



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ



ΤΙΤΛΟΣ ΕΡΓΑΣΙΑΣ

Παραλλαγές του Αλγορίθμου κ-Μέσων

Constrained k-Means Classification

ΑΠΑΛΛΑΚΤΙΚΗ ΕΡΓΑΣΙΑ - ΧΕΙΜΕΡΙΝΟ ΕΞΑΜΗΝΟ 2023-2024

ΟΝΟΜΑΤΕΠΩΝΥΜΟ: ΚΑΔΡΕΒΗ ΜΑΡΙΑ-ΕΛΕΝΗ

Επιβλέπων: Μάρκος Τσίπουρας

Περιεχόμενα

ΕΙΣΑΓΩΓΗ

ΕΠΕΞΗΓΗΣΗ K-MEANS ALGORITHM

ΕΠΕΞΗΓΗΣΗ C-K-MEANS ALGORITHM

ΑΝΑΛΥΣΗ ΤΕΧΝΙΚΗΣ ΓΙΑ ΤΗΝ ΥΛΟΠΟΙΗΣΗ ΤΟΥ C-K-MEANS
ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ

- ΤΡΟΠΟΠΟΙΗΣΗ ΣΥΝΟΛΩΝ ΔΕΔΟΜΕΝΩΝ
- IRIS DATASET
- WINE DATASET
- BREAST CANCER WISCONSIN (DIAGNOSTIC) DATASET

ΑΠΟΤΕΛΕΣΜΑΤΑ

- IRIS DATASET
- WINE DATASET
- BREAST CANCER WISCONSIN (DIAGNOSTIC) DATASET

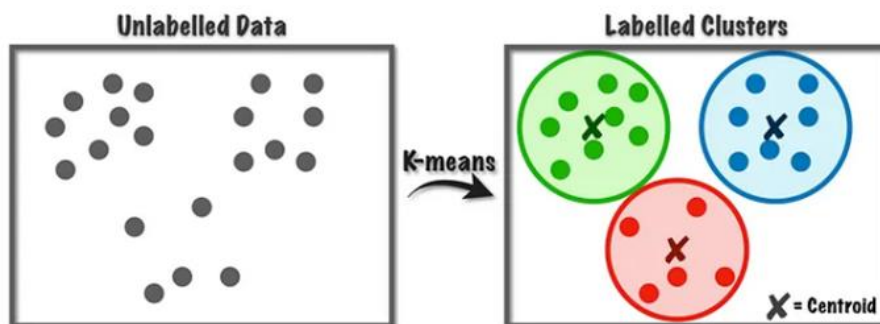
ΕΙΣΑΓΩΓΗ

Δύο βασικές μέθοδοι για την αξιολόγηση και την κατανόηση δεδομένων στους τομείς της μηχανικής μάθησης και της εξόρυξης δεδομένων είναι η ομαδοποίηση και η ταξινόμηση. Αυτές οι μέθοδοι είναι απαραίτητες για την οργάνωση των δεδομένων σε δομές με νόημα, γεγονός που τις καθιστά χρήσιμες σε ένα ευρύ φάσμα εφαρμογών σε πολλούς τομείς.

Ο αλγόριθμος K-means είναι ένας από τους πιο δημοφιλείς αλγορίθμους ομαδοποίησης επειδή είναι εύκολος στη χρήση και αποτελεσματικός. Λειτουργεί διαιρώντας ένα σύνολο δεδομένων σε διακριτές ομάδες (συστάδες) με βάση την ιδέα ότι τα σημεία δεδομένων σε μια ομάδα είναι πιο όμοια μεταξύ τους απ' ό,τι σε άλλες ομάδες. Η σταθμισμένη ευκλείδεια απόσταση χρησιμοποιείται ως κριτήριο ομοιότητας στη μέθοδο C-K-means, η οποία είναι μια παραλλαγή του κανονικού αλγορίθμου K-means που θέτει περιορισμούς στην κίνηση των κεντροειδών. Αυτή η τροποποίηση παρέχει μια πιο εξελιγμένη μέθοδο ομαδοποίησης, ενώ παράλληλα αντιμετωπίζει ορισμένα από τα μειονεκτήματα του αλγορίθμου K-means.

