ανάλογα.*

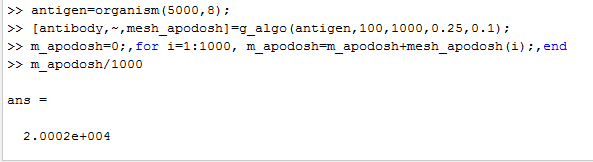






Άρα η μέση απόδοση είναι **20010**. Και αυτό το αποτέλεσμα είναι αναμενόμενο, αφού με 5000 αντιγόνα τα οποία έχουν μήκος 8, η απόδοση θεωρητικά κυμαίνεται από 0 μέχρι 5000\*8=40000 (Το 20010 είναι και πάλι περίπου στη μέση!)

1. Όμοια με πριν, προκύπτει ότι η μέση απόδοση είναι **20002**



Συνοπτικά τα αποτελέσματα παρουσιάζονται στον παρακάτω πίνακα:

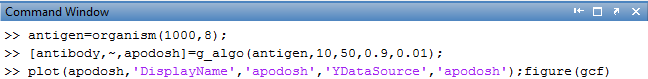
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Α/α** | **αριθμοσ γενιων** | **πιθανοτητα διαστ/σης** | **πιθανοτητα ΜΕΤΑ/ΞΗΣ** | **ΑΡΙΘΜΟΣ ΑΝΤΙΓΟΝΩΝ /ΑΝΤΙΣΩΜΑΤΩΝ** | **ΜΕΣΗ ΑΠΟΔΟΣΗ ΠΛΗΘΥΣΜΟΥ** |
| 1 | 50 | 0.9 | 0.01 | 1000 / 10 | **4003** |
| 2 | 50 | 0.25 | 0.1 | 1000 / 10 | **3996** |
| 3 | 100 | 0.9 | 0.01 | 1000 / 50 | **4012** |
| 4 | 100 | 0.25 | 0.1 | 1000 / 50 | **4001** |
| 5 | 1000 | 0.9 | 0.01 | 5000 / 100 | **20010** |
| 6 | 1000 | 0.25 | 0.1 | 5000 / 100 | **20002** |

Συνολικά παρατηρούμε ότι για τον ίδιο αριθμό γενεών, όταν η πιθανότητα διασταύρωσης είναι μεγάλη και η πιθανότητα μετάλλαξης μικρή, η απόδοση φαίνεται να είναι καλύτερη. Όταν οι πιθανότητες αυτές είναι οι ίδιες, για περισσότερες γενεές, η απόδοση φαίνεται να βελτιώνεται ελάχιστα.

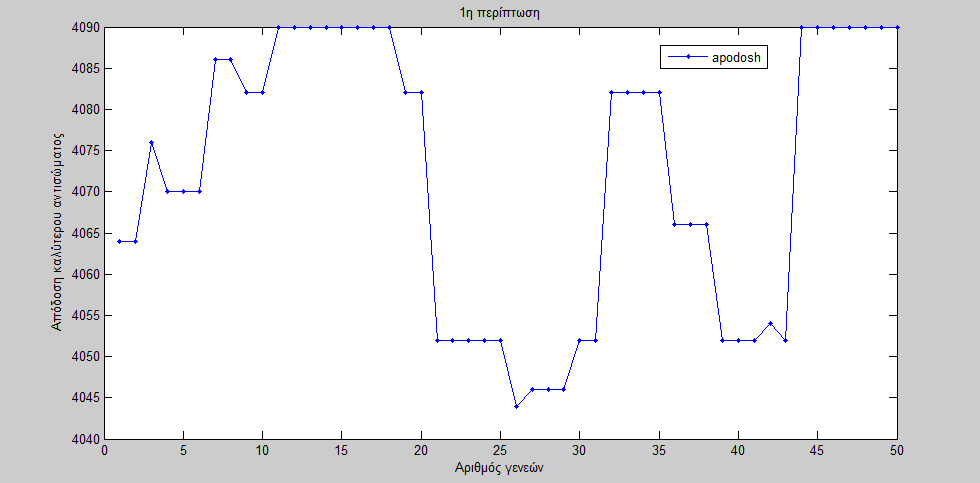
**(BONUS) ΑΞΙΟΛΟΓΗΣΗ ΓΕΝΕΤΙΚΟΥ ΜΟΝΤΕΛΟΥ:**

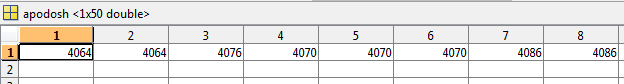
Για κάθε μια από τις έξι παραπάνω περιπτώσεις, θα κάνω την καμπύλη εξέλιξης του καλύτερου αντισώματος (απόδοση /αριθμό γενεών)

