



Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής  
Πρόγραμμα Μεταπτυχιακών Σπουδών  
«Πληροφορική»

**Μεταπτυχιακή Διατριβή**

Τίτλος Διατριβής	<b>Μονοταξικοί ταξινομητές βασισμένοι σε τεχνητά ανοσοποιητικά συστήματα. One class classifiers based on artificial immune systems.</b>
Όνοματεπώνυμο Φοιτητή	<b>Αρώνης Δημήτριος</b>
Πατρώνυμο	<b>Μιχαήλ</b>
Αριθμός Μητρώου	<b>ΜΠΠΛ/16002</b>
Επιβλέπων	<b>Γεώργιος Τσιχριντζής, Καθηγητής</b>

**Τριμελής Εξεταστική Επιτροπή**

(υπογραφή)

(υπογραφή)

(υπογραφή)

Γεώργιος Τσιχριντζής  
Καθηγητής

Σακκόπουλος Ευάγγελος  
Επίκουρος Καθηγητής

Διονύσιος  
Σωτηρόπουλος  
Διδάκτωρ

**ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ**

1	Εισαγωγή .....	5
2	Μονοταξική Ταξινόμηση.....	7
3	Τεχνητά Ανοσοποιητικά Συστήματα .....	8
3.1	Negative Selection .....	9
3.2	Real Valued Negative Selection.....	11
3.3	Ο Αλγόριθμος V-Detector.....	12
4	Δημιουργία Τεχνητών Δεδομένων .....	15
4.1	Η συνάρτηση CreateData.....	21
4.2	Η συνάρτηση CreateKernels.....	22
5	Πειράματα και Αποτελέσματα .....	23
5.1	Πείραμα 1 <sup>ο</sup> .....	25
5.2	Πείραμα 2 <sup>ο</sup> .....	31
6	Συμπεράσματα και Μελλοντική Έρευνα .....	37
6.1	Γράφημα 1 <sup>ο</sup> .....	39
6.2	Γράφημα 2 <sup>ο</sup> .....	41
6.3	Γράφημα 3 <sup>ο</sup> .....	43
7	Βιβλιογραφικές Πηγές .....	47

## **Abstract**

### **Ελληνικά**

Σε αυτή τη μεταπτυχιακή διατριβή θα αναφερθούμε σε βασικά στοιχεία της μηχανικής μάθησης και θα δώσουμε τους ορισμούς για τις βασικές της έννοιες. Στη συνέχεια θα επικεντρώσουμε το ενδιαφέρον μας στην μονοταξική ταξινόμηση. Ακόμα, θα αναφερθούμε σε εφαρμογές βιολογικών διαδικασιών στον κλάδο της επιστήμης των υπολογιστών καθώς και το τι αποτελεί ένα τεχνητό ανοσοποιητικό σύστημα. Στη συνέχεια θα προσπαθήσουμε να περιγράψουμε το πως μπορεί ένα τεχνητό ανοσοποιητικό σύστημα να χρησιμοποιηθεί στο πρόβλημα της μονοταξικής ταξινόμησης. Θα αναφερθούμε σε αλγόριθμους Negative Selection, Real-Valued Negative Selection και θα επικεντρωθούμε στον V-Detector, με χρήση του οποίου θα πραγματοποιηθούν τα πειράματά μας. Για τα πειράματα θα χρειαστεί να παράγουμε τεχνητά δεδομένα σύμφωνα με μια γεωμετρική δομή την οποία θα αναλύσουμε στην ενότητα 4. Τέλος, θα παραθέσουμε τα αποτελέσματα των πειραμάτων και θα προσπαθήσουμε να αξιολογήσουμε τον αλγόριθμο και να εξάγουμε κάποια συμπεράσματα.

### **English**

In this Master's Thesis we are going to talk about some basic principles of machine learning and we are going to give the definitions about some basic concepts. Then we will focus our interest to the one class classification problem. Furthermore, we will talk about some biological processes, their use in computer science and what is an artificial immune system. Then, we will try to describe how could, an artificial immune system, be used in the problem of one class classification. We will mention the Negative Selection, Real-Valued Negative Selection algorithms and we will focus our attention on the V-Detector algorithm, with whom we are going to perform our experiments later. For our experiments we will have to generate our own artificial data according to a specific geometric structure. This procedure is analyzed in section 4. Finally, we are going to present the results of our experiments and we will try to evaluate the algorithm and draw some conclusions.

## 1 Εισαγωγή

Η Μηχανική Μάθηση αποτελεί έναν κλάδο της επιστήμης των υπολογιστών η οποία αναπτύχθηκε από τη μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη<sup>[1]</sup>. Από τον Tom M. Mitchell διατυπώθηκε ένας σχετικά αυστηρός ορισμός ο οποίος χρησιμοποιείται αρκετά συχνά. Ο συγκεκριμένος ορισμός αναφέρει<sup>[2]</sup> :

*Ένα πρόγραμμα υπολογιστή, λέγεται ότι μαθαίνει από την εμπειρία  $E$  σε σχέση με κάποια ομάδα εργασιών  $T$  και μέτρηση απόδοσης  $P$  εάν η απόδοσή του σε εργασίες στο  $T$ , όπως μετράται με  $P$ , βελτιώνεται με την εμπειρία  $E$ .*

Η άνθιση στον συγκεκριμένο κλάδο παρουσιάστηκε προς το τέλος της δεκαετίας του 1980 και στις αρχές της επόμενης καθώς άλλαξε ο τρόπος αντιμετώπισης των προβλημάτων από του επιστήμονες και αντί να δημιουργούνται προγράμματα τα οποία έχουν ήδη την γνώση που απαιτείται άρχισε η δημιουργία αλγορίθμων που εκμεταλλεύονται δεδομένα για να δημιουργούν συσχετίσεις μεταξύ τους ή αλλιώς να «μαθαίνουν» από αυτά. Αυτό δημιούργησε μεγάλο ενθουσιασμό για μελέτη και έρευνα και γρήγορα αναπτύχθηκαν πολλοί αλγόριθμοι οι οποίοι χρησιμοποιούνται μέχρι και σήμερα είτε αυτούσιοι είτε παραλλαγές τους.

Οι αλγόριθμοι της μηχανικής μάθησης μπορούν να χωριστούν σε διάφορες κατηγορίες είτε ως προς την φύση της εκπαίδευσης που πρέπει να υποστούν για να παράγουν αποτελέσματα είτε ως προς το επιθυμητό αποτέλεσμα που περιμένουμε να εξάγουμε από αυτούς.

Ως προς την φύση της εκπαίδευσης χωρίζονται συνήθως στις εξής 3 μεγάλες κατηγορίες :

### a) Επιτηρούμενη Μάθηση (Supervised Learning).

Οι αλγόριθμοι αυτής της κατηγορίας δέχονται τα έναν αριθμό από δεδομένα που έχουμε συλλέξει καθώς και τα αποτελέσματα τους και προσπαθούν να βρουν έναν κανόνα που να τα συσχετίζει μεταξύ τους. Στη συνέχεια αυτός ο κανόνας μπορεί να χρησιμοποιηθεί για δεδομένα που δεν έχει «δει» ο αλγόριθμος και να λάβουμε μια εκτίμηση για αυτά τα δεδομένα.

### b) Μη Επιτηρούμενη Μάθηση (Unsupervised Learning).

Σε αυτή την κατηγορία οι αλγόριθμοι που χρησιμοποιούνται δεν περιέχουν κάποια φάση εκπαίδευσης και ουσιαστικά προσπαθούν χωρίς καμία γνώση των δεδομένων να ανακαλύψουν έναν τρόπο για να οργανωθούν σε ομάδες ως προς κάποιο κριτήριο ομοιότητας.

### c) Ενισχυτική Μάθηση (Reinforcement Learning).

Το πρόγραμμα έρχεται σε αλληλεπίδραση με ένα δυναμικό περιβάλλον με στόχο την επίτευξη ενός στόχου. Σε αυτή τη διαδικασία δεν παρέχονται οδηγίες στο πρόγραμμα ως προς το πως θα γίνει αυτό αλλά αυτό μέσα από μια διαδικασία επαναλήψεων προσπαθεί να βελτιώνει προηγούμενες αποφάσεις ώστε να αποφέρουν ένα καλύτερο αποτέλεσμα.

Ενώ αντίστοιχα η κατηγοριοποίηση που προκύπτει ως προς το επιθυμητό αποτέλεσμα είναι η εξής :

#### a) Ταξινόμηση (Classification).

Τα διακριτά δεδομένα εισόδου χωρίζονται σε δύο ή παραπάνω κλάσεις και ο αλγόριθμος προσπαθεί να κατασκευάσει ένα μοντέλο το οποίο θα αντιστοιχίζει τα δεδομένα σε μια ή περισσότερες κλάσεις. Συνήθως η ταξινόμηση ανήκει στην κατηγορία της επιτηρούμενης μάθησης. Ένα παράδειγμα ταξινόμησης είναι η εφαρμογή φίλτρων spam στα e-mails ή η αναγνώριση χειρόγραφων ψηφίων.

#### b) Παλινδρόμηση (Regression).

Όπως και η ταξινόμηση, η παλινδρόμηση ανήκει παραδοσιακά στην κατηγορία της επιτηρούμενης μάθησης με την διαφορά ότι τα δεδομένα είναι συνεχή. Ένα παράδειγμα παλινδρόμησης αποτελεί η πρόβλεψη της τιμής ενός σπιτιού βάσει διαφόρων παραμέτρων όπως δωμάτια, περιοχή, όροφος, τετραγωνικά μέτρα, πρόσθετες παροχές κλπ.

**c) Συσταδοποίηση (Clustering).**

Σε αυτή την κατηγορία ο αλγόριθμος δέχεται ένα σύνολο από δεδομένα τα οποία προσπαθεί να χωρίσει σε ομάδες χωρίς να γνωρίζει το κάθε στοιχείο σε ποια ομάδα ανήκει. Ανήκει στην κατηγορία της μη επιτηρούμενης μάθησης.

**d) Εκτίμηση Πυκνότητας (Density Estimation).**

Ο αλγόριθμος δέχεται τα δεδομένα και προσπαθεί να βρει την κατανομή που ακολουθούν σε κάποιο χώρο.

**e) Μείωση Διαστασιμότητας (Dimensionality Reduction).**

Σε αυτή την κατηγορία ο αλγόριθμος δέχεται τα δεδομένα και προσπαθεί να τα απλοποιήσει αντιστοιχίζοντας τα έτσι σε έναν χώρο λιγότερων διαστάσεων.

Σύντομα φάνηκε ότι κάποια προβλήματα του πραγματικού κόσμου ήταν πολύ δύσκολο να αντιμετωπιστούν με αποτελεσματικότητα από τις παραδοσιακές επιστημονικές μεθόδους. Η υψηλή περιπλοκότητα πολλών προβλημάτων καθιστούσε μη αποτελεσματική την αντιμετώπιση τους μέσω των πατροπαράδοτων μεθόδων που βασίζονται στη στατιστική μοντελοποίηση και τον τρόπο που δημιουργούμε τα δεδομένα. Αυτή η μη αποτελεσματικότητα οφείλεται σε πολλούς παράγοντες μεταξύ των οποίων η μη γραμμικότητα του προβλήματος, η ύπαρξη πολλών διαστάσεων, χαοτικά φαινόμενα και πληθώρα από βαθμούς ελευθερίας και άγνωστες παραμέτρους στον μηχανισμό παραγωγής των δεδομένων. Αυτό έχει σαν αποτέλεσμα να χάνεται πληροφορία η οποία είναι απαραίτητη και κατά συνέπεια δεν μπορούμε να χρησιμοποιήσουμε παραδοσιακές μαθηματικές μεθόδους όταν θέλουμε να επιλύσουμε το πρόβλημα<sup>[3.1]</sup>.

Αυτά τα προβλήματα όμως τα οποία δεν μπορούμε να αντιμετωπίσουμε αποτελεσματικά με μαθηματικές μεθόδους, η φύση, σε διάφορα βιολογικά συστήματα, έχει βρει διαδικασίες για να αντιμετωπίζει προβλήματα με παρόμοια χαρακτηριστικά. Έτσι η επιστημονική κοινότητα προσπάθησε να μιμηθεί αυτή τη συμπεριφορά δημιουργώντας αλγόριθμους εμπνευσμένους από την λειτουργία αυτών των βιολογικών συστημάτων όπως :

- Τα βιολογικά νευρωνικά δίκτυα , δηλαδή οι διασυνδεδεμένοι βιολογικοί νευρώνες στο νευρικό σύστημα πολλών οργανισμών τους επιτρέπουν να μαθαίνουν, να απομνημονεύουν και να αναγνωρίζουν πρότυπα όπως εικόνες, ήχους, μυρωδιές κλπ.
- Τα μυρμήγκια παρουσιάζουν μια συλλογική και αυτό-οργανωμένη νοημοσύνη η οποία τους επιτρέπει να βρίσκουν τον συντομότερο δρόμο προς την τροφή που αναζητούν με πολύ αποτελεσματικό τρόπο.

Ένα τρίτο παράδειγμα αποτελεί και το βιολογικό ανοσοποιητικό σύστημα των σπονδυλωτών οργανισμών. Τα ανοσοποιητικά συστήματα αυτών των οργανισμών επιδεικνύουν μια υψηλού επιπέδου νοημοσύνη όταν πρόκειται για την άμυνα του οργανισμού καθώς πρέπει να είναι σε θέση να ξεχωρίζουν παθογόνα μέσα στον οργανισμό ώστε να ενεργοποιούνται οι αμυντικές λειτουργίες του οργανισμού και να εξασφαλίζεται ότι δεν θα ασθενήσει ο οργανισμός. Αντιλαμβανόμαστε έτσι ότι ο οργανισμός είναι σε θέση με κάποιο τρόπο να ξεχωρίζει ουσίες, ιούς και ξένα σώματα γενικότερα τα οποία δεν αντιστοιχίζονται με κάποιο τρόπο στον εαυτό του. Είναι πρακτικά αδύνατο ένας οργανισμός να απομνημονεύσει όλα αυτά τα ξένα στοιχεία καθώς αυτό το σύνολο είναι ασύλληπτα μεγάλο. Έτσι σε μια προσπάθεια να μιμηθούμε αυτή τη συμπεριφορά αυτών των οργανισμών δημιουργήθηκε μια ομάδα αλγορίθμων στους οποίους αναφερόμαστε ως «Τεχνητά Ανοσοποιητικά Συστήματα», με τους οποίους προσπαθούμε να αντιμετωπίσουμε τα προβλήματα που αναφέραμε παραπάνω πως οι παραδοσιακές μαθηματικές μέθοδοι δεν είναι τόσο αποτελεσματικές.

Σε αυτή την εργασία θα ασχοληθούμε με έναν τέτοιο αλγόριθμο και θα τον εφαρμόσουμε σε τεχνητά δεδομένα στην οποία η δημιουργία θα αναφερθούμε σε ξεχωριστή παράγραφο και τέλος θα παρουσιάσουμε τα συμπεράσματα μας.

## 2 Μονοταξική Ταξινόμηση

Σε αυτή την εργασία θα αναφερθούμε σε μια κατηγορία αλγόριθμων εμπνευσμένων από βιολογικά συστήματα και θα επικεντρωθούμε κυρίως στον αλγόριθμο V-Detector ο οποίος ανήκει στην κατηγορία των αλγορίθμων Negative Selection. Θα επικεντρώσουμε δε την προσοχή μας σε αλγόριθμους επιτηρούμενης μάθησης που έχουν ως επιθυμητό αποτέλεσμα την ταξινόμηση.

<sup>[3.2]</sup>Όπως αναφέραμε και παραπάνω στην επιτηρούμενη μάθηση τα δεδομένα τροφοδοτούνται στον αλγόριθμο σε ζευγάρια εισόδου-εξόδου. Πιο συγκεκριμένα, το κάθε δεδομένο-μέτρηση αποτελείται από ένα διάνυσμα  $N$  διαστάσεων, δηλαδή όσες και οι παράμετροι του προβλήματος που προσπαθούμε να επιλύσουμε, και το σχετικό αποτέλεσμα που αντιστοιχεί σε αυτό το διάνυσμα δεδομένων. Ο αλγόριθμος προσπαθεί ουσιαστικά να εντοπίσει μια σχέση ανάμεσα στα δεδομένα εισόδου και εξόδου ώστε να μπορεί να αντιστοιχίζει άλλα τέτοια διανύσματα εισόδου με μια ή περισσότερες τιμές από το σύνολο των απαντήσεων που έχουν παρασχεθεί.

Ένα βασικό στοιχείο που πρέπει να αναφερθεί εδώ είναι ότι οι παραδοσιακοί αλγόριθμοι ταξινόμησης λαμβάνουν δεδομένα εισόδου-εξόδου για κάθε μια από τις τάξεις για να μπορέσουν να δημιουργήσουν τη σχέση που αντιστοιχίζει τα δεδομένα εισόδου σε μια ή παραπάνω εξόδους.

Υπάρχουν περιπτώσεις όμως που δεν μας ενδιαφέρει η ταξινόμηση να γίνεται σε περισσότερες από 2 κλάσεις. Ουσιαστικά εξετάζουμε αν ένα στοιχείο ανήκει σε ένα σύνολο (σύνολο-εαυτός) ή αν δεν ανήκει σε αυτό. Σε πολλές τέτοιες περιπτώσεις είναι πολύ δύσκολο να συλλέξουμε δεδομένα που να εκφράζουν πλήρως την κλάση στην οποία δεν θα ανήκει το στοιχείο μας καθώς οι περιπτώσεις του μη-εαυτού είναι πρακτικά αμέτρητες. Συνεπώς η δυσκολία αυτή μεταφέρεται και στην «εκπαίδευση» του αλγόριθμου και στην αποτελεσματική αναγνώριση αυτών των στοιχείων σύμφωνα με την διαδικασία που περιγράψαμε παραπάνω. Ακόμα και αν σε κάποιο πρόβλημα καταφέρουμε να συλλέξουμε ένα πλήρες δείγμα του συνόλου μη-εαυτού το πιο πιθανόν είναι τα δεδομένα μας να είναι πολύ περισσότερα από αυτά του συνόλου εαυτού. Αυτό θα είχε ως αποτέλεσμα να υπάρχει κάποια μεροληψία κατά την διαδικασία της «εκπαίδευσης» του αλγόριθμου και συνεπώς θα ήταν πιθανόν να προκύψουν σημαντικά σφάλματα κατά την εφαρμογή του αλγορίθμου σε νέα δεδομένα που δεν έχει «ξαναδεί» ποτέ. Έτσι για να αντιμετωπιστεί αυτή η ιδιαιτερότητα αυτών των προβλημάτων δημιουργήθηκαν αλγόριθμοι κάποιοι από τους οποίους είναι εμπνευσμένοι από βιολογικές διαδικασίες όπως το ανοσοποιητικό σύστημα των σπονδυλωτών που αναφέραμε στην εισαγωγή. Η λειτουργία αυτών των αλγόριθμων βασίζεται στη λογική ότι δεν χρειάζεται να βρεθεί σχέση που να μπορεί να αναγνωρίζει και τα δύο σύνολα (εαυτού και μη-εαυτού). Αντ' αυτού μπορεί να βρεθεί μια σχέση αναγνώρισης των δεδομένων που θεωρούμε ότι ανήκουν στον σύνολο εαυτός και συνεπώς να αξιοποιηθεί αυτή ως μέσο ταξινόμησης μιας σειράς δεδομένων ενός στοιχείου σε μια από τις δύο κλάσεις. Δηλαδή να γίνεται έλεγχος αν αυτό το στοιχείο ανήκει στο σύνολο εαυτού και αν δεν ικανοποιεί την σχέση που έχει δημιουργηθεί τότε θα ανήκει στο σύνολο μη-εαυτός.

Πιο συγκεκριμένα, έχουν προταθεί διάφοροι τρόποι για την αντιμετώπιση του προβλήματος της μονοταξικής ταξινόμησης οι οποίοι μπορούν να συνοψιστούν στις παρακάτω κατηγορίες.

- Μέθοδοι εκτίμησης πυκνότητας.
- Μέθοδοι καθορισμού συνόρων.
- Μέθοδοι ανακατασκευής.

<sup>[3.3]</sup> Για κάθε μία από τις παραπάνω μεθόδους έχουν δημιουργηθεί διάφοροι αλγόριθμοι που προσεγγίζουν τα δεδομένα του κάθε προβλήματος με σκοπό να αναδείξουν διαφορετικά χαρακτηριστικά της δομής των δεδομένων όπως η κλιμάκωση τους, η ομαδοποίηση των δεδομένων σε συστάδες, την κυρτότητα της σχετικής κατανομής τους και η τοποθέτησή τους σε υποχώρους. Όλες αυτές οι μέθοδοι όμως διαθέτουν δύο κοινά στοιχεία. Το πρώτο είναι ένα μέτρο απόστασης  $d(\mathbf{z})$  ή ομοιότητας  $p(\mathbf{z})$  ενός δεδομένου  $\mathbf{z}$  από την κλάση του συνόλου εαυτού. Το δεύτερο στοιχείο είναι ένα κατώφλι  $\theta$  επάνω σε αυτό το μέτρο απόστασης ή ομοιότητας. Τα

νέα δεδομένα γίνονται δεκτά στο σύνολο εαυτός όταν η απόσταση τους ως προς το σύνολο εαυτός είναι μικρότερη από αυτό το κατώφλι ή η ομοιότητα τους με το σύνολο εαυτός ξεπερνάει αυτό το κατώφλι.