ΕΠΕΞΗΓΗΣΗ K-Means Algorithm

Ο αλγόριθμος K-means είναι μια επαναληπτική τεχνική που χρησιμοποιεί χαρακτηριστικά για να διαχωρίσει ένα σύνολο N σημείων δεδομένων σε K μη επικαλυπτόμενες υποομάδες ή συστάδες. Ο κύριος στόχος είναι η ελαχιστοποίηση του αθροίσματος των τετραγωνικών αποστάσεων μεταξύ κάθε σημείου και του κεντροειδούς της συστάδας ή του αθροίσματος τετραγώνων εντός της συστάδας.



Στη συνέχεια, εκτελεί επαναληπτικά δύο βήματα.

1. Βήμα ανάθεσης:

Κάθε σημείο δεδομένων στο σύνολο δεδομένων εκχωρείται στο πλησιέστερο κέντρο. Το «πλησιέστερο» καθορίζεται συνήθως με βάση την Ευκλείδεια απόσταση. Μετά από αυτό το βήμα, σχηματίζονται k συστάδες με βάση τις τρέχουσες θέσεις των κεντροειδών.

2. Βήμα ενημέρωσης:

Για κάθε ένα από τα k συμπλέγματα, το κεντροειδές ενημερώνεται ώστε να είναι ο μέσος όρος (μέσος όρος) όλων των σημείων δεδομένων που έχουν εκχωρηθεί σε αυτό το σύμπλεγμα. Αυτό το βήμα ουσιαστικά μετακινεί το κέντρο στο κέντρο του συμπλέγματός του.

Αυτή η διαδικασία συνεχίζεται μέχρι να μην υπάρξει σημαντική αλλαγή στις θέσεις των κεντροειδών, υποδηλώνοντας ότι οι συστάδες έχουν σταθεροποιηθεί.

ΕΠΕΞΗΓΗΣΗ C-K-Means Algorithm

Σύμφωνα με το σχετικό paper, η ταξινόμηση Constrained K-Means (C-K-Means) είναι ένας επαναστατικός αλγόριθμος CnC που δημιουργήθηκε ρητά για να αντιμετωπίσει ορισμένα από τα

μειονεκτήματα και τις ελλείψεις που εντοπίστηκαν σε προηγούμενες προσεγγίσεις.

Η τεχνική αυτή παρουσιάζει δύο κύριες αλλαγές σε σχέση με τη συμβατική μέθοδο K-Means. Αυτό το επιτυγχάνει εφαρμόζοντας πρώτα περιορισμούς στη διαδικασία ομαδοποίησης και στη συνέχεια χρησιμοποιώντας ένα σταθμισμένο μέτρο ευκλείδειας απόστασης.

Αυτή η μέθοδος είναι ιδιαίτερη στο ότι κάθε υπερκύβος περιορισμών παράγεται από δεδομένα για μία μόνο κλάση, επομένως κάθε κεντροειδές συνδέεται φυσικά με την κλάση του αντίστοιχου υπερκύβου.

Είναι ενδιαφέρον ότι τόσο για τη διαδικασία συσταδοποίησης όσο και για την κατασκευή των συστάδων χρησιμοποιούνται δεδομένα δοκιμής και όχι σύνολο εκπαίδευσης. Αυτή η μέθοδος βελτιώνει την ταξινόμηση των δοκιμαστικών παραδειγμάτων με τη δυναμική ενημέρωση των κεντροειδών κατά τη διάρκεια της κατασκευής των συστάδων του δοκιμαστικού συνόλου, εκτός από τη συνεκτίμηση της απόστασης κάθε δεδομένου από τα κεντροειδή. Αυτή η νέα προσέγγιση προσφέρει αυξημένη αποτελεσματικότητα και ακρίβεια για την ταξινόμηση των περιπτώσεων δοκιμής.