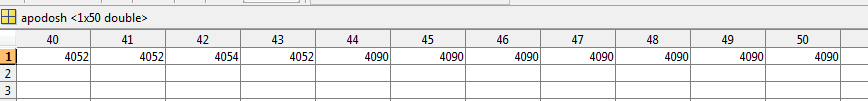
Για την απάντηση αυτού του ερωτήματος, τροποποιώ λίγο τον αρχικό μου κώδικα και δημιουργώ το μητρώο *apodosh* που έχει τόσα στοιχεία όσες και οι γενιές την προκειμένη περίπτωση. Στην θέση **i** του μητρώου, κρατάμε την απόδοση του καλύτερου αντισώματος για την γενιά **i** . Για να το καταφέρω αυτό, χρησιμοποιώ τη μεταβλητή max, η οποία θα έχει τελικά την τιμή αξιολόγησης του καλύτερου αντισώματος για την εκάστοτε γενιά.

1. Εκτελώ τις εντολές

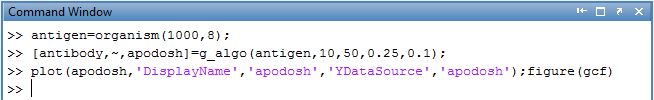
Η καμπύλη εξέλιξης είναι η ακόλουθη



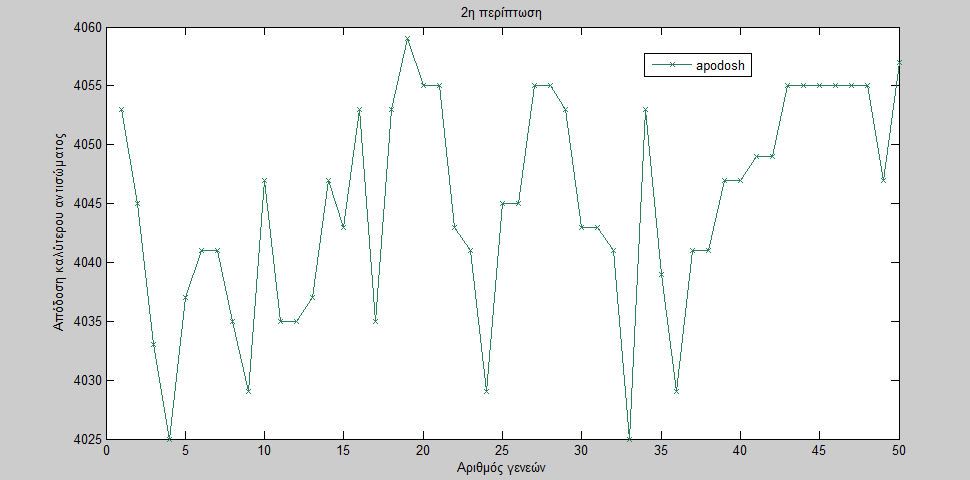


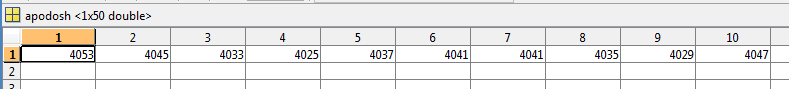


Πέρα από το διάγραμμα παραθέτω και τον πίνακα για τις πρώτες 8 γενιές και τις 10 τελευταίες. Βλέπουμε ότι η απόδοση του πληθυσμού αρχικά αυξάνεται μέχρι περίπου την γενιά 11, όπου παίρνει και τη μέγιστη τιμή και στη συνέχεια πέφτει λίγο η απόδοση, έως ότου τελικά κοντά στη γενιά 50 (συγκεκριμένα στη γενιά 44) λάβει και πάλι την τιμή 4090, η οποία είναι και η μέγιστη απόδοση.