### 3 Τεχνητά Ανοσοποιητικά Συστήματα

<sup>[3.4]</sup>Ένα από τα βασικά χαρακτηριστικά του ανθρώπου είναι η περιέργεια του στο να παρατηρεί τον κόσμο γύρω του και να προσπαθεί να εξάγει θεωρίες οι οποίες προσπαθούν να τον περιγράψουν. Στον κόσμο που ζούμε και με τα φαινόμενα που συμβαίνουν καθημερινά δεν θα ήταν σωστό να αρκεστούμε μόνο στην παρατήρηση ή την επεξήγηση τους. Για χιλιάδες χρόνια η φύση καταφέρνει και διατηρεί τη ζωή στον πλανήτη και να εξελίσει τους οργανισμούς μέσα από διάφορες διαδικασίες. Αυτές οι διαδικασίες, οι οποίες είναι αποδεδειγμένο ότι έχουν εφαρμοστεί αποτελεσματικά, μπορούν να αποτελέσουν έμπνευση ώστε να τις μιμηθούμε ή να τις τροποποιήσουμε για να αντιμετωπίσουμε ένα πρόβλημα που με τους παραδοσιακούς τρόπους επίλυσης που διαθέτουμε, δε μπορούμε να τα λύσουμε με ικανοποιητική αποτελεσματικότητα. Έτσι οι πρόσφατες εξελίξεις που προέκυψαν στους τομείς της πληροφορικής, της μηχανικής και της τεχνολογίας γενικότερα οδήγησαν σε μια βαθύτερη κατανόηση του κόσμου γύρω μας και ειδικότερα, των βιολογικών συστημάτων. Γρήγορα αυτή η γνώση χρησιμοποιήθηκε σε πολλούς τομείς με αποτελέσματα όπως τα τεχνητά νευρωνικά δίκτυα, του εξελικτικού αλγόριθμους, κυτταρικά αυτοματοποιημένα συστήματα κλπ. Έτσι, προέκυψε μια νέα κατηγορία έρευνας γνωστή ως βιο-πληροφορική η οποία ασχολείται με την διαχείριση και επεξεργασία βιολογικών δεδομένων με εφαρμογές από υπολογιστική νοσημοσύνη, ρομποτική μέχρι και ανάλυση γονιδιομάτων. Από την άλλη η βιοιατρική μηχανική ξεκίνησε ως μια προσπάθεια για να συμπεριληφθεί η εφαρμογή των αρχών της μηχανικής σε βιολογικά και ιατρικά προβλήματα. Αυτή η αλληλεπίδραση μεταξύ της πληροφορικής και της βιολογίας μπορεί να εντοπιστεί κυρίως στις παρακάτω προσεγγίσεις :

- **Υπολογιστικές διαδικασίες και συστήματα εμπνευσμένα από την βιολογία.**

Σε αυτή την κατηγορία η βιολογία λειτουργεί ως πηγή των μοντέλων και έμπνευση ώστε να αναπτυχθούν υπολογιστικά συστήματα όπως για παράδειγμα τα τεχνητά νευρωνικά δίκτυα ή τα τεχνητά ανοσοποιητικά συστήματα.

- **Βιολογικές εφαρμογές εμπνευσμένες από υπολογιστικά συστήματα.**

Σε αυτή την κατηγορία η πληροφορική αξιοποιείται για να μπορέσει να παρέχει μοντέλα και έμπνευση για την βιολογία όπως για παράδειγμα κυτταρικά αυτοματοποιημένα συστήματα.

- **Υπολογιστικές διαδικασίες με βιολογικούς μηχανισμούς**

Σε αυτή την κατηγορία περιέχονται περιπτώσεις στις οποίες χρησιμοποιείται η υπολογιστική δυνατότητα των βιολογικών συστημάτων για να αντικαταστήσουν ή να συμπληρώσουν υπολογιστές βασισμένους σε πυρίτιο (όπως κβαντική ή DNA υπολογιστική).

Όπως αναφέραμε και παραπάνω έχουν δημιουργηθεί πολλοί αλγόριθμοι οι οποίοι είναι εμπνευσμένοι από βιολογικές διαδικασίες όπως το ανοσοποιητικό σύστημα των σπονδυλωτών οργανισμών. Για να μπορέσουμε να καταλάβουμε καλύτερα όμως αυτούς τους αλγόριθμους θα πρέπει να κατανοήσουμε την λειτουργία του ανοσοποιητικού μας συστήματος.

<sup>[3.5]</sup>Το ανοσοποιητικό σύστημα αποτελεί ένα φυσικό, γρήγορο και αποτελεσματικό μηχανισμό άμυνας ενός οργανισμού απέναντι σε μολύνσεις. Αποτελείται από μια διπλή γραμμή άμυνας, το έμφυτο και το προσαρμοστικό ανοσοποιητικό σύστημα. Και τα δύο συστήματα εξαρτώνται από την δραστηριότητα ενός μεγάλου αριθμού κυττάρων ανοσίας όπως τα λευκά αιμοσφαίρια. Πιο συγκεκριμένα, η έμφυτη ανοσία σχετίζεται κυρίως με κοκκιοκύτταρα και μακροφάγα, ενώ η προσαρμοστική ανοσία με λεμφοκύτταρα. Τα κύτταρα της έμφυτης γραμμής άμυνας του οργανισμού είναι διαθέσιμα άμεσα για την αντιμετώπιση ενός μεγάλου αριθμού διαφορετικών ειδών αντιγόνων ακόμα και χωρίς να έχουν εκτεθεί στο παρελθόν σε αυτά και να έχουν μάθει να τα αναγνωρίζουν.



Η παραγωγή αντισωμάτων κατά τον εντοπισμό ενός αναγνωρισθέντος μολυσματικού οργανισμού ονομάζεται *ειδική ανοσοαπόκριση* ή *αλλιώς προσαρμοστική ανοσοαπόκριση*. Η διαδικασία παραγωγής αντισωμάτων από το προσαρμοστικό ανοσοποιητικό σύστημα εκτελείται αποκλειστικά ως «απάντηση» σε συγκεκριμένες λοιμώξεις. Συνεπώς η ύπαρξη αντισωμάτων σε έναν οργανισμό αντικατοπτρίζει τις λοιμώξεις στις οποίες είχε εκτεθεί ο οργανισμός στο παρελθόν. Μια πολύ ενδιαφέρουσα πτυχή του προσαρμοστικού ανοσοποιητικού συστήματος αποτελεί η δυνατότητα των κυττάρων αυτών να αναπτύσσουν μνήμη ώστε να μπορούν να αναγνωρίζουν το ίδιο αντιγονικό ερέθισμα στον οργανισμό ξανά. Αυτή η λειτουργία επιτρέπει στον οργανισμό να μην νοσήσει ξανά από την ίδια ασθένεια. Συνεπώς, η προσαρμοστική ανοσοαπόκριση επιτρέπει στο ανοσοποιητικό σύστημα να βελτιώνεται σε κάθε καινούργια επαφή με κάποιο αντιγόνο.

Το ανοσοποιητικό σύστημα αποτελείται από διάφορα συστατικά μεταξύ των οποίων και τα λεμφοκύτταρα. Αυτά αποτελούν τους πρωτογενείς διαμεσολαβητές της προσαρμοστικής ανοσοαπόκρισης και είναι υπεύθυνα για την αναγνώριση και καταστροφή των παθογόνων παραγόντων. Αυτοί οι παράγοντες με την σειρά τους προάγουν την ανοσοποιητική μνήμη που εμφανίζεται μετά την έκθεση σε ασθένεια ή εμβολιασμό. Τα λεμφοκύτταρα συνήθως ενεργοποιούνται όταν υπάρχει κάποιο είδος αλληλεπίδρασης με ένα αντιγονικό ερέθισμα που οδηγεί στην ενεργοποίηση και τον πολλαπλασιασμό των λεμφοκυττάρων. Υπάρχουν δύο κατηγορίες λεμφοκυττάρων

- Β-λεμφοκύτταρα.
- Τ-λεμφοκύτταρα.

οι οποίες χαρακτηρίζονται από την ύπαρξη αντιγονικών υποδοχών, στην επιφάνεια τους, οι οποίες είναι ιδιαίτερα εξειδικευμένες για να αναγνωρίζουν ένα δεδομένο αντιγόνο. Συγκεκριμένα οι υποδοχές αυτές στα Β-λεμφοκύτταρα μπορούν να αναγνωρίσουν κομμάτια των αντιγόνων που κυκλοφορούν ελεύθερα στο αίμα. Οι υποδοχές αυτές αντιδρούν όταν συναντήσουν ένα αντιγόνο και το Β-λεμφοκύτταρο ενεργοποιείται, πολλαπλασιάζεται και τροποποιείται σε κύτταρα πλάσματος τα οποία μεταφέρουν μεγάλες ποσότητες αντισωμάτων στο αντιγόνο με αποτέλεσμα να το εξουδετερώνουν και τελικά να καταστρέφεται. Μερικά από αυτά τα ενεργοποιημένα Β-λεμφοκύτταρα θα παραμείνουν ενεργοποιημένα στο αίμα για μεγάλες χρονικές περιόδους ώστε να εξασφαλιστεί η άμυνα του οργανισμού απέναντι σε αντίστοιχα παθογόνα που μπορεί να τον προσβάλουν στο μέλλον.

Έτσι, ενώ το προσαρμοστικό ανοσοποιητικό σύστημα εξελίσσεται συνεχώς με αποτέλεσμα να δημιουργείται ανοσία απέναντι σε λοιμώξεις από παρόμοια αντιγόνα που έχει συναντήσει ο οργανισμός, η έμφυτη ανοσοαπόκριση παραμένει σταθερή κατά τη διάρκεια της ζωής ενός ατόμου ανεξάρτητα από την έκθεση του σε αντιγόνα. Συνδυάζοντας αυτά τα δύο συστήματα οι σπονδυλωτοί οργανισμοί καταφέρνουν να έχουν μια πολύ αποτελεσματική γραμμή άμυνας απέναντι σε λοιμώξεις.

Από την μελέτη αυτού του μηχανισμού η επιστημονική κοινότητα εμπνεύστηκε αλγόριθμους που μιμούνται αυτές τις διαδικασίες και δημιουργήθηκαν αλγόριθμοι για την αναγνώριση ενός συνόλου εαυτός έναντι ενός συνόλου μη εαυτός το οποίο μπορεί να περιέχει ακόμα και πολλά περισσότερα στοιχεία από το σύνολο εαυτός.

### 3.1 Negative Selection

<sup>[3.6]</sup>Ο μηχανισμός Negative Selection ήταν από τους πρώτους, ανάμεσα στους μηχανισμούς του ανοσοποιητικού μας συστήματος, που ερευνηθήκε ως υπολογιστικό παράδειγμα για την ανάπτυξη ενός τεχνητού ανοσοποιητικού συστήματος. Συχνά αναφερόμαστε σε αυτόν και ως *negative detection* καθώς το σύνολο των παραγόμενων ανιχνευτών προορίζεται να αναγνωρίζει γεγονότα που υποδεικνύουν ότι ένα δεδομένο σύνολο στοιχείων αποκλίνει από έναν δεδομένο κανόνα. Παρά τις σημαντικές εξελίξεις και διαφοροποιήσεις που έχουν εφαρμοστεί στην αρχική εκδοχή του αλγόριθμου τα βασικά του χαρακτηριστικά παραμένουν αμετάβλητα και έτσι δημιουργούν τη βάση για κάθε αλγόριθμο βασισμένο σε λογική του negative selection.

Η αρχική διατύπωση του αλγόριθμου περιλαμβάνει τον ορισμό μιας συλλογής  $S$  από αυτοχαρακτηριζόμενες συμβολοσειρές ίδιου μήκους πάνω σε ένα πεπερασμένο αλφάβητο  $\Sigma$  έτσι ώστε  $S \subset \Sigma^L$ . Αυτή η συλλογή συνήθως παρακολουθείται ή προστατεύεται από αλλαγές. Για παράδειγμα το  $S$  μπορεί να είναι ένα πρόγραμμα, ένα αρχείο δεδομένων ή ένα κανονικό πρότυπο δραστηριότητας, το οποίο είναι κατακερματισμένο σε μέρη ίσου μεγέθους. Ο αλγόριθμος εκτελείται σε δύο φάσεις, σε αυτή της εκπαίδευσης και έπειτα της δοκιμής ως εξής :

### **1. Φάση εκπαίδευσης**

Αυτή η φάση συνήθως αναφέρεται και ως φάση παραγωγής (generation phase) του αλγόριθμου καθώς περιλαμβάνει την δημιουργία ενός συνόλου  $R$  από ανιχνευτές οι οποίοι αποτυγχάνουν στο να ανιχνεύουν όλες τις αυτοχαρακτηριζόμενες συμβολοσειρές που ανήκουν στο  $S$ . Πιο συγκεκριμένα, κατά τη διαδικασία της φάσης δημιουργίας των ανιχνευτών δημιουργείται ένα αρχικό σύνολο εκπαιδευτών  $R_0$  μέσω μιας τυχαίας διαδικασίας, το οποίο στη συνέχεια συγκρίνεται με το σύνολο  $S$  των συμβολοσειρών. Έτσι παράγεται το σύνολο  $R$  με τους τελικούς αποδεκτούς ανιχνευτές του συνόλου μη εαυτού. Ένας σημαντικός παράγοντας του αλγόριθμου είναι ότι απαιτεί ακριβή ή τέλεια αντιστοίχιση για ένα δεδομένο ζεύγος συμβολοσειρών. Αντίθετα, ο βαθμός αντιστοίχισης αξιολογείται μερικώς, έτσι ώστε δύο συμβολοσειρές χαρακτηριστικών να θεωρούνται ότι ταιριάζουν όταν η συνάφεια τους υπερβαίνει ένα προκαθορισμένο όριο. Στην περίπτωση που επιλεγεί ένα δυαδικό αλφάβητο  $\Sigma = \{0, 1\}$  για την αναπαράσταση των δεδομένων, η ιδέα της μερικής αντιστοίχισης πραγματοποιείται με χρήση του κανόνα αντιστοίχισης ενός αριθμού από συνεχόμενα bits. Αυτός ο αριθμός επιλέγεται κατάλληλα ανάλογα το πρόβλημα που προσπαθούμε να επιλύσουμε.

### **2. Φάση Δοκιμής**

Αυτή η φάση συνήθως αναφέρεται και ως φάση παρακολούθησης (monitoring phase) ή φάση ανίχνευσης (detection phase). Εδώ μια επόμενη έκδοση  $S^*$  του συνόλου εαυτός  $S$  ελέγχεται συνεχώς έναντι του συνόλου ανιχνευτών  $R$  για αλλαγές, υποδεικνύοντας την ύπαρξη θορύβου ή διαφθοράς του αρχικού συνόλου δεδομένων. Συγκεκριμένα, κατά τη φάση ανίχνευσης του αλγόριθμου, το κάθε στοιχείο του συνόλου  $S^*$  ελέγχεται έναντι κάθε στοιχείου του συνόλου  $R$ . Αν οποιοσδήποτε ανιχνευτής αντιστοιχιστεί με κάποιο στοιχείο τότε είναι γνωστό ότι έχει συμβεί κάποια αλλαγή από τη στιγμή που οι ανιχνευτές έχουν σχεδιαστεί έτσι ώστε να μην ταιριάζουν με καμία από τις αρχικές συμβολοσειρές του  $S$ .

Στην παρακάτω λίστα συνοψίζουμε τις πιο σημαντικές πτυχές του αλγόριθμου δικαιολογώντας την εγκυρότητα τους ως εναλλακτική προσέγγιση ανίχνευση ανωμαλιών.

1. Κάθε αντίγραφο ενός αλγόριθμου ανίχνευσης είναι μοναδικό. Οι περισσότεροι αλγόριθμοι ανίχνευσης χρειάζεται να παρακολουθούν πολλές τοποθεσίες. Σε αυτά τα καταναμημένα περιβάλλοντα, οποιοδήποτε απλό σύστημα ανίχνευσης είναι απίθανο να είναι αποτελεσματικό, αφού η αποτυχία ανίχνευσης σε μία από τις τοποθεσίες έχει ως αποτέλεσμα τη συνολική αποτυχία όλου του συστήματος. Η προσέγγιση του αλγόριθμου αρνητικής επιλογής, από την άλλη, εξασφαλίζει την προστασία της κάθε τοποθεσίας με ένα μοναδικό σύνολο από ανιχνευτές για την κάθε μία. Έτσι ακόμα και αν μια τοποθεσία εκτεθεί σε αλλαγές οι υπόλοιπες παραμένουν προστατευμένες.
2. Η ανίχνευση εμπλέκει τυχαιότητα. Μια συνέπεια της χρήσης διαφορετικών συνόλων ανιχνευτών για την προστασία κάθε οντότητας, αποτελεί ότι είναι εφικτές οι πιθανοτικές μέθοδοι ανίχνευσης. Αυτό είναι δυνατό, επειδή μια διείσδυση σε μια τοποθεσία είναι απίθανο να είναι επιτυχής σε πολλαπλές τοποθεσίες. Η χρήση πιθανοτικών μεθόδων ανίχνευσης εξασφαλίζει υψηλή αξιοπιστία σε επίπεδο συστήματος σε σχετικά χαμηλό υπολογιστικό κόστος και απαιτήσεις μνήμης. Το τίμημα της χρήσης αυτής της τεχνικής είναι μια κάπως μεγαλύτερη πιθανότητα διείσδυσης σε κάθε τοποθεσία.
3. Η λειτουργία ανίχνευσης του αλγόριθμου βασίζεται στην ικανότητα του να αναγνωρίζει πιθανώς οποιαδήποτε ξένα δραστηριότητα, παρά να αναζητά γνωστά πρότυπα ανωμαλίας. Τα παραδοσιακά συστήματα προστασίας όπως η ανίχνευση ιών ή εισχωρήσεων λειτουργούν αναζητώντας μοναδικά πρότυπα τα οποία είναι γνωστά τη

χρονική στιγμή που διανέμεται το πρόγραμμα ανίχνευσης και κατά συνέπεια τα συστήματα μένουν ευάλωτα σε επιθέσεις νέων μέσων. Αντίθετα, η προσέγγιση του αλγόριθμου Negative Selection μαθαίνει τα πρότυπα τα οποία έρχονται σε αντίθεση με το σύνολο εαυτός και έτσι μπορεί να διακρίνει αποκλίσεις από αυτό.

Τέλος, θα αναφέρουμε τα πιο **σημαντικά χαρακτηριστικά** αυτής της διαφορετικής προσέγγισης για το πρόβλημα της ταξινόμησης.

1. Η αρνητική αναπαράσταση της πληροφορίας καθώς ο αλγόριθμος χρησιμοποιεί την συμπληρωματική έννοια από αυτή που χρησιμοποιείται παραδοσιακά στην μηχανική μάθηση. Παρά το γεγονός ότι είναι σαφώς διαφορετικός από άλλους μηχανισμούς μάθησης, έχει αποδείξει την αποτελεσματικότητα και εφαρμοσιμότητα του σε πολλές εφαρμογές.
2. Η χρήση ενός συνόλου ανιχνευτών ως μηχανισμό αναγνώρισης-ταξινόμησης. Η δημιουργία ενός κατάλληλου συνόλου ανιχνευτών αποτελεί ένα σημαντικό στοιχείο του αλγορίθμου ρυθμίζοντας έτσι την κάλυψη του συνόλου μη εαυτός και την απόδοση της ταξινόμησης.
3. Το πιο σημαντικό στοιχείο του αλγόριθμου, το οποίο μας ενδιαφέρει ιδιαίτερω σε αυτή την εργασία, είναι η εγκυρότητα του ως μια εναλλακτική προσέγγιση για προβλήματα μονοταξικής ταξινόμησης και ιδιαίτερω σε προβλήματα που υπάρχει απόκλιση μεταξύ της ποσότητας των δεδομένων που διαθέτουμε για την κάθε κλάση. Αυτό συμβαίνει καθώς οι ανιχνευτές δημιουργούνται βάσει των δεδομένων της κλάσης εαυτός που συνήθως είναι και αυτή με το λιγότερο πλήθος δεδομένων.

### 3.2 Real Valued Negative Selection.

<sup>[3.7]</sup>Το μεγαλύτερο μέρος της έρευνας στους αλγόριθμους αρνητικής επιλογής αξιοποιεί το γενικότερο πλαίσιο που παρέχεται από τους δυαδικούς χώρους. Η δυαδική αναπαράσταση του προβλήματος αποτελεί μια εύκολη λύση καθώς παρέχει ένα πεπερασμένο χώρο και έτσι είναι πιο εύκολο να γίνει η ανάλυση του προβλήματος χρησιμοποιώντας μεθόδους απαριθμητικής συνδυαστικής και είναι πιο εύκολο να χρησιμοποιηθεί για κατηγοριοποιημένα δεδομένα. Παρόλα αυτά, στον πραγματικό κόσμο τα περισσότερα προβλήματα απαιτούν τη χρήση ενός συνεχούς χώρου πραγματικών τιμών καθώς ένας τέτοιος χώρος παρέχει περισσότερες δυνατότητες αναπαράστασης της πληροφορίας σε σχέση με έναν δυαδικό χώρο. Πιο συγκεκριμένα, υπάρχει ένας μεγάλος αριθμός προβλημάτων ταξινόμησης τα οποία ανήκουν από τη φύση τους σε έναν συνεχή χώρο πραγματικών τιμών και έτσι η υλοποίηση ενός αλγορίθμου που αξιοποιεί αυτόν τον χώρο θα μπορούσε να διαθέτει τη μεγαλύτερη ευστάθεια και να παρέχει τα πιο αποδοτικά αποτελέσματα.