ΑΝΑΛΥΣΗ ΤΕΧΝΙΚΗΣ ΤΗΣ ΥΛΟΠΟΙΗΣΗΣ ΤΟΥ C-K-Means Algorithm

Το εξελιγμένο μοντέλο μηχανικής μάθησης constrained k-means classification (c-k-means) χρησιμοποιεί μια ξεχωριστή μέθοδο για την ομαδοποίηση και την ταξινόμηση.

Το μοντέλο δημιουργεί έναν υπερκύβο για κάθε κλάση κατά τη

διάρκεια της εκπαίδευσης και εκχωρεί βάρη σε κάθε χαρακτηριστικό της κλάσης.

Ο αλγόριθμος ξεκινά λαμβάνοντας ένα σύνολο δεδομένων εκπαίδευσης με "a" χαρακτηριστικά και "c" κλάσεις, στη συνέχεια, για κάθε κλάση, δημιουργεί έναν υπερκυβο καθορίζοντας τα ελάχιστα και μέγιστα όρια για κάθε χαρακτηριστικό.

$$b_{a,c}^{min} = avg_{a,c} - l * s_{a,c}$$
$$b_{a,c}^{max} = avg_{a,c} + l * s_{a,c}$$

Εικόνα 1: ΤΥΠΟΣ ΚΑΘΟΡΙΣΜΟΥ ΕΛΑΧΙΣΤΩΝ ΚΑΙ ΜΕΓΙΣΤΩΝ ΟΡΙΩΝ ΥΠΕΡΚΥΒΟΥ

Τα τρία κύρια μέρη της υλοποίησης του αλγορίθμου c-k-means περιστρέφονται γύρω από μια τροποποιημένη μέθοδο k-means.

Τα αρχικά «K» κεντροειδή αρχικοποιούνται τυχαία μέσα στους υπερκύβους, εξασφαλίζοντας ένα κέντρο ανά υπερκύβο.

$$b_{a,c}^{min} \leq m_{a,c} \leq b_{a,c}^{max}$$

Εικόνα 2: ΣΧΕΣΗ ΟΡΙΩΝ ΚΑΙ ΚΕΝΤΡΩΝ

Το βήμα ανάθεσης χρησιμοποιεί στη συνέχεια τη σταθμισμένη Ευκλείδεια απόσταση ως κριτήριο ομοιότητας για να αντιστοιχίσει κάθε δεδομένο στο πλησιέστερο κέντρο της.

Η σταθμισμένη Ευκλείδεια απόσταση υπολογίζεται, λαμβάνοντας υπόψη τα βάρη που αποδίδονται κατά τη διάρκεια της διαδικασίας εκπαίδευσης.

$$w_{a,c} = 1 - \frac{s_{a,c}}{\max_{a,c} - \min_{a,c}}$$

Εικόνα 3: ΤΥΠΟΣ ΥΠΟΛΟΓΙΣΜΟΥ ΒΑΡΩΝ

$$d_c = \sqrt{\sum_{a=1}^A w_{a,c} * (x_{a,i} - m_{a,c})^2}$$

Εικόνα 4:ΣΤΑΘΜΙΣΜΕΝΗ ΕΥΚΛΕΙΔΕΙΑ ΑΠΟΣΤΑΣΗ

Το βήμα ενημέρωσης του αλγορίθμου c-k-means είναι ένα κρίσιμο στοιχείο. Εάν ένα κέντρο πρέπει να τοποθετηθεί έξω από τον υπερκύβο περιορισμού της αρχικής του κατηγορίας, αναγκάζεται στις δεσμευμένες τιμές του υπερκύβου.

Εάν η τιμή ενός κέντρου πέσει κάτω από το ελάχιστο όριο (b_{\min_ac}) της κατηγορίας του, ορίζεται σε αυτήν την ελάχιστη τιμή και ομοίως, εάν υπερβαίνει το μέγιστο όριο (b_{\max_ac}), προσαρμόζεται σε αυτήν τη μέγιστη τιμή.

ΠΑΡΑΜΕΤΡΟΣ ΧΑΛΑΡΩΣΗΣ

Ένα επιπλέον βασικό χαρακτηριστικό του αλγορίθμου C-K-Means είναι η χρήση υπερκύβων για τον καθορισμό ορίων ή περιορισμών για κάθε κατηγορία στον χώρο χαρακτηριστικών. Το μέγεθος αυτών των υπερκύβων επηρεάζεται από μια παράμετρο χαλάρωσης.