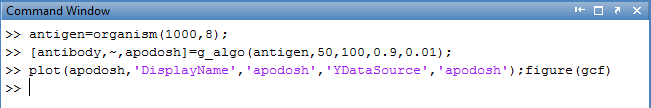
1. Εκτελώ τις εντολές

Τα δεδομένα είναι τα ίδια με πριν, μόνο που τώρα η πιθανότητα διασταύρωσης είναι πολύ μικρή (0.25) και η πιθανότητα μετάλλαξης αρκετά μεγαλύτερη (10πλάσια)

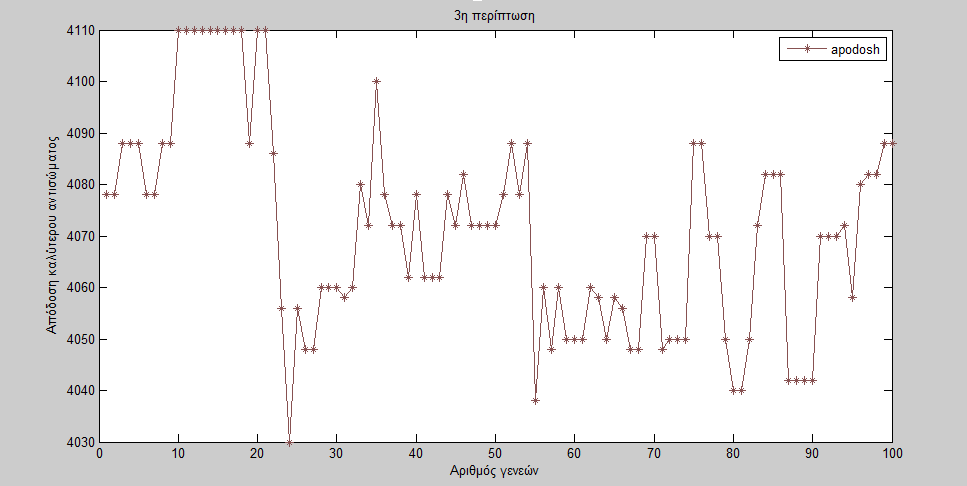


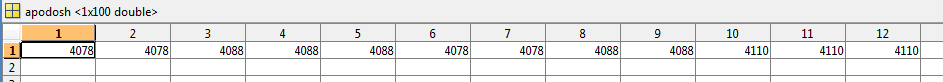


Οι τιμές αυτές των πιθανοτήτων είναι και ο λόγος που παρατηρούμε τόσο απότομες διακυμάνσεις (αυξομειώσεις) στην απόδοση ανά γενιά (σε σχέση τουλάχιστον με την προηγούμενη περίπτωση που η καμπύλη ήταν πιο ομοιόμορφη).

1. Εκτελώ τις εντολές

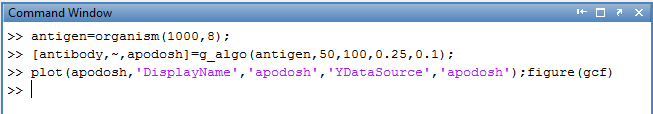
Η καμπύλη που παίρνω είναι η ακόλουθη:

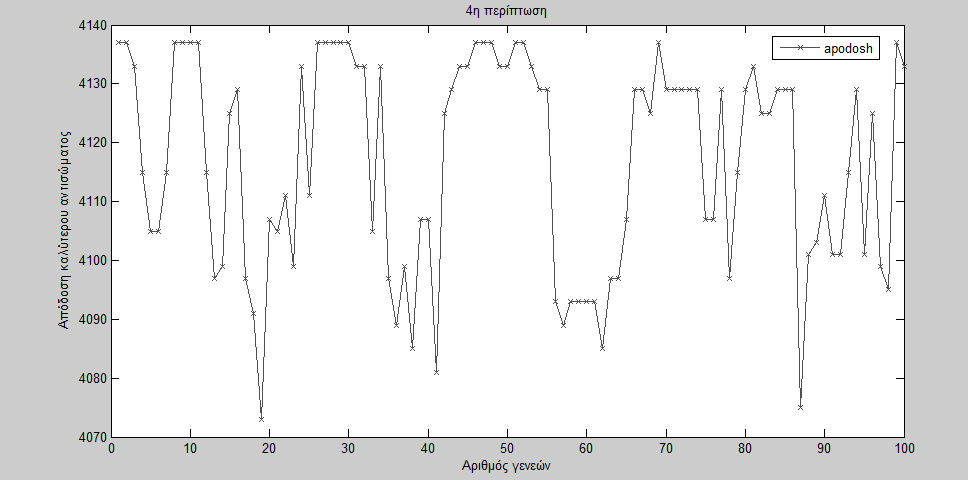




Σε αυτή την περίπτωση, αυξήσαμε τον αριθμό των γενιών σε 100. Όπως φαίνεται και από το πινακάκι, στην 10η γενιά η τιμή του καλύτερου αντισώματος είναι 4110. Αυτή η τιμή είναι η καλύτερη, που όμως στη συνέχεια μειώνεται, έως ότου καταλήξει στην γενιά 100 στην τιμή 4088.