Οι Real-Valued Negative Selection αλγόριθμοι λειτουργούν βασιζόμενοι στις ίδιες αρχές όπως και οι αρχικοί Negative Selection αλγόριθμοι σε δυαδικούς χώρους. Οι ανιχνευτές δημιουργούνται τυχαία και στη συνέχεια αφαιρούνται όσοι χρειάζεται σύμφωνα με την φιλοσοφία του αρχικού αλγορίθμου με μόνη διαφορά την αξιοποίηση ενός υπέρ-αριθμήσιμου χώρου. Έτσι οι ανιχνευτές αυτοί αναπαρίστανται σε αυτό τον χώρο ως σημεία με μια σταθερή ακτίνα ή ισοδύναμα ως N-σφαίρες σταθερού όγκου στον χώρο του προβλήματος.

Ανεξάρτητα από το μέτρο ομοιότητας (ή ανομοιότητας) που θα χρησιμοποιηθεί, και τον αντίστοιχο κανόνα αντιστοίχισης, οι ανιχνευτές που δημιουργούνται έχουν ορισμένα κοινά χαρακτηριστικά όπως ο αριθμός των κοινών ψηφίων που απαιτούνται για την αντιστοίχιση σε ένα δυαδικό χώρο συμβολοσειρών, ή το κατώφλι της απόστασης που καθορίζει την αναγνώριση όταν εκτιμούμε δεδομένα πραγματικών τιμών. Γενικότερα υπάρχει ένα όριο συνάφειας μεταξύ των ανιχνευτών που καθορίζει την κατάσταση αναγνώρισης μεταξύ ενός δεδομένου ζεύγους χαρακτηριστικών. Κατά συνέπεια τα χαρακτηριστικά των ανιχνευτών μπορούν να επεκταθούν ώστε να ξεπεράσουν αυτόν τον περιορισμό. Η ιδέα αυτή εκφράστηκε μέσω του αλγορίθμου V-Detector τον οποίο θα μελετήσουμε σε αυτή την εργασία.

Η βασική ιδέα πίσω από αυτόν τον αλγόριθμο είναι η ανάθεση διαφορετικού χώρου αναγνώρισης γύρω από τον κάθε ανιχνευτή μέσω της παραμετροποίησης του κατώφλιου  $\theta$ . Πιο συγκεκριμένα ο κάθε ανιχνευτής συνεχίζει να αποτελεί μια σφαίρα στις  $N$  διαστάσεις όπως και στην απλή εκδοχή του αλγόριθμου Real-Valued Negative Selection με την διαφορά ότι η ακτίνα του κάθε ανιχνευτή μπορεί να είναι διαφορετική από τις υπόλοιπες. Άλλες διαφοροποιήσεις που μπορούν να χρησιμοποιηθούν είναι η αλλαγή της συνάρτησης που χρησιμοποιείται για την μέτρηση της απόστασης των ανιχνευτών από τα δεδομένα. Σε αυτή την εργασία χρησιμοποιείται η γενικευμένη Ευκλείδεια απόσταση. Η χρήση διαφορετικών μέτρων σύγκρισης μπορεί να οδηγήσει σε διαφοροποίηση του γεωμετρικού σχήματος που ορίζει ο κάθε ανιχνευτής. Έτσι μπορούμε με τις κατάλληλες συναρτήσεις να έχουμε ανιχνευτές οι οποίοι καλύπτουν  $N$ -παραλληλόγραμμα ή  $N$ -υπερβολές, σχήματα τα οποία αποτελούν τη γενίκευση των αντίστοιχων γεωμετρικών σχημάτων στις 2 διαστάσεις.

### 3.3 Ο Αλγόριθμος V-Detector

<sup>[3.8]</sup>Ο αλγόριθμος που θα χρησιμοποιηθεί σε αυτή την διπλωματική εργασία για την ταξινόμηση των δεδομένων είναι ο V-Detector . Ο συγκεκριμένος αλγόριθμος ανήκει στην οικογένεια των Real Valued Negative Selection Algorithms με μεταβλητή ακτίνα ανιχνευτών. Στην αρχική εκδοχή του αλγόριθμου η λειτουργία του επικεντρώνεται στη τυχαία δημιουργία ενός συνόλου ανιχνευτών μέχρις ότου η εκτιμώμενη κάλυψη του χώρου που καταλαμβάνει το σύνολο μη-εαυτός να είναι περίπου ίση με μια προκαθορισμένη τιμή  $\Theta_0$  και ταυτόχρονα η εκτιμώμενη κάλυψη του χώρου που καταλαμβάνει το σύνολο εαυτός να βρίσκεται κάτω από ένα προκαθορισμένο κατώφλι  $\Theta_{self}$ . Ο αλγόριθμος δέχεται ως ορίσματα το σύνολο  $S$  του συνόλου εαυτός, την ακτίνα ανίχνευσης  $R_{self}$ , καθώς και τον μέγιστο αριθμό από ανιχνευτές που μπορούν να δημιουργηθούν. Η εκτιμώμενη κάλυψη των χώρων που ορίζουν τα σύνολα εαυτός και μη εαυτός αποτελούν και τις βασικές παραμέτρους ελέγχου για τον αλγόριθμο V-Detector καθώς παρέχουν επιπρόσθετα κριτήρια διακοπής κατά τη φάση δημιουργίας των ανιχνευτών. Η εκτιμώμενη κάλυψη είναι ένα υποπροϊόν των ανιχνευτών και βασίζεται στην ιδέα ότι, όταν λάβουμε ένα δείγμα  $M$  σημείων στον χώρο και δεν καλύπτεται ένα μόνο σημείο τότε η εκτιμώμενη κάλυψη θα είναι δίνεται από την ποσότητα  $1 - 1/M$ . Αν εκτελεστούν  $L$  τυχαία πειράματα χωρίς να βρεθεί ένα μη καλυπτόμενο σημείο τότε μπορούμε να συμπεράνουμε ότι η εκτιμώμενη κάλυψη του χώρου είναι τουλάχιστον  $A = 1 - 1/M$ . Έτσι, καταλήγουμε στο ότι ο απαιτούμενος αριθμός προσπαθειών ώστε η απαιτούμενη κάλυψη να είναι  $A$  δίνεται από τον τύπο  $M = 1/(1-A)$ .

Η φάση δημιουργίας των ανιχνευτών αποτελεί μια διαδικασία συνεχούς δειγματοληψίας επάνω στον χώρο που εξετάζουμε, σύμφωνα με την ομοιόμορφη κατανομή μέχρις ότου το σύνολο των ανιχνευτών να φτάσει το προκαθορισμένο όριο. Παρόλα αυτά η φάση δημιουργίας μπορεί να τερματιστεί πριν επιτευχθεί ο προκαθορισμένος αριθμός ανιχνευτών εάν η κάλυψη των χώρων που αντιστοιχούν στο σύνολο εαυτός και μη εαυτός ξεπεράσουν ένα προκαθορισμένο κατώφλι. Ειδικότερα, κατά τη φάση δημιουργίας κάθε υποψήφιος ανιχνευτής ελέγχεται ως προς το αν καλύπτει τα δεδομένα του συνόλου εαυτός και τους ανιχνευτές που έχουν ήδη γίνει αποδεκτοί. Κάθε φορά που ένας ανιχνευτής ο οποίος βρίσκεται εντός των ήδη καλυμμένων χώρων του συνόλου εαυτός ή μη-εαυτός καταγράφεται το γεγονός σε έναν μετρητή για να ληφθεί υπόψη στον υπολογισμό της τρέχουσας εκτίμησης της κάλυψης των δύο χώρων. Επομένως, σε περίπτωση που η κάλυψη αυτή υπερβεί τις δεδομένες ποσοότητες που έχουν οριστεί ως κατώφλι, η φάση δημιουργίας τερματίζεται και έχουμε στη διάθεση μας το σύνολο των ανιχνευτών. Εάν ο υποψήφιος ανιχνευτής ούτε βρίσκεται σε καλυπτόμενη περιοχή του συνόλου μη εαυτός, ούτε αποτελεί στοιχείο του συνόλου εαυτός, αποθηκεύεται ως έγκυρος ανιχνευτής του συνόλου μη εαυτός και η αντίστοιχη ακτίνα ανίχνευσης βασίζεται στην ελάχιστη απόσταση από κάθε ανιχνευτή που θα διατηρηθεί.

Είναι σημαντικό να σημειωθεί ότι εάν ο αλγόριθμος υπερβεί την προκαθορισμένη κάλυψη του συνόλου εαυτός τότε θα προκύψει ένας μη ομαλός πρόωρος τερματισμός του. Αυτό μπορεί να συμβεί όταν τα δεδομένα του συνόλου εαυτός που παρέχονται καλύπτουν σχεδόν πλήρως τον

χώρο του προβλήματος ή όταν οι ακτίνες αναγνώρισης των στοιχείων του συνόλου εαυτός είναι αρκετά μεγάλες και έτσι ολόκληρος ο χώρος αντιστοιχίζεται σε αυτή την κλάση.

Στη συνέχεια ακολουθεί αναλυτική περιγραφή του αλγόριθμου καθώς και των μεταβλητών και πινάκων που θα χρησιμοποιηθούν.

- $S \in M_n \times L$  ορίζουμε τον πίνακα που αποθηκεύονται τα πρότυπα προς ανάλυση. Η κάθε γραμμή αυτού του πίνακα αντιστοιχεί στα  $n$  δεδομένα  $L$  διαστάσεων που είναι διαθέσιμα για κάθε πρότυπο.
- $D \in M_m \times L$  ορίζουμε τον πίνακα που αποθηκεύονται οι ανιχνευτές. Αντίστοιχα και εδώ η κάθε γραμμή του πίνακα αντιστοιχεί σε έναν ανιχνευτή  $L$  διαστάσεων.
- $R \in M_1 \times m$  ορίζουμε το διάνυσμα στο οποίο αποθηκεύονται οι αντίστοιχη ακτίνα του κάθε ανιχνευτή. Η ακτίνα σε μια γραμμή  $k$  αντιστοιχίζεται με τον ανιχνευτή που ορίζεται από την αντίστοιχη γραμμή  $k$  στον πίνακα  $D$ .
- $D_d \in M_1 \times m$  ορίζουμε το διάνυσμα στο οποίο αποθηκεύουμε τις αποστάσεις μεταξύ του ανιχνευτή που δημιουργείται σε ένα συγκεκριμένο βήμα και όλων των υπόλοιπων ανιχνευτών.
- $I_d \in M_1 \times k$ , όπου  $k \leq m$ , ορίζουμε το διάνυσμα που αποθηκεύει τις θέσεις (στον πίνακα) των ανιχνευτών που έχουμε ήδη αξιολογήσει και αποδεχτεί για τους οποίους η απόσταση από τον ανιχνευτή προς εξέταση είναι μικρότερη από την αντίστοιχη ακτίνα.
- $D_s \in M_1 \times n$  ορίζουμε το διάνυσμα που θα αποθηκεύονται οι αποστάσεις του ανιχνευτή που εξετάζουμε από όλα τα πρότυπα του συνόλου εαυτός.
- SCS (Self-Coverage Status) είναι μια λογική μεταβλητή που δείχνει εάν η εκτιμώμενη κάλυψη του χώρου που ορίζει το σύνολο εαυτός έχει ξεπεράσει το προκαθορισμένο κατώφλι.
- NCS (Non-Self-Coverage Status) είναι μια λογική μεταβλητή που δείχνει εάν η εκτιμώμενη κάλυψη του χώρου που ορίζει το σύνολο μη-εαυτός έχει ξεπεράσει το προκαθορισμένο κατώφλι.
- COVERED είναι μια λογική μεταβλητή που μας δείχνει εάν ο υποψήφιος ανιχνευτής που δημιουργείται καλύπτεται ήδη από κάποιον άλλον.
- SELF είναι μια λογική μεταβλητή που μας δείχνει εάν ο υποψήφιος ανιχνευτής βρίσκεται μέσα στον χώρο που ορίζουν τα πρότυπα του συνόλου εαυτός.
- $C_0$  είναι η τιμή που ορίζουμε ως στόχο για να θεωρήσουμε ότι ο χώρος που ορίζει το σύνολο μη εαυτός έχει καλυφθεί.
- $C_{self}$  είναι το μέγιστο επιτρεπτό ποσοστό της κάλυψης του χώρου που ορίζει το σύνολο εαυτός.
- $T_{max}$  είναι ο μέγιστος αριθμός ανιχνευτών προς δημιουργία.
- $R_{self}$  είναι η ακτίνα που αντιστοιχεί στα σημεία του συνόλου εαυτός. Δηλαδή τα σημεία αυτά σχηματίζουν σφαίρες  $L$  διαστάσεων με ακτίνα  $R_{self}$  και αθροίζοντας αυτές τις (υπέρ)σφαίρες προκύπτει ο χώρος του συνόλου εαυτός.

Στη συνέχεια ακολουθεί η αλγοριθμική εκδοχή του V-Detector.

$m \leftarrow 0$

$D \leftarrow []$

$R \leftarrow []$

SCS  $\leftarrow$  FALSE

NCS ← FALSE

While ( $m \leq T_{max}$ ) and (NCS = FALSE) and (SCS = FALSE) do:

$t \leftarrow 0$

$T \leftarrow 0$

  COVERED ← FALSE

  SELF ← FALSE

  While (COVERED != TRUE) or (SELF != TRUE) do:

- Υπολόγισε διάνυσμα  $x \in M_1 \times L$  ως ένα τυχαίο σημείο στον χώρο  $[0, 1]^L$ , σύμφωνα με την ομοιόμορφη κατανομή.

- Υπολόγισε πίνακα  $D_d \in M_1 \times m$  σύμφωνα με την εξίσωση

$D_d(i) = D(x, D(i, :))$ , για κάθε  $i \in [m]$ .

- Υπολόγισε διάνυσμα  $I_d \in M_1 \times k$ , όπου  $k \leq m$ , σύμφωνα με την εξίσωση

$I_d = \{i \in [m] : D_d(i) \leq R(i)\}$ .

- If ( $k = 0$ ) then:

  COVERED ← TRUE

else:

$t \leftarrow t + 1$

  if ( $t > 1/(1-C_0)$ ) then:

    NCS ← TRUE

- If (COVERED = FALSE) and (NCS = FALSE) then:

- Υπολόγισε το διάνυσμα  $D_s \in M_1 \times n$  σύμφωνα με την εξίσωση

$D_s(i) = D(x, S_i)$ , για κάθε  $i \in [n]$

- Υπολόγισε την ακτίνα του υποψήφιου ανιχνευτή σύμφωνα με την εξίσωση

$r = \min\{D_s(i) - R_{self}\}$

- If ( $r > R_{self}$ ) then:

- $m \leftarrow m + 1$

- $R \leftarrow [R, r]$

- $D \leftarrow [D; x]$

else:

- SELF ← TRUE

- $T \leftarrow T + 1$

- If ( $T > 1/(1-C_{self})$ ) then:

    SELF ← TRUE

Η πολυπλοκότητα του αλγόριθμου ως προς τον χρόνο μπορεί να υπολογιστεί αν λάβουμε υπόψη μας ότι το σημείο που απαιτεί τον περισσότερο χρόνο είναι αυτό που υπολογίζεται η απόσταση του υποψήφιου ανιχνευτή από όλα τα πρότυπα του συνόλου εαυτός. Ο κάθε ένας τέτοιος υπολογισμός μπορεί να γίνει σε χρόνο τάξης  $O(n)$  όπου  $n$  ο αριθμός των προτύπων του

συνόλου που τροφοδοτείται για την εκπαίδευση του αλγόριθμου. Συνεπώς, αν χρειαζόμαστε  $m$  ανιχνευτές η συνολική πολυπλοκότητα που προκύπτει είναι  $O(mn)$ .

Οι απαιτήσεις του αλγόριθμου σε μνήμη μπορούν να υπολογιστούν αθροίζοντας τη μνήμη που απαιτούν οι πιο απαιτητικές μεταβλητές, δηλαδή οι  $S$ ,  $D$ ,  $R$ ,  $D_d$ ,  $D_s$ ,  $I_d$ . Συνολικά προκύπτει ότι η απαιτούμενη μνήμη είναι  $(m+n)L+2m+k+n$ .

#### 4 Δημιουργία Τεχνητών Δεδομένων

Για την εφαρμογή του αλγόριθμου V-Detector σε αυτή την εργασία θα χρειαστεί να παράγουμε δικά μας τεχνητά δεδομένα. Για να επιτευχθεί αυτό θα χρησιμοποιήσουμε τη συνάρτηση `mnrand` του MATLAB η οποία παράγει έναν αριθμό από σημεία  $N$  διαστάσεων, σύμφωνα με την κανονική κατανομή, γύρω από έναν δεδομένο πυρήνα με έναν πίνακα συσχέτισης. Σε αυτή την εργασία θα εξετάσουμε μόνο δεδομένα τα οποία είναι ασυσχέτιστα μεταξύ τους. Καθώς θέλαμε να δημιουργήσουμε δεδομένα τα οποία προσομοιώνουν πολλά προβλήματα του πραγματικού κόσμου δώσαμε ιδιαίτερη προσοχή στην πολυτροπικότητα (multimodality) των δεδομένων που θα παράγουμε. Έτσι η συνάρτηση `CreateData` που δημιουργήσαμε δέχεται ως ορίσματα τα παρακάτω.

- Τον αριθμό των συντεταγμένων του κάθε σημείου δηλαδή την διάσταση των δεδομένων του προβλήματος.
- Τον αριθμό των πυρήνων της κλάσης εαυτός.
- Τον αριθμό των πυρήνων της κλάσης μη-εαυτός.
- Τον αριθμό των δεδομένων της κλάσης εαυτός που θέλουμε να παράγουμε.
- Τον αριθμό των δεδομένων της κλάσης μη-εαυτός που θέλουμε να παράγουμε.

Η παραπάνω συνάρτηση επιστρέφει σε δύο πίνακες τα δεδομένα σημεία των κλάσεων εαυτός και μη-εαυτός. Για την δημιουργία των υπόλοιπων πυρήνων δημιουργούνται ακτίνες με τιμές ανάμεσα σε μια ελάχιστη και μέγιστη τιμή. Αυτή η ελάχιστη και μέγιστη τιμή είναι διαφορετική για τους πυρήνες του συνόλου εαυτός και για τους πυρήνες του συνόλου μη-εαυτός. Στη συνέχεια για τον αριθμό των πυρήνων που μας ενδιαφέρουν δημιουργούνται ακτίνες ανάμεσα σε αυτές τις 2 τιμές. Η ακτίνες αυτές τροφοδοτούνται στην συνάρτηση `CreateKernels` η οποία και δημιουργεί τους πυρήνες με χρήση σφαιρικών συντεταγμένων. Η συνάρτηση `CreateKernels` θα αναλυθεί στη συνέχεια. Η διαδικασία παραγωγής των δεδομένων αυτών συνοψίζεται στα παρακάτω βήματα :

1. Δημιουργείται ο πίνακας `Sigma` ο οποίος είναι ένας τετραγωνικός πίνακας με τόσες γραμμές και στήλες όσες και οι διαστάσεις των δεδομένων. Περιέχει την τιμή μηδέν σε όλες τις θέσεις εκτός από αυτές της κύριας διαγωνίου, όπου και περιέχει την τιμή 1.
2. Αρχικοποιούμε τον πίνακα `SelfKernels` με μηδενικά.
3. Δημιουργούμε τον πρώτο πυρήνα και τον εισάγουμε στην πρώτη γραμμή του πίνακα `SelfKernels`. Οι τιμές για τις συντεταγμένες του πρώτου πυρήνα είναι τυχαίοι ακέραιοι αριθμοί στο διάστημα  $[1, 9]$
4. Ορίζεται η μέγιστη και ελάχιστη ακτίνα για τους πυρήνες του συνόλου εαυτός και μη-εαυτός και δημιουργούνται οι τυχαίες ακτίνες για τον κάθε πυρήνα.
5. Δημιουργούνται πρώτα οι υπόλοιποι πυρήνες του συνόλου εαυτός, και μαζί με τον κεντρικό πυρήνα σχηματίζουν τον πίνακα `SelfKernels`. Αντίστοιχα οι πυρήνες του συνόλου μη-εαυτός που δημιουργούνται σχηματίζουν τον πίνακα `NonSelfKernels`.
6. Δημιουργούνται τυχαίοι ακέραιοι αριθμοί, τόσο όσοι και το πλήθος των δεδομένων του συνόλου εαυτός. Ο καθένας από αυτούς παίρνει ως τιμή έναν ακέραιο ο οποίος αντιστοιχίζεται με έναν πυρήνα. Έτσι, στη συνέχεια βρίσκουμε πόσες φορές έχει δημιουργηθεί η κάθε τιμή και αντιστοιχίζουμε αυτό το πλήθος με τον αριθμό των δεδομένων που θα δημιουργηθούν για τον κάθε πυρήνα. Τέλος, δημιουργούμε τα

δεδομένα που αντιστοιχούν στον κάθε πυρήνα τα οποία σχηματίζουν τον πίνακα SelfSamples.

7. Η ίδια διαδικασία ακολουθείται και για τη δημιουργία των στοιχείων του συνόλου μη-εαυτός, τα οποία αποθηκεύονται στον πίνακα NonSelfSamples.
8. Τέλος, εμφανίζουμε ένα γράφημα με τους πυρήνες που δημιουργήθηκαν.

Με την παραπάνω διαδικασία έχουμε έναν έλεγχο στο πόσο «κοντά» ή «μακριά» θα είναι τα δεδομένα μας και συνεπώς έτσι ρυθμίζεται η δυσκολία του προβλήματος που καλούμαστε να αντιμετωπίσουμε. Ταυτόχρονα με αυτό τον τρόπο μπορούμε να δημιουργήσουμε κατανομές δεδομένων πολλών μεταβλητών γύρω από πολλούς πυρήνες όπως ήταν και ο αρχικός μας στόχος, ώστε να μπορούμε να ελέγξουμε διαφορετικές περιπτώσεις δόμησης της πληροφορίας.