Μια μεγαλύτερη παράμετρος χαλάρωσης οδηγεί σε μεγαλύτερους υπερκύβους, παρέχοντας μεγαλύτερη ευελιξία για κεντροειδείς κινήσεις, που μπορεί να είναι πλεονεκτικό για την προσαρμογή σε δεδομένα, αλλά μπορεί να οδηγήσει σε αλληλεπικαλυπτόμενους υπερκύβους σε λιγότερο διαχωρισμένες κατηγορίες.

Αντίθετα, μια μικρότερη παράμετρος χαλάρωσης αποδίδει

μικρότερους υπερκύβους, επιβάλλοντας αυστηρότερους περιορισμούς στην κεντροειδή κίνηση. Αυτή η ισορροπημένη προσέγγιση του C-K-Means επιτρέπει μια πιο λεπτή και προσαρμοστική διαδικασία ομαδοποίησης.

Στον κώδικα η παράμετρος χαλάρωσης έχει οριστεί σε 0.1 .

ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ

Η χρήση του αλγορίθμου ταξινόμησης Constrained K-Means (C-K-Means) σε τρία διαφορετικά σύνολα δεδομένων είναι απαραίτητη για τη συγκεκριμένη εργασία. Προκειμένου ο αλγόριθμος να λειτουργήσει όσο το δυνατόν καλύτερα και να παράγει αξιόπιστα αποτελέσματα, κάθε σύνολο δεδομένων πρέπει να προετοιμαστεί και να επεξεργαστεί προσεκτικά. Αυτή η προετοιμασία περιλαμβάνει την δημιουργία ενός script () για να διασφαλιστεί ότι τα σύνολα δεδομένων είναι συμβατά με τη μέθοδο C-K-Means, ιδίως όσον αφορά τη μορφοποίηση των δεδομένων και τη διαμόρφωση των παραμέτρων. Μέσα σε αυτό πραγματοποιείται και η κλήση της κλάσης CKMeans με αρχικοποιημένες παραμέτρους σύμφωνα με το κάθε σύνολο δεδομένων και οι συναρτήσεις εμφάνισης αποτελεσμάτων.

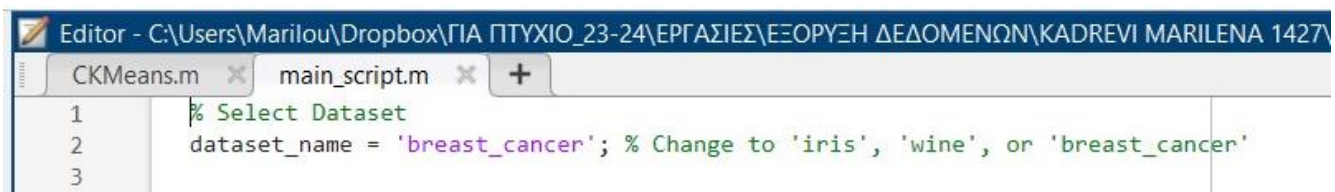
ΤΡΟΠΟΠΟΙΗΣΗ ΣΥΝΟΛΩΝ ΔΕΔΟΜΕΝΩΝ

Τα σύνολα δεδομένων πρέπει, πρώτα απ' όλα, να αναπαρίστανται αριθμητικά, επειδή οι υπολογισμοί του C-K-Means εξαρτώνται από αριθμητικές τιμές. Αυτό σημαίνει ότι όλα τα δεδομένα κατηγοριών πρέπει να κωδικοποιηθούν ή να μετατραπούν σε αριθμητική αναπαράσταση. Σε σύνολα δεδομένων όπου οι ετικέτες ή τα

χαρακτηριστικά δίνονται ως κείμενο ή κατηγορίες, θα πρέπει να χρησιμοποιούνται κωδικοποιημένα διανύσματα ή αριθμητικοί δείκτες.