1. Στην 4η περίπτωση (σε σχέση με την τρίτη) αλλάζουμε και πάλι τις πιθανότητες διασταύρωσης (σε 0.25, δηλαδή τη μειώνουμε αρκετά) και μετάλλαξης (σε 0.1, δηλαδή την δεκαπλασιάζουμε)

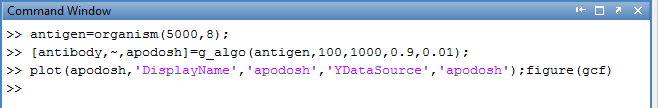
Εκτελώ τις εντολές που ακολουθούν: 



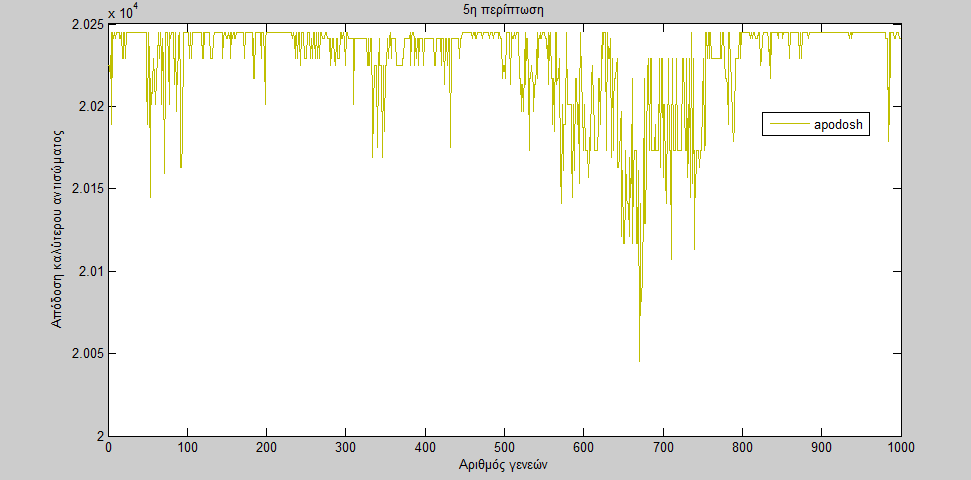
Και εδώ παρατηρούμε τις απότομες αυξομειώσεις στην καλύτερη απόδοση, με το πέρασμα των γενιών. Αυτό συμβαίνει κυρίως γιατί αφού γίνονται σχετικά λίγες διασταυρώσεις (πιθανότητα=0.25), ο πληθυσμός περνάει τις περισσότερες φορές αναλλοίωτος και στη συνέχεια συμβαίνει μετάλλαξη σε κάποια από τα χρωμοσώματα (πιθανότητα=0.1, δηλαδή αρκετά μεγάλη για μετάλλαξη). Έτσι παρατηρείται αυτή η περίεργη και απότομη διακύμανση στις τιμές της απόδοσης (με μέγιστη τιμή απόδοσης την 4137).

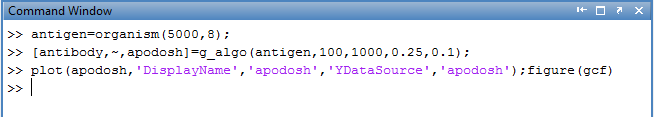
1. Σε αυτή την περίπτωση, τα αντιγόνα είναι 5000, τα αντισώματα 100 και οι γενιές 1000

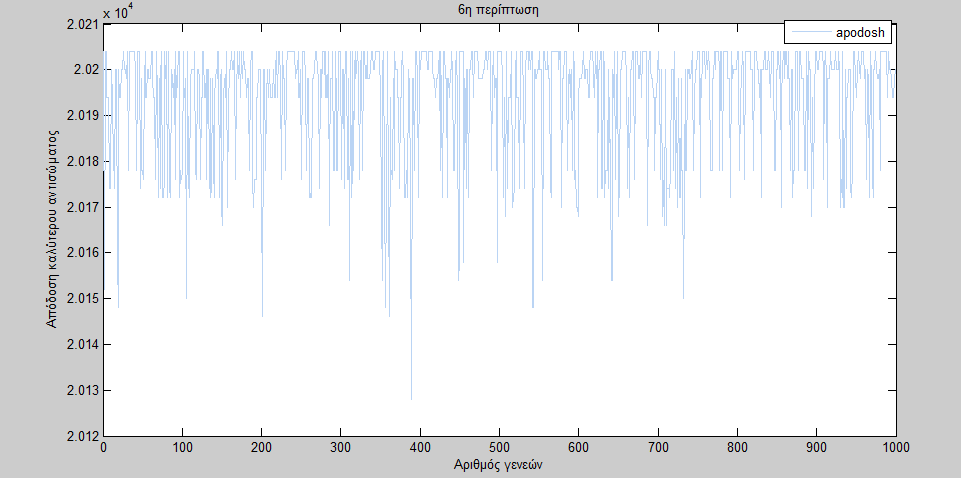
**Παρατήρηση***:(Με 10000 αντιγόνα ο αλγόριθμος γινόταν πολύ αργός και δύσκολος στην μελέτη των αποτελεσμάτων, γι αυτό θεώρησα πως και τα 5000 αντιγόνα ήταν υπεραρκετά για τις παρατηρήσεις μου!)*