Στη συνέχεια θα αναφερθούμε στη συνάρτηση CreateKernels και στη διαδικασία δημιουργίας των πυρήνων. Η παραπάνω συνάρτηση δέχεται ως ορίσματα:

- Τις διαστάσεις των δεδομένων.
- Τον αριθμό των πυρήνων που θέλουμε να παράγουμε.
- Τον πίνακα των ακτινών R.
- Τον κεντρικό πυρήνα γύρω από τον οποίο θα δημιουργηθούν οι υπόλοιποι.

Οι συντεταγμένες των νέων πυρήνων προκύπτουν με τον υπολογισμό σημείων επάνω σε μια σφαίρα N-διαστάσεων, όπου N οι διαστάσεις του προβλήματος που μελετάμε στην εκάστοτε περίπτωση. Η N-σφαίρα θα έχει ως κέντρο τον αρχικό πυρήνα που παρέχεται ως όρισμα και η ακτίνα θα μεταβάλλεται ανά πυρήνα σύμφωνα με τις τιμές του πίνακα R. Οι τύποι που χρησιμοποιήθηκαν είναι οι εξής :

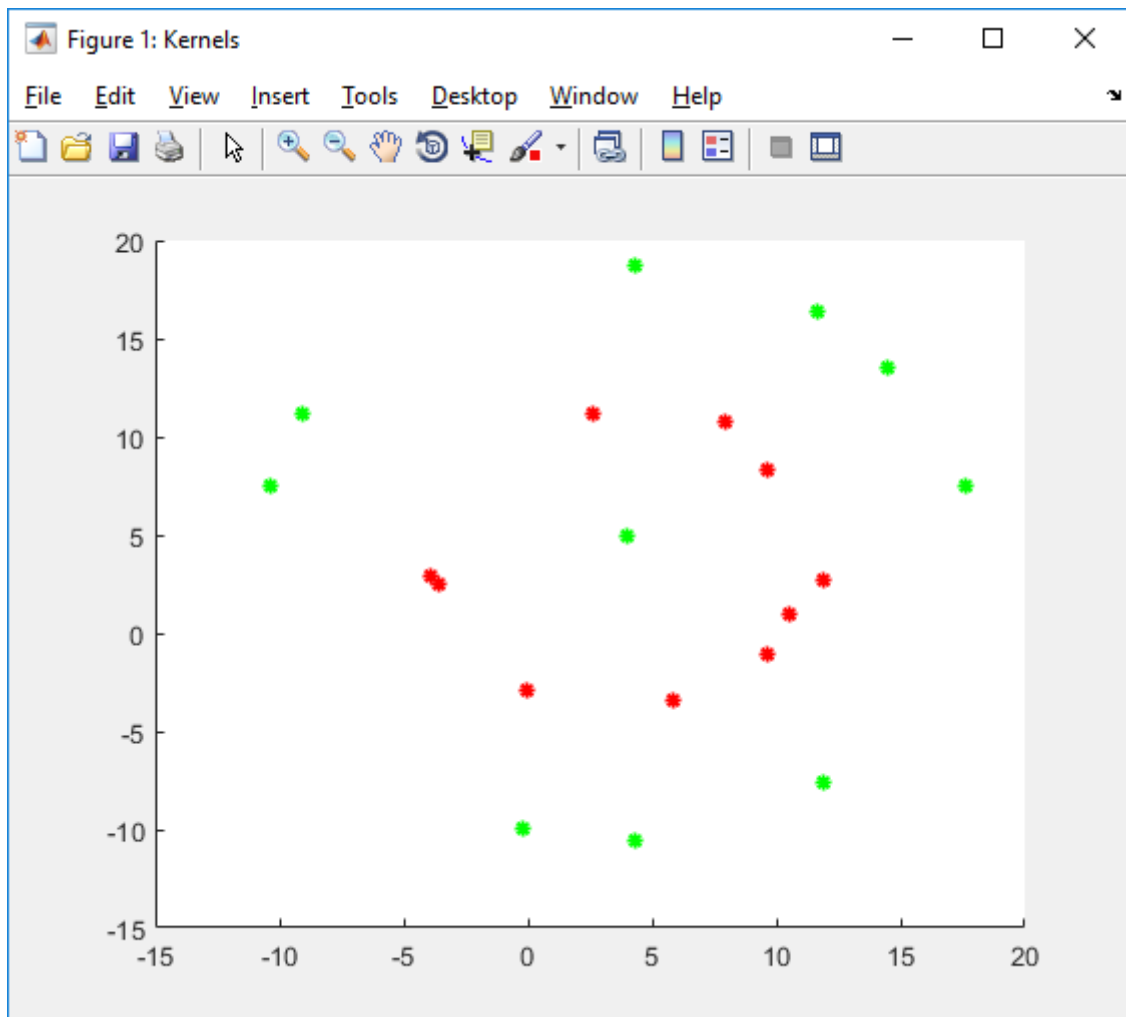
$$\begin{aligned}x_1 &= r \cos(\phi_1) \\x_2 &= r \sin(\phi_1) \cos(\phi_2) \\x_3 &= r \sin(\phi_1) \sin(\phi_2) \cos(\phi_3) \\&\vdots \\x_{n-1} &= r \sin(\phi_1) \cdots \sin(\phi_{n-2}) \cos(\phi_{n-1}) \\x_n &= r \sin(\phi_1) \cdots \sin(\phi_{n-2}) \sin(\phi_{n-1}).\end{aligned}$$

Όπου  $\phi_1, \phi_2, \dots, \phi_{n-2}$  ανήκουν στο διάστημα  $[0, \pi]$  και η γωνία  $\phi_{n-1}$  στο  $[0, 2\pi)$ .

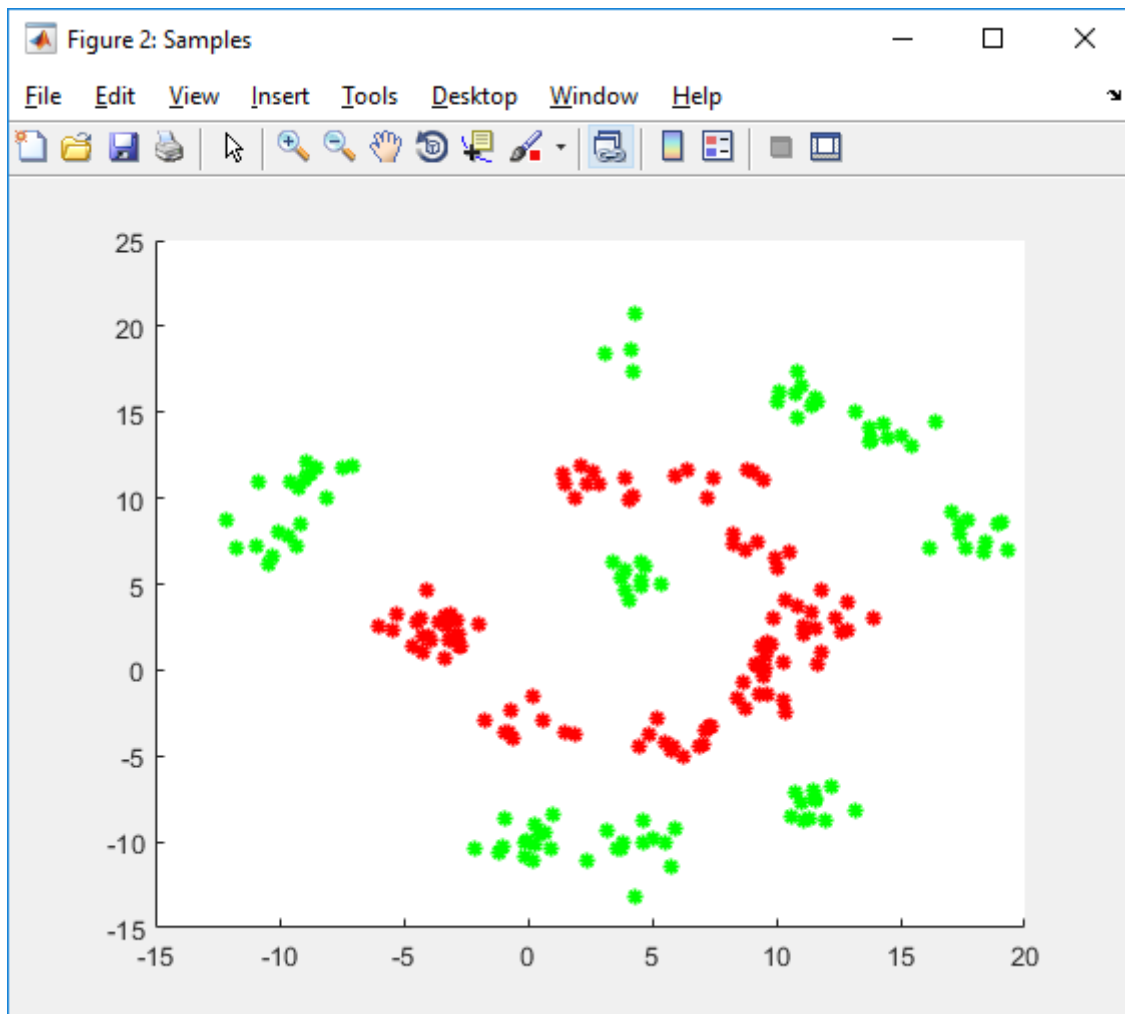
Η συνάρτηση CreateKernels υπολογίζει τις παραπάνω ποσότητες δημιουργώντας τυχαίες γωνίες και στη συνέχεια προσθέτοντας το παραπάνω διάνυσμα στον κεντρικό πυρήνα προκύπτει ένα σημείο επάνω στην υπέρ-σφαίρα με κέντρο τον κεντρικό πυρήνα και ακτίνα r. Ενδεικτικά παραθέτουμε τα γραφήματα της κατανομής των πυρήνων και των δεδομένων όταν έχουμε 10 πυρήνες σε κάθε μια από τις τάξεις, και ο αριθμός των δεδομένων που δημιουργείται για την κάθε τάξη είναι 100. Παραθέτονται αντίστοιχα τα γραφήματα στις περιπτώσεις των 2 και 3 διαστάσεων. Για τα προβλήματα μεγαλύτερων διαστάσεων η γραφική αναπαράσταση τους βασίζεται στη στατιστική διαδικασία PCA (Principal Component Analysis).

Να σημειωθεί πως η διαδικασία παραγωγής των πυρήνων έγινε μέσω πολικών συντεταγμένων για να αντιμετωπίσουμε το πρόβλημα της διαστασιμότητας (Curse of Dimensionality), στο οποίο καθώς μεταβαίνουμε σε μεγαλύτερες διαστάσεις οι αποστάσεις μεταξύ των σημείων αυξάνονται και έτσι το πρόβλημα ταξινόμησης ουσιαστικά γίνεται πιο εύκολο. Έτσι χρησιμοποιώντας πολικές συντεταγμένες η απόσταση καθορίζεται από την ακτίνα του κύκλου και όχι από τα ίδια τα σημεία, συνεπώς μπορούμε να την παραμετροποιήσουμε όπως εμείς χρειαζόμαστε.

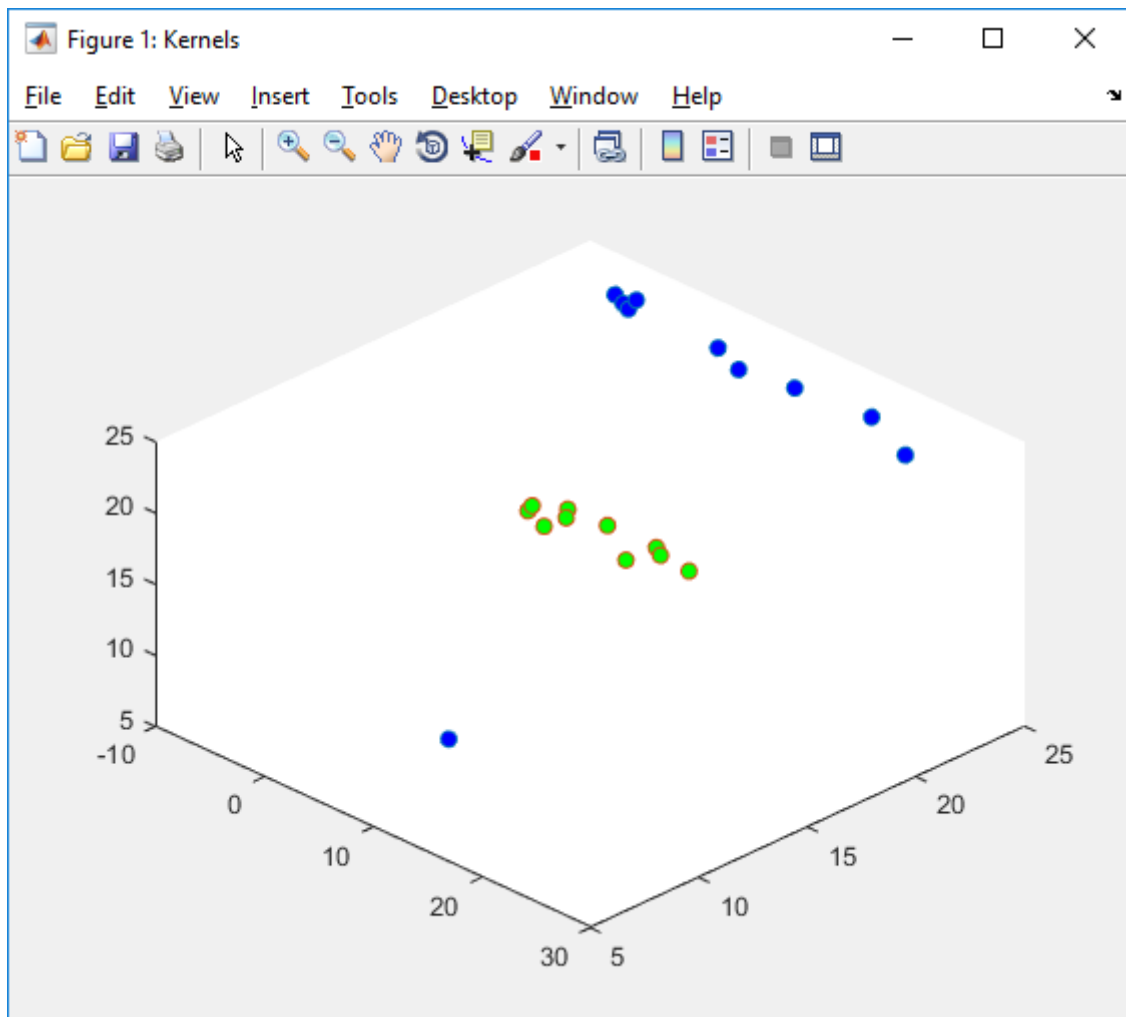




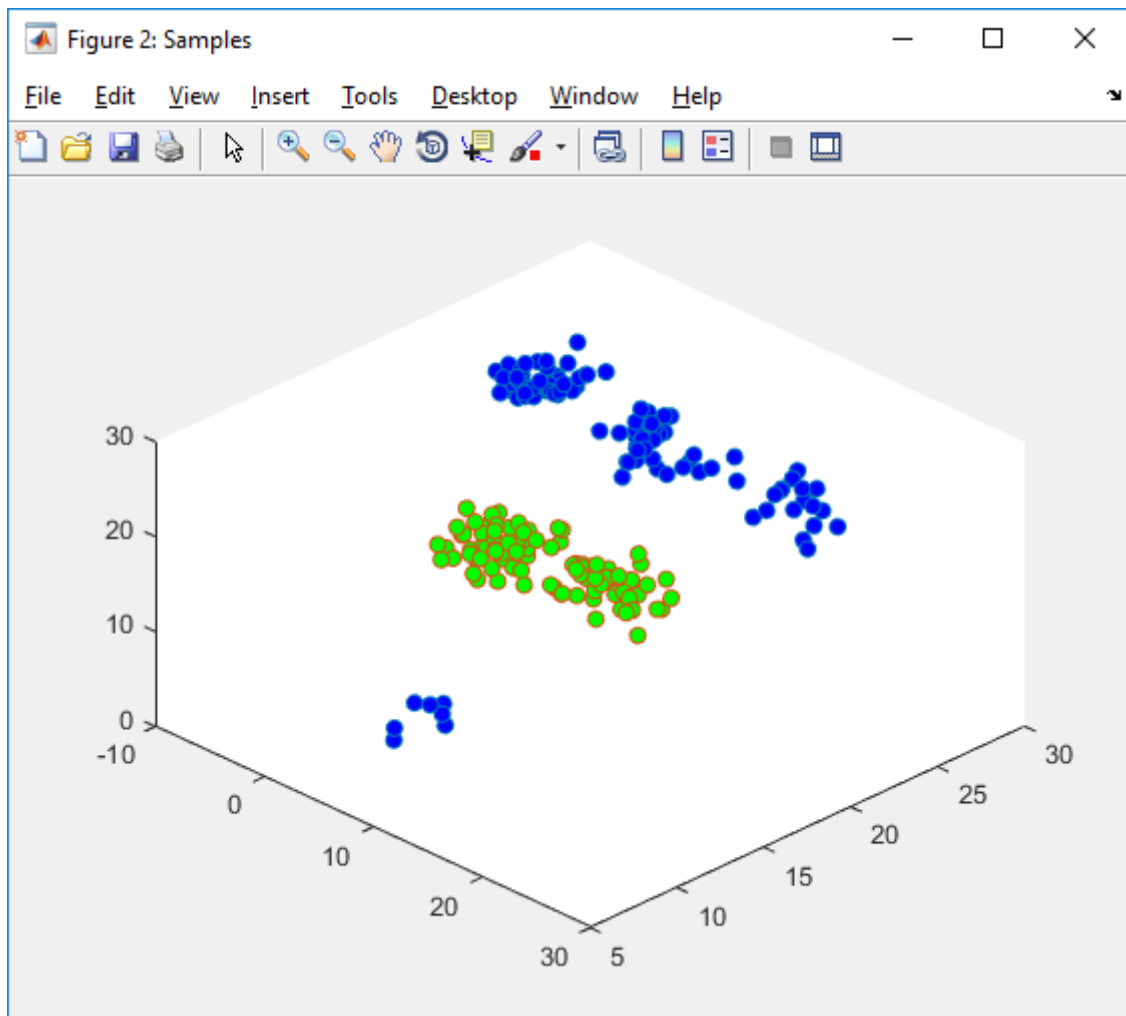
Εικόνα 4.1 – Δημιουργία των πυρήνων σε 2 διαστάσεις.



Εικόνα 4.2 – Δημιουργία δεδομένων σε 2 διαστάσεις.



Εικόνα 4.3 – Δημιουργία των πυρήνων σε 3 διαστάσεις.



**Εικόνα 4.4 – Δημιουργία δεδομένων σε 3 διαστάσεις.**

Στις επόμενες υπό-ενότητες παραθέτουμε τον κώδικα για τις συναρτήσεις που δημιουργούν τα δεδομένα τα οποία αναπαρίστανται στα παραπάνω γραφήματα.

#### 4.1 Η συνάρτηση CreateData.

Εδώ παραθέτουμε τον κώδικα της συνάρτησης CreateData.

```
function [SelfSamples, NonSelfSamples] = CreateData(Dimensionality, SelfKernelsNum,
NonSelfKernelsNum, SelfSamplesNum, NonSelfSamplesNum)
```

```
% Define the number of training and testing patterns generated from both
% classes and the parameters of the multivariate gaussian distributions.
```

```
Sigma = eye(Dimensionality);
SelfKernels = zeros(SelfKernelsNum, Dimensionality);
```

```
% Define the center of the first SelfSample Kernel.
```

```
for i = 1:Dimensionality
    SelfKernels(1,i) = randi([1 9],1 ,1);
end
```

```
% Set the distances for the rest of the Self Kernels.
```

```
SelfKernelsRmax = 15;
SelfKernelsRmin = 14;
SelfKernelsR = SelfKernelsRmin+(SelfKernelsRmax-SelfKernelsRmin)
    *rand(SelfKernelsNum-1,1);
```

```
% Set the distances for the Non-Self Kernels.
```

```
NonSelfKernelsRmax = 5;
NonSelfKernelsRmin = 4;
NonSelfKernelsR = NonSelfKernelsRmin+(NonSelfKernelsRmax-
    NonSelfKernelsRmin)*rand(NonSelfKernelsNum,1);
```

```
% Generate the rest of the Kernels
```

```
RestSelfKernels = CreateKernels(Dimensionality, SelfKernelsNum-1,
    SelfKernelsR, SelfKernels(1,:));
SelfKernels = [SelfKernels(1,:); RestSelfKernels];
NonSelfKernels = CreateKernels(Dimensionality, NonSelfKernelsNum,
    NonSelfKernelsR, SelfKernels(1,:));
```

```
% Generate patterns for Self Class.
```

```
SelfDataIndexes = randi([1 SelfKernelsNum], SelfSamplesNum, 1);
SelfSamples = zeros(SelfSamplesNum, Dimensionality);
EachSelfSampleNum = zeros(SelfKernelsNum,1);
EachSelfSampleNum(1) = length(find(SelfDataIndexes==1));
SelfSamples(1:EachSelfSampleNum(1),:) =
    mvnrnd(SelfKernels(1,:),Sigma, EachSelfSampleNum(1));
s = EachSelfSampleNum(1)+1;
for i = 2:SelfKernelsNum
    EachSelfSampleNum(i) = length(find(SelfDataIndexes==i));
    SelfSamples(s:s + EachSelfSampleNum(i)-1,:) =
        mvnrnd(SelfKernels(i,:),Sigma, EachSelfSampleNum(i));
    s = s + EachSelfSampleNum(i);
end
```

```
% Generate patterns for Non-Self Class.
```

```
NonSelfDataIndexes = randi([1 NonSelfKernelsNum], NonSelfSamplesNum, 1);
NonSelfSamples = zeros(NonSelfSamplesNum, Dimensionality);
EachNonSelfSampleNum = zeros(NonSelfKernelsNum,1);
```

```

EachNonSelfSampleNum(1) = length(find(NonSelfDataIndexes==1));
NonSelfSamples(1:EachNonSelfSampleNum(1),:) =
    mvnrnd(NonSelfKernels(1,:),Sigma, EachNonSelfSampleNum(1));
s = EachNonSelfSampleNum(1)+1;
for i = 1:NonSelfKernelsNum
    EachNonSelfSampleNum(i) = length(find(NonSelfDataIndexes==i));
    NonSelfSamples(s:s + EachNonSelfSampleNum(i)-1,:) =
        mvnrnd(NonSelfKernels(i,:),Sigma, EachNonSelfSampleNum(i));
    s = s + EachSelfSampleNum(i);
end

% Plot the created Kernels.
plotData(Dimensionality,SelfKernels,NonSelfKernels,'Kernels');

% Plot the created Samples.
plotData(Dimensionality,SelfSamples,NonSelfSamples,'Samples');

```

## 4.2 Η συνάρτηση CreateKernels.

Εδώ παραθέτουμε τον κώδικα της συνάρτησης CreateData.

```

function [Kernels] = CreateKernels(Dimensionality, KernelsNum, R,
    CentralKernel)

% Create the angles needed in order to compute
% the new Kernels using spherical coordinates.
Rotations = pi*rand(KernelsNum, Dimensionality-2);
Rotations = [Rotations, 2*pi*rand(KernelsNum, 1)];

% Initialization of the Kernel Array
Kernels = zeros(KernelsNum, Dimensionality);

% For Each Kernel
for k = 1:KernelsNum
    % Compute the quantity needed to add to each coordinate
    % to get the new Kernel.
    for i = 1:(Dimensionality-1)
        P = R(k);
        for j = i:1
            if j==i
                P = P*cos(Rotations(k,j));
            else
                P = P*sin(Rotations(k,j));
            end
        end
        Kernels(k,i) = CentralKernel(1,i) + P;
    end
    P = R(k);
    for j = (Dimensionality-1):1
        P = P*sin(Rotations(k,j));
    end
    % Add Vector P to the Central Kernel to get the new Kernel.
    Kernels(k,Dimensionality) = CentralKernel(1,i) + P;
end

```

## 5 Πειράματα και Αποτελέσματα

Τα πειράματα που θα εκτελέσουμε σε αυτή την εργασία αφορούν την παροχή ενός αριθμού από συνδυασμούς δεδομένων τα οποία δημιουργούνται όπως περιγράψαμε παραπάνω. Υπενθυμίζουμε ότι τα δεδομένα μας δημιουργούνται βάσει :

- Διαστάσεων (Dimensionality).
- Πλήθος πυρήνων του συνόλου εαυτός (Number of Kernels).
- Πλήθος πυρήνων του συνόλου μη εαυτός (Number of Non-Self Kernels).
- Πλήθος δεδομένων του συνόλου εαυτός (Number of Self Samples).
- Πλήθος δεδομένων του συνόλου μη εαυτός (Number of Non-Self Samples).
- Αποστάσεις μεταξύ των πυρήνων (SelfKernelsR, NonSelfKernelsR)

Επιπλέον, σε κάθε περίπτωση γίνεται και μια επιλογή της καλύτερης απόδοσης του αλγόριθμου V-Detector ελέγχοντας έναν αριθμό από συνδυασμούς των παραμέτρων που παρέχονται ως ορίσματα στον αλγόριθμο. Τα ορίσματα αυτά είναι :

- **Rself**

Η ακτίνα των σημείων του συνόλου εαυτός τα οποία ορίζουν τον χώρο που καλύπτει το κάθε πρότυπο στο σύνολο εαυτός.

- **Co**

Η τιμή που ορίζουμε ως στόχο για να θεωρήσουμε ότι ο χώρος που ορίζει το σύνολο μη εαυτός έχει καλυφθεί.

- **a**

Η τιμή αυτή αντιστοιχεί στο ποσοστό κάλυψης του στόχου. Χρησιμοποιείται για τον προσδιορισμό των προσπαθειών δημιουργίας ενός ανιχνευτή όταν αυτός δεν γίνεται αποδεκτός και πρέπει να δημιουργηθεί άλλος στη θέση του.

- **DeltaTmax**

Η τιμή αυτή προστίθεται σε μια τιμή αρχικοποιημένη σύμφωνα με το ποσοστό κάλυψης που ζητάμε και έτσι προκύπτει ο μέγιστος αριθμός ανιχνευτών.

Πιο συγκεκριμένα, οι περιπτώσεις που θα εξετάσουμε ως προς τα δεδομένα εισόδου αντιστοιχούν στους συνδυασμούς των στοιχείων των παρακάτω συνόλων, εάν επιλεγεί ένα στοιχείο από το καθένα.

Dimensionality = [2, 3, 100]

SelfKernelsNum = [5, 10, 25]

NonSelfKernelsNum = [5, 10, 25]

SelfSamplesNum = [100, 500, 1500]

NonSelfSamplesNum = [100, 500, 1500]

Καθώς το κάθε ένα από αυτά τα σύνολα έχει πληθάρημο 3 το σύνολο των συνδυασμών προς εξέταση θα είναι  $3^5 = 243$ . Σε κάθε έναν από αυτούς τους συνδυασμούς θα εφαρμοστεί ο αλγόριθμος V-Detector αρκετές φορές χρησιμοποιώντας συνδυασμούς των παραμέτρων που λαμβάνει ως ορίσματα. Η λύση που θα αποθηκεύουμε θα είναι αυτή η οποία προσφέρει τον καλύτερο μέσο όρο ως προς τα ποσοστά επιτυχίας αναγνώρισης των στοιχείων του συνόλου εαυτός και το αντίστοιχο ποσοστό επιτυχίας του συνόλου μη εαυτός. Οι συνδυασμοί των παραμέτρων αυτών προκύπτουν από τον παρακάτω πίνακα :

	<b>Rself</b>	<b>Co</b>	<b>a</b>	<b>DeltaTmax</b>
<b>max</b>	<b>0.1</b>	<b>0.999</b>	<b>0.0001</b>	<b>200</b>
<b>min</b>	<b>0.01</b>	<b>0.99</b>	<b>0.00002</b>	<b>100</b>

step	0.01	0.001	0.00002	100
------	------	-------	---------	-----

Οι συνδυασμοί των παραμέτρων του V-Detector που προκύπτουν σύμφωνα με τα στοιχεία του παραπάνω πίνακα αντιστοιχούν σε 10 επιλογές για το Rself, 10 επιλογές για το C<sub>0</sub>, 5 επιλογές για το a και 2 επιλογές για το DeltaTmax. Συνεπώς, προκύπτει ότι συνολικά έχουμε 10\*10\*5\*2 = 1000 δυνατούς συνδυασμούς για τις παραμέτρους του αλγόριθμου.

Έτσι επιλέγοντας ένα συνδυασμό από ακτίνες, τοποθετούμε τους πυρήνες όσο κοντά ή μακριά θέλουμε, και συνεπώς να ρυθμίζουμε την δυσκολία του προβλήματος προς αντιμετώπιση. Για κάθε ένα από αυτά τα πειράματα ουσιαστικά ελέγχουμε 243\*1000 = 243.000 διαφορετικές περιπτώσεις εκτέλεσης του αλγόριθμου.

Η διαδικασία διεξαγωγής των πειραμάτων μπορεί να περιγραφεί από τα παρακάτω βήματα:

- Το πείραμα επαναλαμβάνεται για τον κάθε ένα από τους 243 συνδυασμούς που ορίζουν τη δομή των δεδομένων του συνόλου εαυτός και μη εαυτός.
- Για κάθε μια από αυτές τις περιπτώσεις εκτελούνται 1000 διαφορετικές επαναλήψεις του V-Detector η κάθε μια με έναν από τους δυνατούς συνδυασμούς των παραμέτρων εισόδου.
- Από αυτές τις 1000 περιπτώσεις δεχόμαστε μόνο την βέλτιστη η οποία προκύπτει ως εκείνη με τον μεγαλύτερο μέσο όρο των ποσοστών επιτυχούς αναγνώρισης των συνόλων εαυτός και μη εαυτός, σύμφωνα με τον παρακάτω τύπο:

$$\sum_{i=1}^2 \frac{f_i x_i}{n}$$

**Όπου:**

- $x_1$  το ποσοστό επιτυχημένης αναγνώρισης των testing στοιχείων του συνόλου εαυτός.
- $x_2$  το αντίστοιχο ποσοστό αναγνώρισης των στοιχείων του συνόλου μη εαυτός
- $f_i$  το πλήθος των στοιχείων της κλάσης  $i$ .
- $n$  το σύνολο των στοιχείων των 2 κλάσεων.
- Αποθηκεύουμε τον καλύτερο μέσο όρο της κάθε περίπτωσης μαζί με τις διάφορες παραμέτρους που οδήγησαν σε αυτό το αποτέλεσμα σε έναν πίνακα 243x10. Οι στήλες αντιστοιχούν με την σειρά στις εξής ποσότητες:
  - Διαστάσεις
  - Πυρήνες Συνόλου Εαυτός
  - Πυρήνες Συνόλου Μη Εαυτός
  - Αριθμός Δεδομένων Συνόλου Εαυτός
  - Αριθμός Δεδομένων Συνόλου Μη Εαυτός
  - Rself
  - C<sub>0</sub>
  - a
  - DeltaTmax
  - Ποσοστό Επιτυχίας

Η παραπάνω διαδικασία επαναλήφθηκε 3 φορές ως πείραμα με στόχο να παρατηρήσουμε την συμπεριφορά του αλγόριθμου και τις διαφοροποιήσεις στα αποτελέσματα που προκύπτουν. Στο πρώτο πείραμα τοποθετήσαμε τους πυρήνες όσο το δυνατόν πιο διακριτά γίνεται ενώ στη συνέχεια αρχίσαμε να πλησιάζουμε τους πυρήνες ώστε το πρόβλημα να γίνεται πιο δύσκολο. Παρακάτω μπορούμε να δούμε για το κάθε πείραμα τις αποστάσεις των πυρήνων καθώς και τα αποτελέσματα που παρήχθησαν από αυτή τη διαδικασία.



## 5.1 Πείραμα 1<sup>ο</sup>

Στο πρώτο πείραμα προσπαθήσαμε να κρατήσουμε τα δεδομένα αρκετά διαχωρίσιμα και έτσι ορίσαμε τις αποστάσεις των πυρήνων του συνόλου μη εαυτός, από τον κεντρικό πυρήνα του συνόλου εαυτός, να παίρνουν τιμές στο διάστημα [7, 8] ενώ για τους αντίστοιχους πυρήνες του συνόλου εαυτός από το διάστημα [14, 15]. Στον παρακάτω πίνακα μπορείτε να δείτε όλους τους συνδυασμούς των παραμέτρων δημιουργίας των δεδομένων καθώς και τον συνδυασμό των παραμέτρων του αλγόριθμου V-Detector που μας παρείχαν το μεγαλύτερο ποσοστό επιτυχίας.

Διαστάσεις	Πυρήνες Συνόλου Εαυτός	Πυρήνες Συνόλου Μη Εαυτός	Αριθμός Δεδομένων Συνόλου Εαυτός	Αριθμός Δεδομένων Συνόλου Μη Εαυτός	Rself	C <sub>0</sub>	a	DeltaTmax	Ποσοστό Επιτυχίας
2	5	5	100	100	0.02	0.987	0.00002	100	0.829091
2	5	5	100	500	0.02	0.987	0.0001	100	0.999412
2	5	5	100	1500	0.02	0.987	0.00002	100	0.999139
2	5	5	500	100	0.02	0.987	0.00002	100	0.983333
2	5	5	500	500	0.02	0.987	0.00006	200	0.994727
2	5	5	500	1500	0.02	0.991	0.00002	100	0.974065
2	5	5	1500	100	0.02	0.987	0.00002	100	0.9688
2	5	5	1500	500	0.02	0.987	0.0001	200	0.734462
2	5	5	1500	1500	0.02	0.987	0.00006	100	0.985455
2	5	10	100	100	0.02	0.987	0.00002	100	0.936364
2	5	10	100	500	0.02	0.987	0.00002	100	0.818431
2	5	10	100	1500	0.02	0.987	0.00006	200	1
2	5	10	500	100	0.02	0.987	0.00002	100	0.922
2	5	10	500	500	0.02	0.987	0.00002	200	0.905091
2	5	10	500	1500	0.02	0.987	0.00002	200	0.969613
2	5	10	1500	100	0.02	0.987	0.00002	200	0.984
2	5	10	1500	500	0.02	0.987	0.00006	100	0.964154
2	5	10	1500	1500	0.02	0.987	0.00002	200	0.970364
2	5	25	100	100	0.02	0.987	0.0001	100	0.969091
2	5	25	100	500	0.02	0.987	0.00002	100	1
2	5	25	100	1500	0.02	0.987	0.00002	200	0.883046
2	5	25	500	100	0.02	0.987	0.00002	100	0.993333
2	5	25	500	500	0.02	0.987	0.00002	100	0.954909
2	5	25	500	1500	0.02	0.987	0.00002	200	0.977097
2	5	25	1500	100	0.02	0.987	0.00002	200	0.9844
2	5	25	1500	500	0.02	0.987	0.00006	100	0.936154
2	5	25	1500	1500	0.02	0.987	0.0001	100	0.978788
2	10	5	100	100	0.02	0.987	0.00002	100	1
2	10	5	100	500	0.02	0.987	0.00006	100	0.908235
2	10	5	100	1500	0.02	0.987	0.00002	100	0.999934
2	10	5	500	100	0.02	0.987	0.00006	200	0.802

2	10	5	500	500	0.02	0.987	0.00006	100	0.775636
2	10	5	500	1500	0.02	0.987	0.00006	100	0.828774
2	10	5	1500	100	0.02	0.987	0.00006	100	0.9896
2	10	5	1500	500	0.02	0.987	0.00002	100	0.777385
2	10	5	1500	1500	0.02	0.987	0.00002	100	0.918727
2	10	10	100	100	0.02	0.987	0.00006	100	0.982727
2	10	10	100	500	0.02	0.987	0.0001	200	0.971961
2	10	10	100	1500	0.02	0.987	0.00002	200	0.941192
2	10	10	500	100	0.02	0.987	0.00002	200	0.972
2	10	10	500	500	0.02	0.987	0.00002	100	0.985273
2	10	10	500	1500	0.02	0.987	0.00006	100	0.814129
2	10	10	1500	100	0.02	0.987	0.00002	200	0.898
2	10	10	1500	500	0.02	0.987	0.00006	200	0.747538
2	10	10	1500	1500	0.02	0.987	0.00006	100	0.900848
2	10	25	100	100	0.02	0.987	0.00002	200	0.997273
2	10	25	100	500	0.02	0.987	0.0001	100	0.989608
2	10	25	100	1500	0.02	0.987	0.0001	100	0.951325
2	10	25	500	100	0.02	0.987	0.00002	200	0.974
2	10	25	500	500	0.02	0.987	0.00006	100	0.984182
2	10	25	500	1500	0.02	0.987	0.00002	200	0.785226
2	10	25	1500	100	0.02	0.987	0.00006	200	0.8928
2	10	25	1500	500	0.02	0.987	0.00002	100	0.802154
2	10	25	1500	1500	0.02	0.987	0.00006	200	0.729939
2	25	5	100	100	0.02	0.987	0.00002	100	0.989091
2	25	5	100	500	0.02	0.987	0.00006	200	0.99902
2	25	5	100	1500	0.02	0.991	0.00002	200	0.979139
2	25	5	500	100	0.02	0.987	0.00002	100	0.984667
2	25	5	500	500	0.02	0.987	0.00002	200	0.941455
2	25	5	500	1500	0.02	0.987	0.0001	200	0.802194
2	25	5	1500	100	0.02	0.987	0.00002	200	0.9728
2	25	5	1500	500	0.02	0.991	0.00002	100	0.931538
2	25	5	1500	1500	0.02	0.987	0.00002	200	0.95097
2	25	10	100	100	0.02	0.987	0.00006	200	0.99
2	25	10	100	500	0.02	0.987	0.0001	200	0.896078
2	25	10	100	1500	0.02	0.991	0.00006	100	0.995099
2	25	10	500	100	0.02	0.991	0.00006	100	0.907333
2	25	10	500	500	0.02	0.991	0.00006	200	0.737091
2	25	10	500	1500	0.02	0.987	0.00006	100	0.827097
2	25	10	1500	100	0.02	0.987	0.00006	100	0.9588
2	25	10	1500	500	0.02	0.987	0.00002	200	0.861077

2	25	10	1500	1500	0.02	0.987	0.00006	200	0.772182
2	25	25	100	100	0.02	0.991	0.00006	200	0.994545
2	25	25	100	500	0.02	0.987	0.00002	100	0.87549
2	25	25	100	1500	0.02	0.987	0.0001	200	0.992517
2	25	25	500	100	0.02	0.987	0.00002	100	0.926
2	25	25	500	500	0.02	0.987	0.00002	100	0.781455
2	25	25	500	1500	0.02	0.987	0.00002	200	0.853355
2	25	25	1500	100	0.02	0.987	0.00006	200	0.9196
2	25	25	1500	500	0.02	0.987	0.00006	200	0.952308
2	25	25	1500	1500	0.02	0.987	0.00002	200	0.918182
3	5	5	100	100	0.02	0.987	0.00002	100	1
3	5	5	100	500	0.02	0.987	0.00002	200	0.998627
3	5	5	100	1500	0.02	0.987	0.00002	100	0.999536
3	5	5	500	100	0.02	0.987	0.00002	200	0.983333
3	5	5	500	500	0.02	0.987	0.00002	200	0.996909
3	5	5	500	1500	0.02	0.987	0.00002	200	0.999613
3	5	5	1500	100	0.02	0.987	0.00002	200	0.9832
3	5	5	1500	500	0.02	0.987	0.00002	100	0.999846
3	5	5	1500	1500	0.02	0.987	0.00002	100	0.99697
3	5	10	100	100	0.02	0.987	0.00002	100	1
3	5	10	100	500	0.02	0.987	0.00002	100	0.999804
3	5	10	100	1500	0.02	0.987	0.00002	200	1
3	5	10	500	100	0.02	0.987	0.00002	100	1
3	5	10	500	500	0.02	0.987	0.00002	200	0.992
3	5	10	500	1500	0.02	0.987	0.00002	100	0.999935
3	5	10	1500	100	0.02	0.987	0.00002	100	1
3	5	10	1500	500	0.02	0.987	0.00002	100	0.999385
3	5	10	1500	1500	0.02	0.987	0.00002	100	0.999636
3	5	25	100	100	0.02	0.987	0.00002	100	0.991818
3	5	25	100	500	0.02	0.987	0.00002	100	0.998235
3	5	25	100	1500	0.02	0.987	0.00002	100	0.999934
3	5	25	500	100	0.02	0.987	0.00002	100	1
3	5	25	500	500	0.02	0.987	0.00002	100	0.999455
3	5	25	500	1500	0.02	0.987	0.00002	200	0.996968
3	5	25	1500	100	0.02	0.987	0.00002	100	1
3	5	25	1500	500	0.02	0.987	0.00002	100	0.998308
3	5	25	1500	1500	0.02	0.987	0.00006	100	0.997758
3	10	5	100	100	0.02	0.987	0.00006	100	0.993636
3	10	5	100	500	0.02	0.987	0.00002	100	1
3	10	5	100	1500	0.02	0.987	0.00006	100	0.999868

3	10	5	500	100	0.02	0.987	0.00002	100	1
3	10	5	500	500	0.02	0.987	0.00002	100	1
3	10	5	500	1500	0.02	0.987	0.00002	200	0.999742
3	10	5	1500	100	0.02	0.987	0.00002	200	1
3	10	5	1500	500	0.02	0.987	0.00002	200	0.992923
3	10	5	1500	1500	0.02	0.987	0.0001	100	0.996182
3	10	10	100	100	0.02	0.987	0.00002	200	1
3	10	10	100	500	0.02	0.987	0.00002	100	1
3	10	10	100	1500	0.02	0.987	0.00002	100	1
3	10	10	500	100	0.02	0.987	0.00002	100	1
3	10	10	500	500	0.02	0.987	0.00002	100	0.999091
3	10	10	500	1500	0.02	0.987	0.00002	100	0.999871
3	10	10	1500	100	0.02	0.987	0.00002	100	1
3	10	10	1500	500	0.02	0.987	0.00002	200	0.998769
3	10	10	1500	1500	0.02	0.987	0.00006	200	0.997818
3	10	25	100	100	0.02	0.987	0.00002	100	1
3	10	25	100	500	0.02	0.987	0.00002	200	0.997255
3	10	25	100	1500	0.02	0.987	0.00002	100	1
3	10	25	500	100	0.02	0.987	0.00002	100	1
3	10	25	500	500	0.02	0.987	0.00002	100	1
3	10	25	500	1500	0.02	0.987	0.00002	100	0.998903
3	10	25	1500	100	0.02	0.987	0.00002	100	1
3	10	25	1500	500	0.02	0.987	0.00002	200	0.996923
3	10	25	1500	1500	0.02	0.987	0.00002	100	0.989091
3	25	5	100	100	0.02	0.987	0.00002	200	0.999091
3	25	5	100	500	0.02	0.987	0.00002	100	1
3	25	5	100	1500	0.02	0.987	0.00002	200	0.999868
3	25	5	500	100	0.02	0.987	0.00002	100	1
3	25	5	500	500	0.02	0.987	0.00002	100	0.999818
3	25	5	500	1500	0.02	0.987	0.00002	100	1
3	25	5	1500	100	0.02	0.987	0.00006	100	0.9988
3	25	5	1500	500	0.02	0.987	0.00002	100	0.998462
3	25	5	1500	1500	0.02	0.987	0.00002	100	1
3	25	10	100	100	0.02	0.987	0.00002	100	0.999091
3	25	10	100	500	0.02	0.987	0.00002	100	1
3	25	10	100	1500	0.02	0.987	0.00002	100	0.999934
3	25	10	500	100	0.02	0.987	0.00002	100	1
3	25	10	500	500	0.02	0.987	0.00002	100	1
3	25	10	500	1500	0.02	0.987	0.00002	100	0.999484
3	25	10	1500	100	0.02	0.987	0.00002	200	0.9972