Εκτελώ τις εντολές που ακολουθούν: 

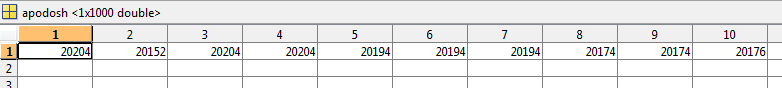
Και παίρνω την καμπύλη



1. Αλλάζουμε και πάλι τις πιθανότητες όπως και στις προηγούμενες περιπτώσεις και για 5000 αντιγόνα και 1000 γενιές έχουμε: 



Για τους ίδιους λόγους με πριν (περιπτώσεις 2 και 4), παρατηρείται αυτή η συνεχής και απότομη αυξομείωση στην τιμή της καλύτερης απόδοσης στην καμπύλη εξέλιξης, η οποία είναι εμφανώς εντονότερη σε σχέση με αυτή της 5ης περίπτωσης. Βέβαια είναι δύσκολο να βγάλουμε λεπτομερή συμπεράσματα για μια συγκεκριμένη γενιά, αφού τα χίλια σημεία είναι πάρα πολύ κοντά το ένα στο άλλο!



Στο παραπάνω πινακάκι φαίνονται οι τιμές του καλύτερου αντισώματος για τις 10 πρώτες γενιές

**Γενική αξιολόγηση του γενετικού μου μοντέλου:**

Το ανοσοποιητικό σύστημα είναι ένα σύστημα οργάνων υπεύθυνο για την άμυνα του οργανισμού. Αποτελείται από πολλά όργανα και αυτά με τη σειρά τους από ακόμα περισσότερα κύτταρα. Εμείς, για το μοντέλο που υλοποίησα, είδαμε πιο πριν την βασική κυτταρική μονάδα του ανοσοποιητικού συστήματος που είναι το **αντίσωμα**.

Γενικά ο γενετικός κώδικας που υπάρχει στους οργανισμούς (γονιδίωμα), μπορεί να προσημειωθεί με την κωδικοποίηση 0/1 που ακολούθησα και εγώ. (Σίγουρα όμως, οτιδήποτε στη φύση είναι αδύνατο να προσομοιωθεί πιστά σε ένα μοντέλο!). Το μοντέλο μου ήταν αρκετά κοντά στο πραγματικό, τουλάχιστον βάση των αποτελεσμάτων. Εφόσον αντιγόνα και αντισώματα δρουν συμπληρωματικά, αν για παράδειγμα έχουμε 100 αντιγόνα μήκους 8, τότε η **μέση απόδοση** θα προκύπτει από τη σύγκριση κάθε χρωμοσώματος του αντισώματος με κάθε χρωμόσωμα του κάθε αντιγόνου και θα είναι:

0<= **μέση απόδοση** <=100\*8

Εμείς, όπως είδαμε στα προηγούμενα, οι τιμές της απόδοσης κυμαίνονταν στο μέσο (ίσως και «λίγο πιο πάνω» τις πιο πολλές φορές) αυτού του διαστήματος. Πάντως, όπως παρατηρήσαμε και στις καμπύλες εξέλιξης, οι συναρτήσεις της απόδοσης ήταν κυρίως αύξουσες που όμως κάποια στιγμή παρουσίαζαν τοπικά βέλτιστα (κάποιο από αυτά θα ήταν μάλλον και το ολικό βέλτιστο!)

Τέλος, η εξέλιξη του πληθυσμού των αντισωμάτων σε κάθε γενιά, είχε ως στόχο την βελτίωση της μέσης απόδοσης. Αυτό πρακτικά στο ανοσοποιητικό μας σύστημα, εκφράζεται με την πρόληψη του αντιγόνου και στην συνέχεια μετά από ειδική επεξεργασία την επανέκφραση τμήματός του στην κυτταρική τους μεμβράνη..