3	25	10	1500	500	0.02	0.987	0.00002	100	0.998769
3	25	10	1500	1500	0.02	0.987	0.00002	100	1
3	25	25	100	100	0.02	0.987	0.00002	100	1
3	25	25	100	500	0.02	0.987	0.00002	100	0.999804
3	25	25	100	1500	0.02	0.987	0.00002	100	0.999934
3	25	25	500	100	0.02	0.987	0.00002	100	0.999333
3	25	25	500	500	0.02	0.987	0.00002	100	1
3	25	25	500	1500	0.02	0.987	0.00002	200	0.995935
3	25	25	1500	100	0.02	0.987	0.00002	100	1
3	25	25	1500	500	0.02	0.987	0.00002	100	1
3	25	25	1500	1500	0.02	0.987	0.00002	200	0.99697
100	5	5	100	100	0.02	0.987	0.00002	100	1
100	5	5	100	500	0.02	0.987	0.00002	100	1
100	5	5	100	1500	0.02	0.987	0.00002	100	1
100	5	5	500	100	0.02	0.987	0.00002	100	1
100	5	5	500	500	0.02	0.987	0.00002	100	1
100	5	5	500	1500	0.02	0.987	0.00002	100	1
100	5	5	1500	100	0.02	0.987	0.00002	100	1
100	5	5	1500	500	0.02	0.987	0.00002	100	1
100	5	5	1500	1500	0.02	0.987	0.00002	100	1
100	5	10	100	100	0.02	0.987	0.00002	100	1
100	5	10	100	500	0.02	0.987	0.00002	100	1
100	5	10	100	1500	0.02	0.987	0.00002	100	1
100	5	10	500	100	0.02	0.987	0.00002	100	1
100	5	10	500	500	0.02	0.987	0.00002	100	1
100	5	10	500	1500	0.02	0.987	0.00002	100	1
100	5	10	1500	100	0.02	0.987	0.00002	100	1
100	5	10	1500	500	0.02	0.987	0.00002	100	1
100	5	10	1500	1500	0.02	0.987	0.00002	100	1
100	5	25	100	100	0.02	0.987	0.00002	100	1
100	5	25	100	500	0.02	0.987	0.00002	100	1
100	5	25	100	1500	0.02	0.987	0.00002	100	1
100	5	25	500	100	0.02	0.987	0.00002	100	1
100	5	25	500	500	0.02	0.987	0.00002	100	1
100	5	25	500	1500	0.02	0.987	0.00002	100	1
100	5	25	1500	100	0.02	0.987	0.00002	100	1
100	5	25	1500	500	0.02	0.987	0.00002	100	1
100	5	25	1500	1500	0.02	0.987	0.00002	100	1
100	10	5	100	100	0.02	0.987	0.00002	100	1
100	10	5	100	500	0.02	0.987	0.00002	100	1

100	10	5	100	1500	0.02	0.987	0.00002	100	1
100	10	5	500	100	0.02	0.987	0.00002	100	1
100	10	5	500	500	0.02	0.987	0.00002	100	1
100	10	5	500	1500	0.02	0.987	0.00002	100	1
100	10	5	1500	100	0.02	0.987	0.00002	100	1
100	10	5	1500	500	0.02	0.987	0.00002	100	1
100	10	5	1500	1500	0.02	0.987	0.00002	100	1
100	10	10	100	100	0.02	0.987	0.00002	100	1
100	10	10	100	500	0.02	0.987	0.00002	100	1
100	10	10	100	1500	0.02	0.987	0.00002	100	1
100	10	10	500	100	0.02	0.987	0.00002	100	1
100	10	10	500	500	0.02	0.987	0.00002	100	1
100	10	10	500	1500	0.02	0.987	0.00002	100	1
100	10	10	1500	100	0.02	0.987	0.00002	100	1
100	10	10	1500	500	0.02	0.987	0.00002	100	1
100	10	10	1500	1500	0.02	0.987	0.00002	100	1
100	10	25	100	100	0.02	0.987	0.00002	100	1
100	10	25	100	500	0.02	0.987	0.00002	100	1
100	10	25	100	1500	0.02	0.987	0.00002	100	1
100	10	25	500	100	0.02	0.987	0.00002	100	1
100	10	25	500	500	0.02	0.987	0.00002	100	1
100	10	25	500	1500	0.02	0.987	0.00002	100	1
100	10	25	1500	100	0.02	0.987	0.00002	100	1
100	10	25	1500	500	0.02	0.987	0.00002	100	1
100	10	25	1500	1500	0.02	0.987	0.00002	100	1
100	25	5	100	100	0.02	0.987	0.00002	100	1
100	25	5	100	500	0.02	0.987	0.00002	100	1
100	25	5	100	1500	0.02	0.987	0.00002	100	1
100	25	5	500	100	0.02	0.987	0.00002	100	1
100	25	5	500	500	0.02	0.987	0.00002	100	1
100	25	5	500	1500	0.02	0.987	0.00002	100	1
100	25	5	1500	100	0.02	0.987	0.00002	100	1
100	25	5	1500	500	0.02	0.987	0.00002	100	1
100	25	5	1500	1500	0.02	0.987	0.00002	100	1
100	25	10	100	100	0.02	0.987	0.00002	100	1
100	25	10	100	500	0.02	0.987	0.00002	100	1
100	25	10	100	1500	0.02	0.987	0.00002	100	1
100	25	10	500	100	0.02	0.987	0.00002	100	1
100	25	10	500	500	0.02	0.987	0.00002	100	1
100	25	10	500	1500	0.02	0.987	0.00002	100	1

100	25	10	1500	100	0.02	0.987	0.00002	100	1
100	25	10	1500	500	0.02	0.987	0.00002	100	1
100	25	10	1500	1500	0.02	0.987	0.00002	100	1
100	25	25	100	100	0.02	0.987	0.00002	100	1
100	25	25	100	500	0.02	0.987	0.00002	100	1
100	25	25	100	1500	0.02	0.987	0.00002	100	1
100	25	25	500	100	0.02	0.987	0.00002	100	1
100	25	25	500	500	0.02	0.987	0.00002	100	1
100	25	25	500	1500	0.02	0.987	0.00002	100	1
100	25	25	1500	100	0.02	0.987	0.00002	100	1
100	25	25	1500	500	0.02	0.987	0.00002	100	1
100	25	25	1500	1500	0.02	0.987	0.00002	100	1

## 5.2 Πείραμα 2<sup>ο</sup>

Στο δεύτερο πείραμα προσπαθήσαμε να κρατήσουμε τα δεδομένα αρκετά διαχωρίσιμα και έτσι ορίσαμε τις αποστάσεις των πυρήνων του συνόλου μη εαυτός, από τον κεντρικό πυρήνα του συνόλου εαυτός, να παίρνουν τιμές στο διάστημα [5, 6] ενώ για τους αντίστοιχους πυρήνες του συνόλου εαυτός από το διάστημα [9, 10]. Στον παρακάτω πίνακα μπορείτε να δείτε όλους τους συνδυασμούς των παραμέτρων δημιουργίας των δεδομένων καθώς και τον συνδυασμό των παραμέτρων του αλγόριθμου V-Detector που μας παρείχαν το μεγαλύτερο ποσοστό επιτυχίας.

Διαστάσεις	Πυρήνες Συνόλου Εαυτός	Πυρήνες Συνόλου Μη Εαυτός	Αριθμός Δεδομένων Συνόλου Εαυτός	Αριθμός Δεδομένων Συνόλου Μη Εαυτός	Rself	C <sub>0</sub>	a	DeltaTmax	Ποσοστό Επιτυχίας
2	5	5	100	100	0.02	0.987	0.00006	100	0.755455
2	5	5	100	500	0.02	0.987	0.00002	200	0.911373
2	5	5	100	1500	0.02	0.987	0.00002	200	0.946424
2	5	5	500	100	0.02	0.987	0.00002	100	0.694667
2	5	5	500	500	0.02	0.987	0.00006	200	0.963636
2	5	5	500	1500	0.02	0.987	0.00002	200	0.637871
2	5	5	1500	100	0.02	0.987	0.00002	200	0.9412
2	5	5	1500	500	0.02	0.987	0.00002	100	0.702769
2	5	5	1500	1500	0.02	0.987	0.00002	100	0.631515
2	5	10	100	100	0.02	0.987	0.00006	200	0.754545
2	5	10	100	500	0.02	0.987	0.00006	100	0.974706
2	5	10	100	1500	0.02	0.987	0.00002	100	0.906159
2	5	10	500	100	0.02	0.987	0.00002	100	0.668667
2	5	10	500	500	0.02	0.987	0.00006	200	0.787636
2	5	10	500	1500	0.02	0.987	0.00006	100	0.894323
2	5	10	1500	100	0.02	0.987	0.00002	100	0.9004
2	5	10	1500	500	0.02	0.987	0.00006	200	0.640154

2	5	10	1500	1500	0.02	0.987	0.0001	100	0.709333
2	5	25	100	100	0.02	0.987	0.00002	200	0.736364
2	5	25	100	500	0.02	0.987	0.00002	100	0.86
2	5	25	100	1500	0.02	0.991	0.00002	100	0.837152
2	5	25	500	100	0.02	0.987	0.00002	200	0.818
2	5	25	500	500	0.02	0.987	0.00002	100	0.797091
2	5	25	500	1500	0.02	0.987	0.00002	100	0.584323
2	5	25	1500	100	0.02	0.987	0.00002	200	0.802
2	5	25	1500	500	0.02	0.991	0.00002	100	0.603231
2	5	25	1500	1500	0.02	0.987	0.00002	100	0.586061
2	10	5	100	100	0.02	0.987	0.0001	200	0.900909
2	10	5	100	500	0.02	0.987	0.0001	100	0.940196
2	10	5	100	1500	0.02	0.991	0.00006	200	0.807682
2	10	5	500	100	0.02	0.995	0.00006	200	0.647333
2	10	5	500	500	0.02	0.987	0.00006	200	0.446182
2	10	5	500	1500	0.02	0.987	0.00002	100	0.605032
2	10	5	1500	100	0.02	0.987	0.00002	100	0.6844
2	10	5	1500	500	0.02	0.987	0.00006	200	0.379077
2	10	5	1500	1500	0.02	0.987	0.00006	100	0.494727
2	10	10	100	100	0.02	0.987	0.00002	100	0.898182
2	10	10	100	500	0.02	0.987	0.00006	100	0.874902
2	10	10	100	1500	0.02	0.987	0.00002	200	0.908146
2	10	10	500	100	0.02	0.987	0.00002	100	0.804667
2	10	10	500	500	0.02	0.991	0.00002	200	0.661636
2	10	10	500	1500	0.02	0.987	0.00006	100	0.779097
2	10	10	1500	100	0.02	0.987	0.00002	200	0.7292
2	10	10	1500	500	0.02	0.987	0.00002	200	0.622462
2	10	10	1500	1500	0.02	0.987	0.00006	200	0.503091
2	10	25	100	100	0.02	0.987	0.00002	100	0.901818
2	10	25	100	500	0.02	0.991	0.00006	100	0.821176
2	10	25	100	1500	0.02	0.987	0.00006	100	0.800464
2	10	25	500	100	0.02	0.991	0.00002	200	0.756667
2	10	25	500	500	0.02	0.987	0.00002	100	0.554182
2	10	25	500	1500	0.02	0.987	0.0001	100	0.466581
2	10	25	1500	100	0.02	0.987	0.00002	200	0.726
2	10	25	1500	500	0.02	0.987	0.00002	100	0.549231
2	10	25	1500	1500	0.02	0.987	0.00006	100	0.640364
2	25	5	100	100	0.02	0.987	0.00002	100	0.812727
2	25	5	100	500	0.02	0.987	0.00002	100	0.793333
2	25	5	100	1500	0.02	0.987	0.0001	100	0.677748



2	25	5	500	100	0.02	0.987	0.00002	100	0.708667
2	25	5	500	500	0.02	0.987	0.00002	100	0.768909
2	25	5	500	1500	0.02	0.987	0.00006	100	0.557032
2	25	5	1500	100	0.02	0.987	0.00006	100	0.7684
2	25	5	1500	500	0.02	0.987	0.00006	200	0.476154
2	25	5	1500	1500	0.02	0.991	0.0001	200	0.479455
2	25	10	100	100	0.02	0.987	0.00002	100	0.775455
2	25	10	100	500	0.02	0.987	0.00002	100	0.86902
2	25	10	100	1500	0.02	0.991	0.00002	100	0.638742
2	25	10	500	100	0.02	0.987	0.0001	100	0.675333
2	25	10	500	500	0.02	0.987	0.00006	200	0.650182
2	25	10	500	1500	0.02	0.987	0.00006	100	0.476903
2	25	10	1500	100	0.02	0.991	0.00002	100	0.6924
2	25	10	1500	500	0.02	0.991	0.0001	100	0.495692
2	25	10	1500	1500	0.02	0.987	0.00002	200	0.316242
2	25	25	100	100	0.02	0.987	0.00002	200	0.66
2	25	25	100	500	0.02	0.991	0.0001	100	0.829804
2	25	25	100	1500	0.02	0.987	0.0001	100	0.788477
2	25	25	500	100	0.02	0.987	0.00006	100	0.672
2	25	25	500	500	0.02	0.991	0.00006	200	0.534364
2	25	25	500	1500	0.02	0.991	0.00002	100	0.434129
2	25	25	1500	100	0.02	0.987	0.00002	200	0.7796
2	25	25	1500	500	0.02	0.987	0.00006	100	0.532308
2	25	25	1500	1500	0.02	0.987	0.0001	100	0.484788
3	5	5	100	100	0.02	0.987	0.00002	200	0.989091
3	5	5	100	500	0.02	0.987	0.00002	100	0.999412
3	5	5	100	1500	0.02	0.987	0.00006	200	0.998013
3	5	5	500	100	0.02	0.987	0.00002	200	0.924667
3	5	5	500	500	0.02	0.987	0.00006	200	0.820727
3	5	5	500	1500	0.02	0.987	0.00002	100	0.927097
3	5	5	1500	100	0.02	0.987	0.00002	100	0.9224
3	5	5	1500	500	0.02	0.987	0.00002	100	0.937692
3	5	5	1500	1500	0.02	0.987	0.00006	200	0.844182
3	5	10	100	100	0.02	0.987	0.00002	100	0.986364
3	5	10	100	500	0.02	0.987	0.00002	100	0.99902
3	5	10	100	1500	0.02	0.987	0.00002	100	0.988013
3	5	10	500	100	0.02	0.987	0.00002	100	0.971333
3	5	10	500	500	0.02	0.987	0.00006	200	0.997636
3	5	10	500	1500	0.02	0.987	0.0001	100	0.941032
3	5	10	1500	100	0.02	0.987	0.00002	200	0.9476

3	5	10	1500	500	0.02	0.987	0.00002	100	0.854462
3	5	10	1500	1500	0.02	0.987	0.00002	100	0.838061
3	5	25	100	100	0.02	0.987	0.0001	100	0.915455
3	5	25	100	500	0.02	0.987	0.00002	100	0.98902
3	5	25	100	1500	0.02	0.987	0.00002	200	0.958477
3	5	25	500	100	0.02	0.987	0.0001	100	0.990667
3	5	25	500	500	0.02	0.987	0.0001	200	0.82
3	5	25	500	1500	0.02	0.987	0.00002	100	0.913548
3	5	25	1500	100	0.02	0.987	0.00002	100	0.9632
3	5	25	1500	500	0.02	0.987	0.00002	200	0.881846
3	5	25	1500	1500	0.02	0.987	0.00002	200	0.935697
3	10	5	100	100	0.02	0.987	0.00006	100	0.950909
3	10	5	100	500	0.02	0.987	0.00006	100	0.969216
3	10	5	100	1500	0.02	0.987	0.00002	200	0.988609
3	10	5	500	100	0.02	0.987	0.0001	200	0.984
3	10	5	500	500	0.02	0.987	0.00002	200	0.948727
3	10	5	500	1500	0.02	0.987	0.00002	200	0.789935
3	10	5	1500	100	0.02	0.987	0.00006	200	0.964
3	10	5	1500	500	0.02	0.987	0.00002	200	0.878308
3	10	5	1500	1500	0.02	0.987	0.00006	200	0.875697
3	10	10	100	100	0.02	0.987	0.00002	100	0.91
3	10	10	100	500	0.02	0.987	0.00006	100	0.984314
3	10	10	100	1500	0.02	0.987	0.00002	100	0.985033
3	10	10	500	100	0.02	0.987	0.00002	100	0.901333
3	10	10	500	500	0.02	0.987	0.00006	100	0.921818
3	10	10	500	1500	0.02	0.987	0.00006	200	0.844968
3	10	10	1500	100	0.02	0.987	0.00002	200	0.9356
3	10	10	1500	500	0.02	0.987	0.00002	100	0.754769
3	10	10	1500	1500	0.02	0.987	0.00002	200	0.846121
3	10	25	100	100	0.02	0.987	0.00006	100	0.953636
3	10	25	100	500	0.02	0.987	0.00006	100	0.992549
3	10	25	100	1500	0.02	0.987	0.00002	200	0.98596
3	10	25	500	100	0.02	0.987	0.00002	200	0.963333
3	10	25	500	500	0.02	0.987	0.00002	200	0.942364
3	10	25	500	1500	0.02	0.987	0.00006	200	0.964581
3	10	25	1500	100	0.02	0.987	0.00002	100	0.976
3	10	25	1500	500	0.02	0.987	0.00002	100	0.89
3	10	25	1500	1500	0.02	0.987	0.00006	100	0.864909
3	25	5	100	100	0.02	0.987	0.00002	100	0.973636
3	25	5	100	500	0.02	0.987	0.00006	100	0.929608

3	25	5	100	1500	0.02	0.987	0.00002	100	0.992649
3	25	5	500	100	0.02	0.987	0.00002	100	0.868667
3	25	5	500	500	0.02	0.987	0.00002	100	0.966364
3	25	5	500	1500	0.02	0.987	0.00002	100	0.879935
3	25	5	1500	100	0.02	0.987	0.00002	200	0.9676
3	25	5	1500	500	0.02	0.987	0.00002	200	0.969692
3	25	5	1500	1500	0.02	0.987	0.00002	200	0.732
3	25	10	100	100	0.02	0.987	0.0001	100	0.98
3	25	10	100	500	0.02	0.987	0.0001	100	0.982745
3	25	10	100	1500	0.02	0.987	0.00006	100	0.925364
3	25	10	500	100	0.02	0.987	0.00002	100	0.920667
3	25	10	500	500	0.02	0.987	0.00002	200	0.955091
3	25	10	500	1500	0.02	0.987	0.00002	100	0.932387
3	25	10	1500	100	0.02	0.987	0.0001	100	0.9728
3	25	10	1500	500	0.02	0.987	0.00006	100	0.902308
3	25	10	1500	1500	0.02	0.987	0.00002	200	0.847394
3	25	25	100	100	0.02	0.987	0.00002	100	0.979091
3	25	25	100	500	0.02	0.987	0.00002	200	0.981765
3	25	25	100	1500	0.02	0.987	0.00006	100	0.981589
3	25	25	500	100	0.02	0.987	0.00002	200	0.902667
3	25	25	500	500	0.02	0.987	0.00006	200	0.986727
3	25	25	500	1500	0.02	0.987	0.00002	100	0.916839
3	25	25	1500	100	0.02	0.987	0.00002	200	0.9484
3	25	25	1500	500	0.02	0.987	0.00002	200	0.638769
3	25	25	1500	1500	0.02	0.987	0.00006	200	0.815394
100	5	5	100	100	0.02	0.987	0.00002	100	1
100	5	5	100	500	0.02	0.987	0.00002	100	1
100	5	5	100	1500	0.02	0.987	0.00002	100	1
100	5	5	500	100	0.02	0.987	0.00002	100	1
100	5	5	500	500	0.02	0.987	0.00002	100	1
100	5	5	500	1500	0.02	0.987	0.00002	100	1
100	5	5	1500	100	0.02	0.987	0.00002	100	1
100	5	5	1500	500	0.02	0.987	0.00002	100	1
100	5	5	1500	1500	0.02	0.987	0.00002	100	1
100	5	10	100	100	0.02	0.987	0.00002	100	1
100	5	10	100	500	0.02	0.987	0.00002	100	1
100	5	10	100	1500	0.02	0.987	0.00002	100	1
100	5	10	500	100	0.02	0.987	0.00002	100	1
100	5	10	500	500	0.02	0.987	0.00002	100	1
100	5	10	500	1500	0.02	0.987	0.00002	100	1

100	5	10	1500	100	0.02	0.987	0.00002	100	1
100	5	10	1500	500	0.02	0.987	0.00002	100	1
100	5	10	1500	1500	0.02	0.987	0.00002	100	1
100	5	25	100	100	0.02	0.987	0.00002	100	1
100	5	25	100	500	0.02	0.987	0.00002	100	1
100	5	25	100	1500	0.02	0.987	0.00002	100	1
100	5	25	500	100	0.02	0.987	0.00002	100	1
100	5	25	500	500	0.02	0.987	0.00002	100	1
100	5	25	500	1500	0.02	0.987	0.00002	100	1
100	5	25	1500	100	0.02	0.987	0.00002	100	1
100	5	25	1500	500	0.02	0.987	0.00002	100	1
100	5	25	1500	1500	0.02	0.987	0.00002	100	1
100	10	5	100	100	0.02	0.987	0.00002	100	1
100	10	5	100	500	0.02	0.987	0.00002	100	1
100	10	5	100	1500	0.02	0.987	0.00002	100	1
100	10	5	500	100	0.02	0.987	0.00002	100	1
100	10	5	500	500	0.02	0.987	0.00002	100	1
100	10	5	500	1500	0.02	0.987	0.00002	100	1
100	10	5	1500	100	0.02	0.987	0.00002	100	1
100	10	5	1500	500	0.02	0.987	0.00002	100	1
100	10	5	1500	1500	0.02	0.987	0.00002	100	1
100	10	10	100	100	0.02	0.987	0.00002	100	1
100	10	10	100	500	0.02	0.987	0.00002	100	1
100	10	10	100	1500	0.02	0.987	0.00002	100	1
100	10	10	500	100	0.02	0.987	0.00002	100	1
100	10	10	500	500	0.02	0.987	0.00002	100	1
100	10	10	500	1500	0.02	0.987	0.00002	100	1
100	10	10	1500	100	0.02	0.987	0.00002	100	1
100	10	10	1500	500	0.02	0.987	0.00002	100	1
100	10	10	1500	1500	0.02	0.987	0.00002	100	1
100	10	25	100	100	0.02	0.987	0.00002	100	1
100	10	25	100	500	0.02	0.987	0.00002	100	1
100	10	25	100	1500	0.02	0.987	0.00002	100	1
100	10	25	500	100	0.02	0.987	0.00002	100	1
100	10	25	500	500	0.02	0.987	0.00002	100	1
100	10	25	500	1500	0.02	0.987	0.00002	100	1
100	10	25	1500	100	0.02	0.987	0.00002	100	1
100	10	25	1500	500	0.02	0.987	0.00002	100	1
100	10	25	1500	1500	0.02	0.987	0.00002	100	1
100	25	5	100	100	0.02	0.987	0.00002	100	1

100	25	5	100	500	0.02	0.987	0.00002	100	1
100	25	5	100	1500	0.02	0.987	0.00002	100	1
100	25	5	500	100	0.02	0.987	0.00002	100	1
100	25	5	500	500	0.02	0.987	0.00002	100	1
100	25	5	500	1500	0.02	0.987	0.00002	100	1
100	25	5	1500	100	0.02	0.987	0.00002	100	1
100	25	5	1500	500	0.02	0.987	0.00002	100	1
100	25	5	1500	1500	0.02	0.987	0.00002	100	1
100	25	10	100	100	0.02	0.987	0.00002	100	1
100	25	10	100	500	0.02	0.987	0.00002	100	1
100	25	10	100	1500	0.02	0.987	0.00002	100	1
100	25	10	500	100	0.02	0.987	0.00002	100	1
100	25	10	500	500	0.02	0.987	0.00002	100	1
100	25	10	500	1500	0.02	0.987	0.00002	100	1
100	25	10	1500	100	0.02	0.987	0.00002	100	1
100	25	10	1500	500	0.02	0.987	0.00002	100	1
100	25	10	1500	1500	0.02	0.987	0.00002	100	1
100	25	25	100	100	0.02	0.987	0.00002	100	1
100	25	25	100	500	0.02	0.987	0.00002	100	1
100	25	25	100	1500	0.02	0.987	0.00002	100	1
100	25	25	500	100	0.02	0.987	0.00002	100	1
100	25	25	500	500	0.02	0.987	0.00002	100	1
100	25	25	500	1500	0.02	0.987	0.00002	100	1
100	25	25	1500	100	0.02	0.987	0.00002	100	1
100	25	25	1500	500	0.02	0.987	0.00002	100	1
100	25	25	1500	1500	0.02	0.987	0.00002	100	1

## 6 Συμπεράσματα και Μελλοντική Έρευνα

Το ίδιο πείραμα που περιγράφηκε παραπάνω επαναλήφθηκε ακόμα μια φορά μειώνοντας ακόμα περισσότερο τις αποστάσεις μεταξύ των πυρήνων. Αυτή την φορά ορίσαμε την απόσταση των πυρήνων του συνόλου μη εαυτός από τον κεντρικό πυρήνα να παίρνει τιμές στο διάστημα [2, 3] ενώ οι τιμές των ακτινών των πυρήνων του συνόλου εαυτός παίρνουν τιμές στο διάστημα [5, 6]. Αυτό που παρατηρήθηκε είναι ότι η διαδικασία δημιουργίας των ανιχνευτών σε αρκετές περιπτώσεις έγινε αρκετά δύσκολη με αποτέλεσμα είτε να καθυστερεί δυσανάλογα περισσότερο από τις προηγούμενες περιπτώσεις και πολλές φορές να μη μπορεί να προσδιορίσει έναν ανιχνευτή με αποτέλεσμα να μην προχωράει η εκτέλεση του αλγόριθμου για απαγορευτικά μεγάλο χρονικό διάστημα. Αυτό είναι λογικό όταν τα δεδομένα βρίσκονται πολύ «κοντά» μεταξύ τους, καθώς ο αλγόριθμος τοποθετεί τυχαία έναν ανιχνευτή στον χώρο μας και ελέγχει τις αποστάσεις από όλες τις N-σφαίρες που ορίζουν τα σημεία του συνόλου εαυτός μαζί με την Rself. Ουσιαστικά όταν αυτός ο ανιχνευτής βρίσκεται στο εσωτερικό αυτής της N-σφαίρας τον απορρίπτουμε, και με έναν μεγάλο αριθμό από πυρήνες και δεδομένα είναι πιθανόν να έχουν διασκορπιστεί στον χώρο τόσο ομοιόμορφα ώστε να είναι πολύ δύσκολο να δημιουργηθεί τυχαία ένας ή περισσότεροι ανιχνευτές οι οποίοι θα είναι αποδεκτοί.

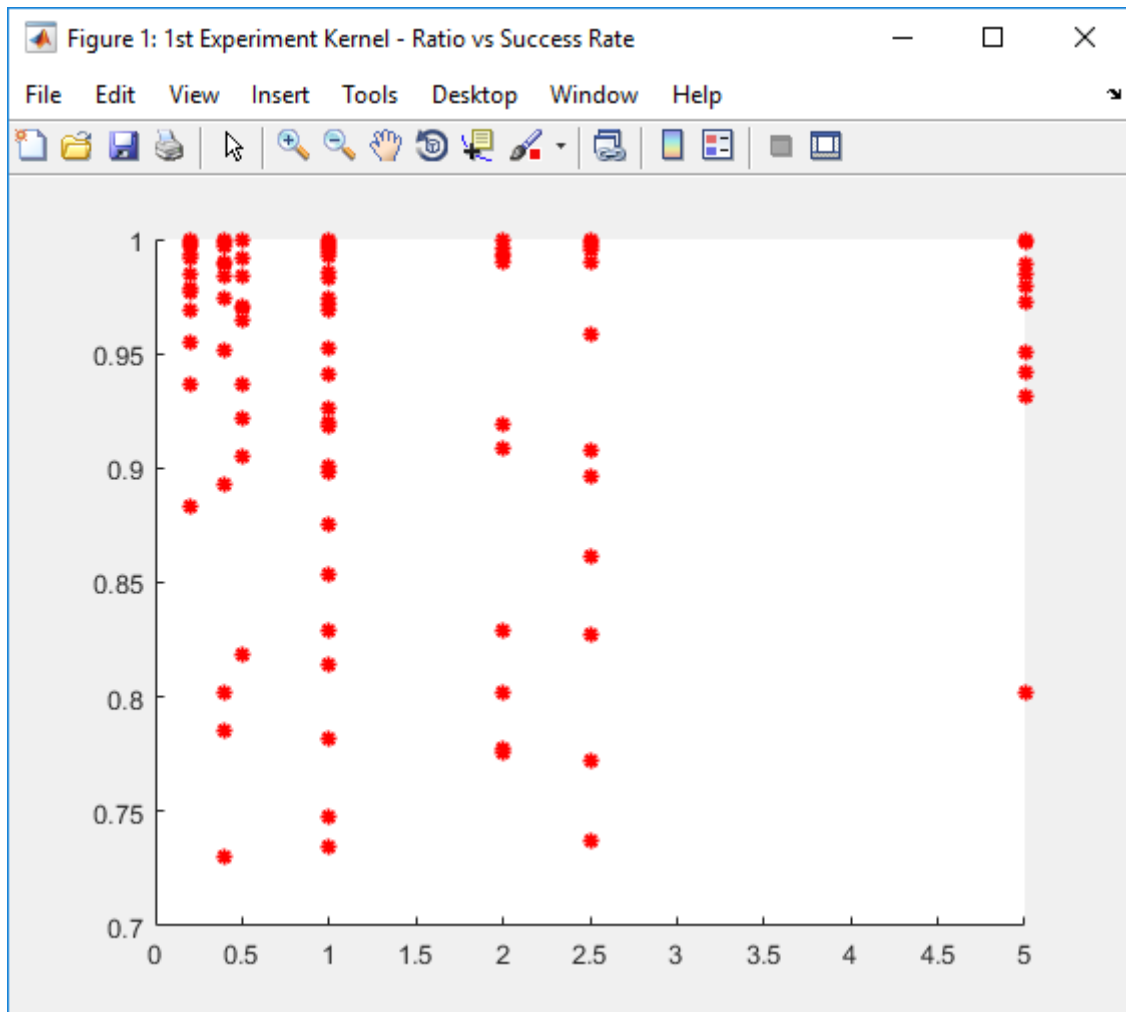
Το παραπάνω πρόβλημα παρατηρήσαμε ότι δημιουργείται όταν υπάρχουν αρκετοί πυρήνες στον χώρο μας και ο αριθμός των δεδομένων της κλάσης είναι μεγάλος. Σημαντικός παράγοντας είναι επίσης και το  $R_{self}$  καθώς όσο μεγαλύτερη είναι αυτή η ποσότητα τόσο μεγαλύτερο μέρος του χώρου θα καταλαμβάνει το σύνολο εαυτός και θα είναι αρκετά δύσκολο να δημιουργήσουμε αποδεκτό ανιχνευτή. Η παραπάνω συμπεριφορά ήταν αναμενόμενη μέχρι ένα σημείο καθώς ο V-Detector είναι ένας αλγόριθμος ο οποίος επηρεάζεται από την γεωμετρική φύση των δεδομένων και την κατανομή τους και συνεπώς όταν τα δεδομένα είναι πυκνά στον χώρο περιορίζεται η αποτελεσματικότητά του.

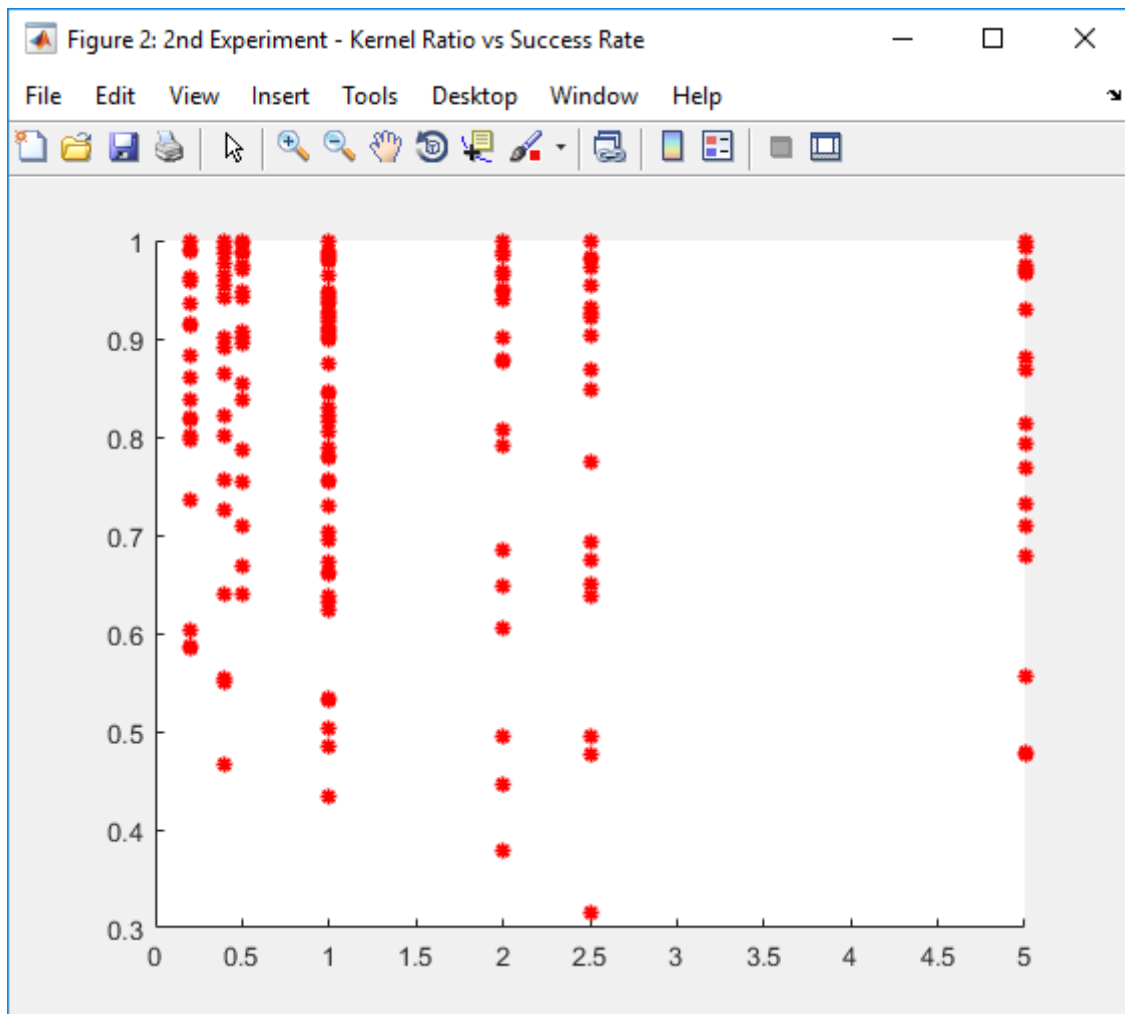
Γενικότερα, παρατηρούμε ότι ο αλγόριθμος διατηρεί υψηλά ποσοστά επιτυχίας ακόμα και όταν αρχίσουν τα δεδομένα να περιπλέκονται σε ένα μικρό βαθμό. Σίγουρα υπάρχει μια μείωση της απόδοσης, κατά μέσο όρο της τάξης του 9% στα αποτελέσματα του δεύτερου πειράματος σε σχέση με το πρώτο. Ακόμα σημαντικό ρόλο παίζει και το  $R_{self}$  καθώς όσο μικρότερο είναι τόσο περισσότερο μας επιτρέπει να δημιουργήσουμε έναν αποδεκτό ανιχνευτή στον χώρο. Πιθανόν, όμως το  $R_{self}$  να μην είναι διαπραγματεύσιμο σε ένα πρόβλημα του πραγματικού κόσμου οπότε θα πρέπει η όποια βελτιστοποίηση να γίνει ως προς τα υπόλοιπα ορίσματα ή την καλύτερη επιλογή των δεδομένων εισόδου. Στη συνέχεια ακολουθούν μερικά γραφήματα ώστε να μπορέσουμε να αντιληφθούμε καλύτερα πληροφορίες που παρέχονται από τους 2 παραπάνω πίνακες και να μπορέσουμε να συγκρίνουμε και οπτικά τα 2 αυτά πειράματα.

Μελλοντικά, η έρευνα θα μπορούσε να επικεντρωθεί στην αναζήτηση ενός τρόπου βελτιστοποίησης της διαδικασίας δημιουργίας των ανιχνευτών για τις περιπτώσεις όπου η δεδομένη μορφή του αλγόριθμου αποτυγχάνει να παράγει τους ανιχνευτές αποτελεσματικά. Ακόμα, θα μπορούσαμε να επικεντρωθούμε σε μια πιο συγκεκριμένη γεωμετρική δομή δεδομένων η οποία διέπει ένα πρόβλημα στον πραγματικό κόσμο και να μελετήσουμε τις διάφορες περιπτώσεις των δεδομένων που παράγονται και αντιστοιχούν σε αυτό με τρόπο αντίστοιχο αυτής της εργασίας.

## 6.1 Γράφημα 1°

Στα παρακάτω διαγράμματα μπορούμε να διακρίνουμε την κατανομή των ποσοστών επιτυχημένων ταξινομήσεων ως προς τον λόγο των πυρήνων της κάθε κλάσης. Οι ποσότητες κάτω από 1 ορίζουν ότι ο αριθμός των πυρήνων του συνόλου εαυτός είναι λιγότεροι από αυτούς του συνόλου μη εαυτός ενώ οι ποσότητες μεγαλύτερες του 1 το αντίθετο.



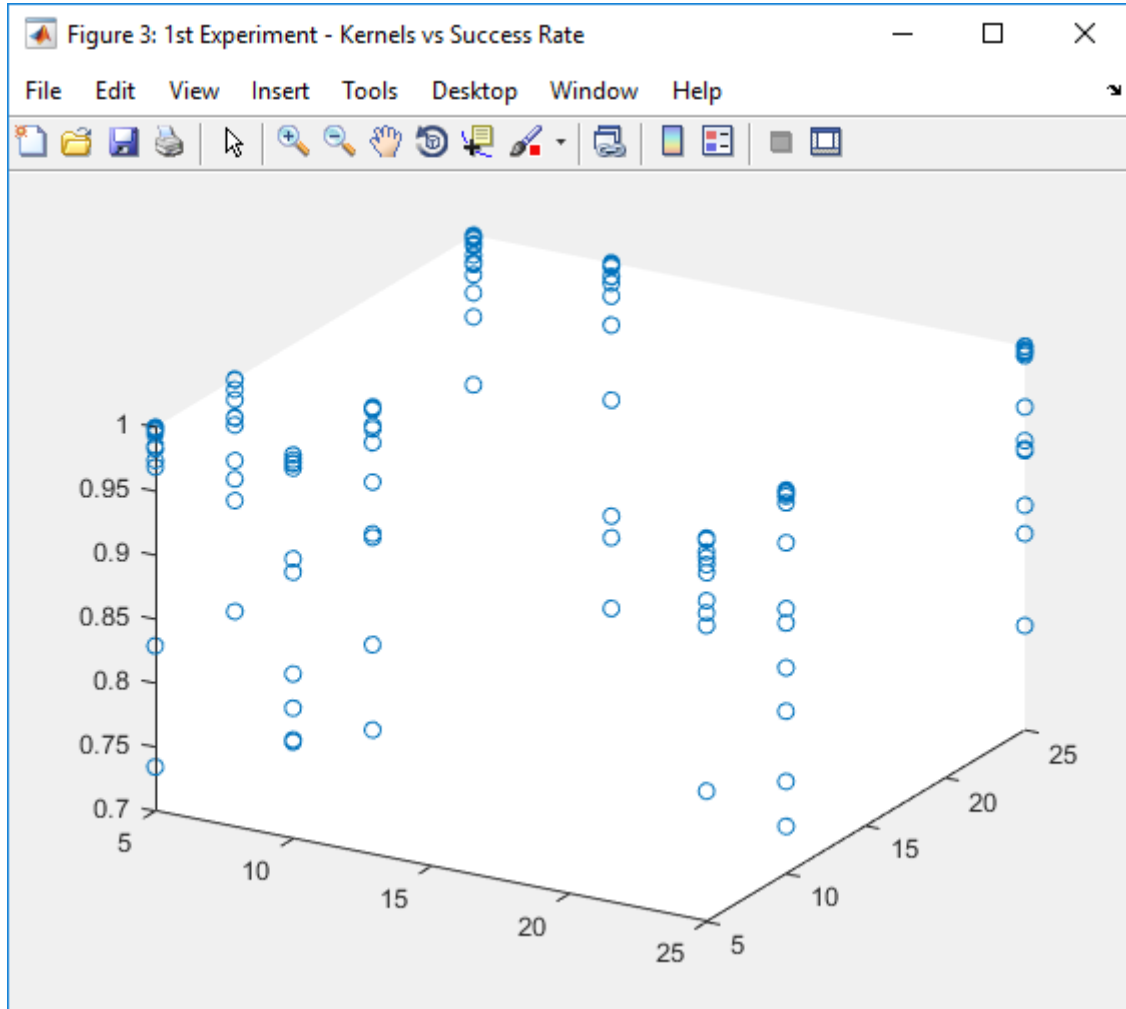


Εικόνα 6.2 - Πείραμα 2<sup>ο</sup>  
Λόγος πυρήνων προς ποσοστό επιτυχημένης ταξινόμησης.

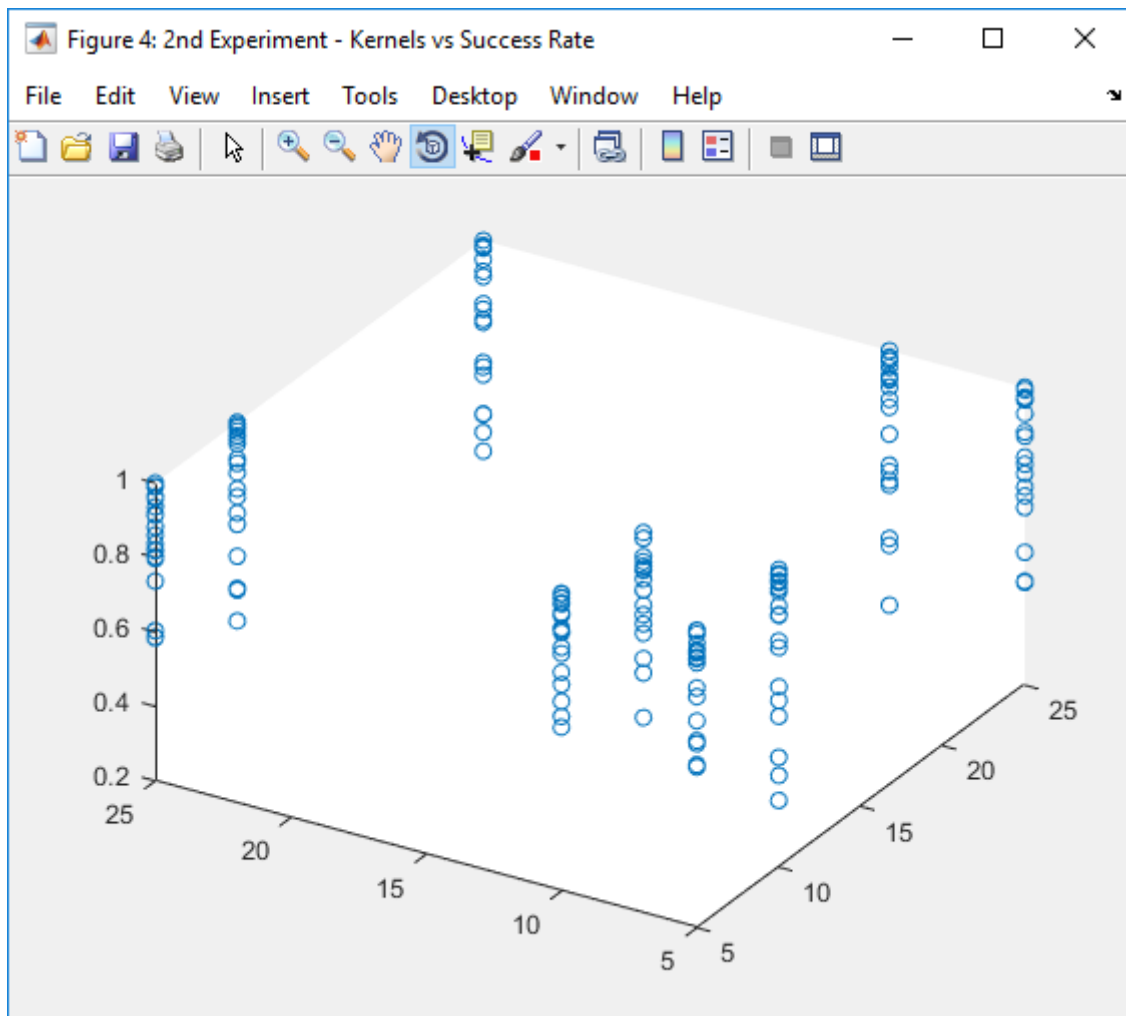


## 6.2 Γράφημα 2°

Στα παρακάτω διαγράμματα μπορούμε να διακρίνουμε την κατανομή των ποσοστών επιτυχημένων ταξινομήσεων ως προς τον αριθμό των πυρήνων της κάθε κλάσης.



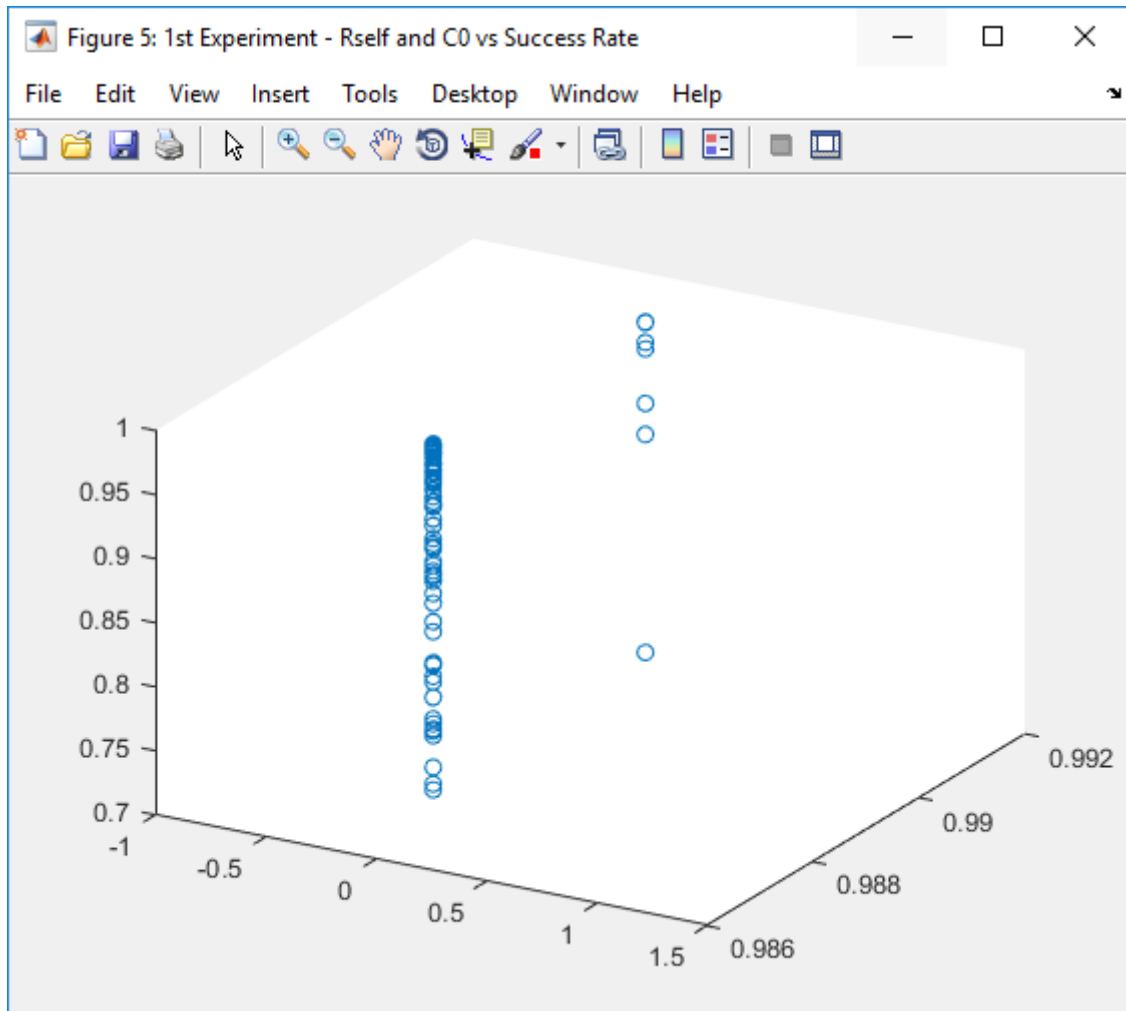
**Εικόνα 6.3 - Πείραμα 1°**  
Αριθμός πυρήνων (εαυτός) προς αριθμό πυρήνων (μη εαυτός)  
προς ποσοστό επιτυχημένης ταξινόμησης.



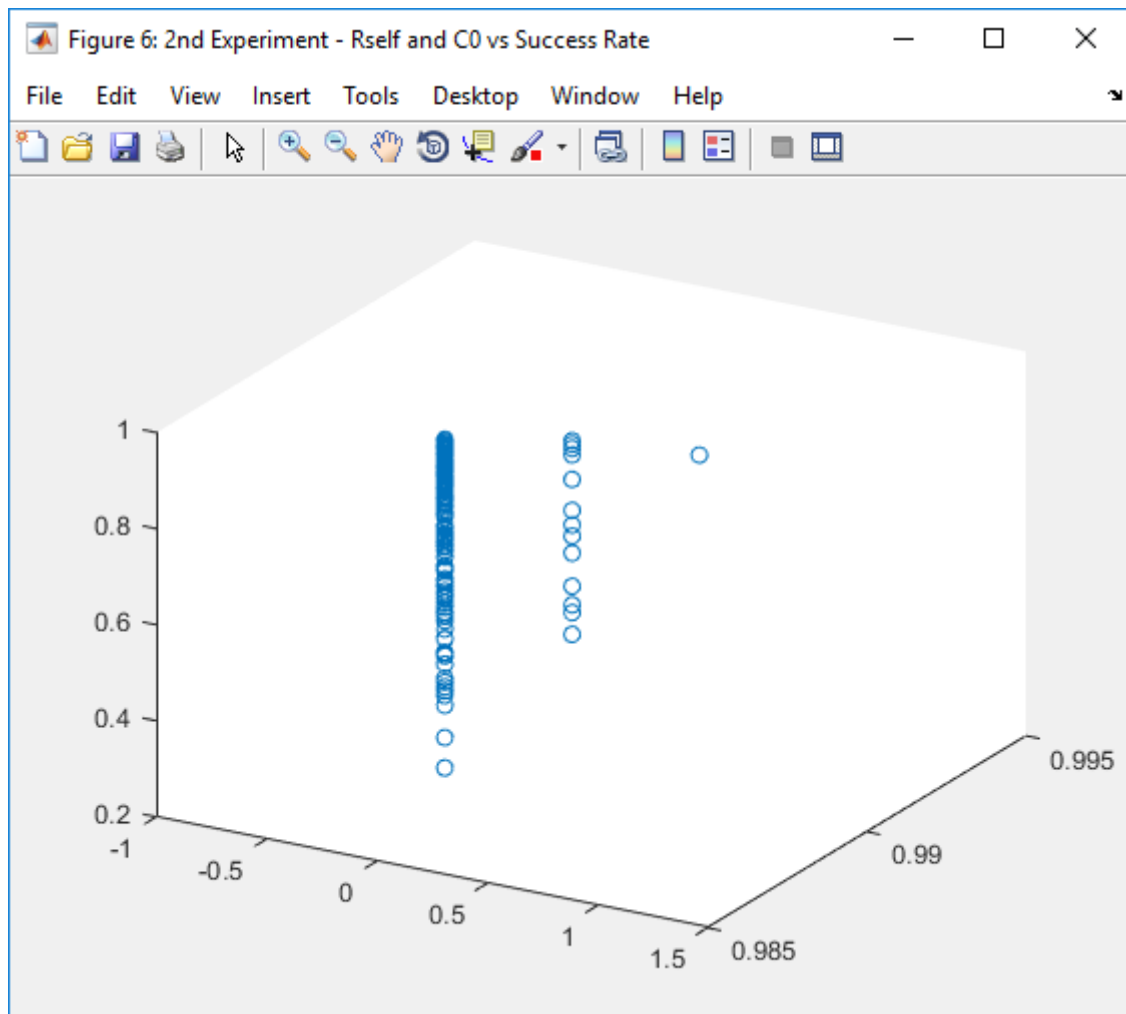
**Εικόνα 6.4 - Πείραμα 2<sup>ο</sup>**  
**Αριθμός πυρήνων (εαυτός) προς αριθμό πυρήνων (μη εαυτός)**  
**προς ποσοστό επιτυχημένης ταξινόμησης.**

### 6.3 Γράφημα 3°

Στα παρακάτω διαγράμματα μπορούμε να διακρίνουμε την κατανομή των ποσοστών επιτυχημένων ταξινομήσεων ως προς το  $R_{self}$  και το  $C_0$ . Παρατηρούμε ότι τα περισσότερα σημεία έχουν προκύψει με ποσοστό κάλυψης μικρότερο του 0.99 ενώ αντίστοιχα τα σημεία με ποσοστό κάλυψης μεγαλύτερο του είναι πολύ λιγότερα.



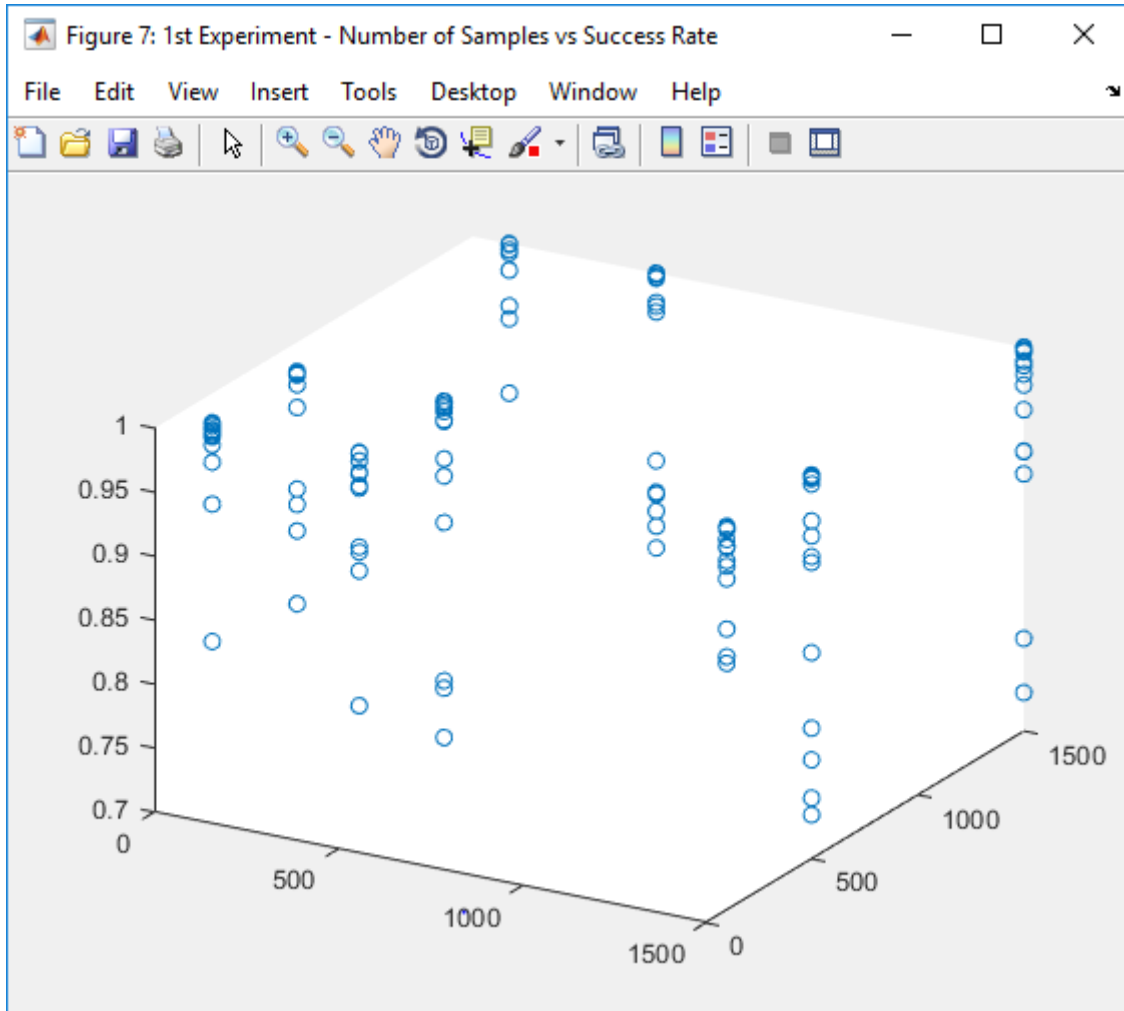
Εικόνα 6.5 - Πείραμα 1°  
Ακτίνα  $R_{self}$  προς ποσοστό κάλυψης  $C_0$   
προς ποσοστό επιτυχημένης ταξινόμησης.



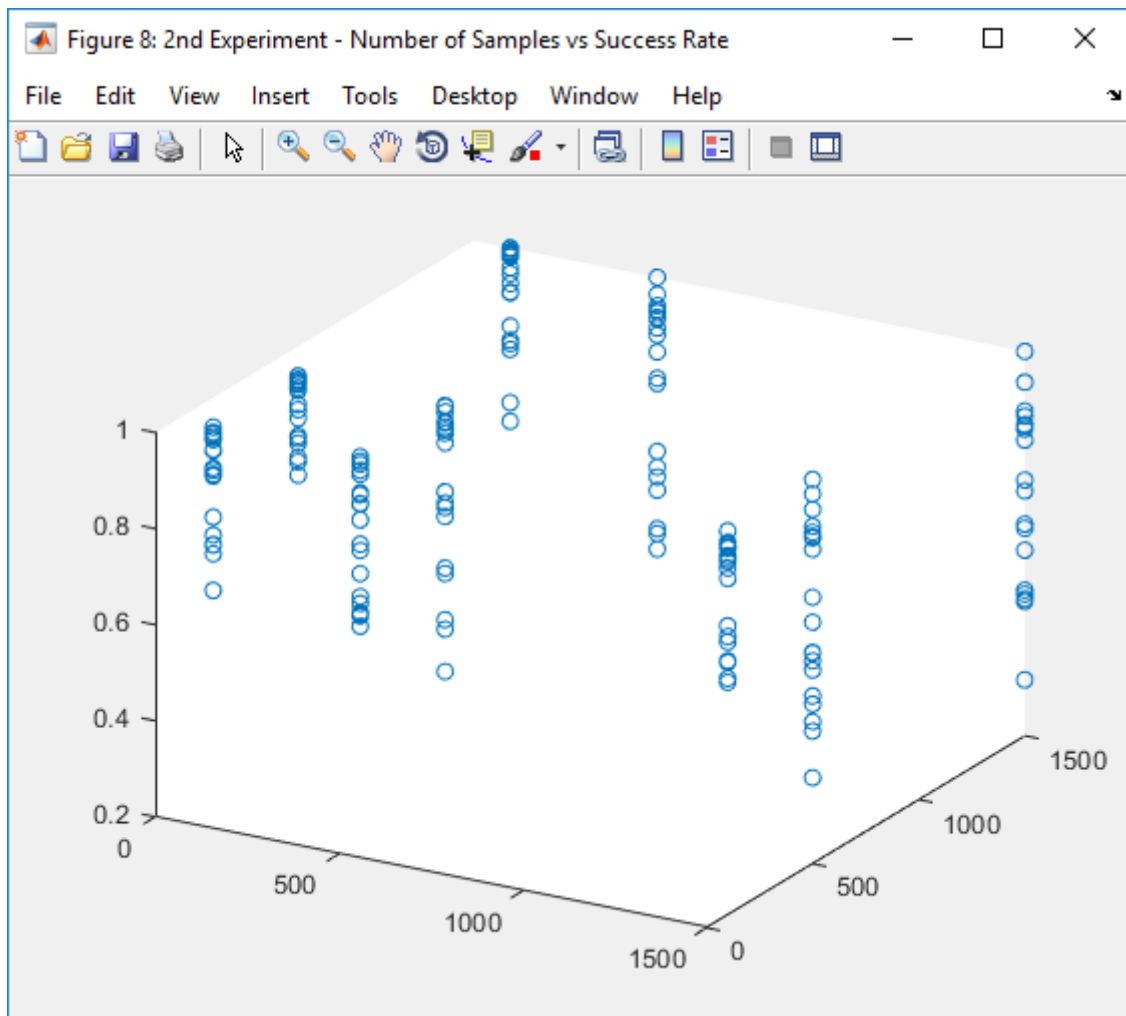
**Εικόνα 6.6 - Πείραμα 2<sup>ο</sup>**  
**Ακτίνα Rself προς ποσοστό κάλυψης C<sub>0</sub>**  
**προς ποσοστό επιτυχημένης ταξινόμησης.**

## 6.4 Γράφημα 4<sup>ο</sup>

Στα παρακάτω διαγράμματα μπορούμε να διακρίνουμε την κατανομή των ποσοστών επιτυχημένων ταξινομήσεων ως προς τον αριθμό των δεδομένων της κάθε κλάσης.



**Εικόνα 6.7 - Πείραμα 1<sup>ο</sup>**  
**Αριθμός δεδομένων (εαυτός) προς αριθμό δεδομένων (μη εαυτός)**  
**προς ποσοστό επιτυχημένης ταξινόμησης.**



**Εικόνα 6.8 - Πείραμα 2<sup>ο</sup>**  
**Αριθμός δεδομένων (εαυτός) προς αριθμό δεδομένων (μη εαυτός)**  
**προς ποσοστό επιτυχημένης ταξινόμησης.**

## 7 Βιβλιογραφικές Πηγές

Για τη σύνταξη της παραπάνω διπλωματικής εργασίας χρησιμοποιήθηκαν οι παρακάτω βιβλιογραφικές πηγές.

1. [https://el.wikipedia.org/wiki/Μηχανική\\_μάθηση](https://el.wikipedia.org/wiki/Μηχανική_μάθηση)
2. [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning)
3. Machine Learning Paradigms, Dionisios N. Sotiropoulos, George A. Tsihrintzis, Intelligent Systems Reference Library 118.
  1. Σελ. vii Foreword.
  2. Σελ. 26, Supervised Learning.
  3. Σελ. 91, One Class Classifiers
  4. Σελ. 133-134, 6.1 Introduction
  5. Σελ. 137-139, 6.3 Fundamentals and Main Components.
  6. Σελ. 214-217, 7.7 AIS-Based Negative Selection
  7. Σελ. 220-221, 7.7.3 Real-Valued Negative Selection with Variable-Sized Detectors.
  8. Σελ. 225-229, 7.7.5 V-Detector Algorithm.
4. <https://en.wikipedia.org/wiki/N-sphere>
5. Artificial Immune Systems for Self-NonSelf Discrimination: Application to Anomaly Detection, Electrical and Computer Engineering Department, Kansas University, Manhattan, KS, USA.  
Authors: Sanjoy Das, Min Gui, Anil Pahwa.