Το σύνολο εκπαίδευσης και το σύνολο δοκιμής είναι τα δύο τελευταία υποσύνολα που δημιουργούνται από κάθε σύνολο δεδομένων. Στην παρούσα υλοποίηση, κάθε σύνολο δεδομένων διαχωρίστηκε 70% σε σετ εκπαίδευσης και 30% σε σετ δοκιμής.

Για να υλοποιηθεί ο αλγόριθμος ckmmeans σε κάθε διαφορετικό σύνολο δεδομένων θα πρέπει να τροποποιηθεί αναλόγως ο παρακάτω κώδικας.



```
Editor - C:\Users\Marilou\Dropbox\ΓΙΑ ΠΤΥΧΙΟ_23-24\ΕΡΓΑΣΙΕΣ\ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ\KADREVI MARILENA 1427\
CKMeans.m  x  main_script.m  x  +
1  % Select Dataset
2  dataset_name = 'breast_cancer'; % Change to 'iris', 'wine', or 'breast_cancer'
3
```

IRIS DATASET

Το πρώτο σύνολο δεδομένων που χρησιμοποιήθηκε ως βάση για αυτήν την υλοποίηση είναι το σύνολο δεδομένων Iris.

Μερικά χαρακτηριστικά για το σύνολο δεδομένων iris:

- Κάθε σειρά στο σύνολο δεδομένων αντιπροσωπεύει ένα φυτό ίριδας.
- Υπάρχουν 3 κατηγορίες, η καθεμία αντιστοιχεί σε ένα είδος φυτού ίριδας.
- Περιλαμβάνει 4 χαρακτηριστικά ανά παρουσία.
- Δεν υπάρχουν κενές τιμές.
- Οι αλφαριθμητικές ετικέτες μετατράπηκαν σε αριθμητικές: "Iris-setosa" ως 1, το "Iris-versicolor" ως 2 και το "Iris-virginica" ως 3.

Iris
Donated on 6/30/1988

A small classic dataset from Fisher, 1936. One of the earliest known datasets used for evaluating classification methods.

Dataset Characteristics	Subject Area	Associated Tasks
Tabular	Biology	Classification
Feature Type	# Instances	# Features
Real	150	4

Dataset Information

What do the instances in this dataset represent?
Each instance is a plant

Additional Information
This is one of the earliest datasets used in the literature on classification methods and widely used in statistics and machine learning. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are not linearly separable from each other....

SHOW MORE

Has Missing Values?
No

Keywords
ecology

Creators
R. A. Fisher

DOI
10.24432/C56C76

License
This dataset is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

WINE

DATASET

Το δεύτερο σύνολο δεδομένων που χρησιμοποιήθηκε ως βάση για αυτήν την υλοποίηση είναι το σύνολο δεδομένων wine.

Μερικά χαρακτηριστικά για το σύνολο δεδομένων wine:

- Όπως το σύνολο δεδομένων Iris, το σύνολο δεδομένων Wine είναι ένα πρόβλημα ταξινόμησης. Περιλαμβάνει την πρόβλεψη της ποικιλίας του κρασιού με βάση διάφορες χημικές αναλύσεις.
- Υπάρχουν 3 κατηγορίες, το καθένα αντιστοιχεί σε διαφορετικό τύπο ποικιλίας κρασιού.
- Περιλαμβάνει 13 χαρακτηριστικά ανά παρουσία.
- Δεν υπάρχουν κενές τιμές.

Wine

For what purpose was the dataset created?
test

Additional Information
These data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.

I think that the initial data set had around 30 variables, but for some reason I only have the 13 dimensional version. I had a list of what the 30 or so variables were, but a.) I lost it, and b.), I would not know which 13 variables are included in the set.

The attributes are (donated by Riccardo Leardi, rleardi@anchem.unige.it)

- 1) Alcohol
- 2) Malic acid
- 3) Ash
- 4) Alkalinity of ash
- 5) Magnesium
- 6) Total phenols
- 7) Flavonoids
- 8) Nonflavanoid phenols
- 9) Proanthocyanins
- 10) Color intensity
- 11) Hue
- 12) OD280/OD315 of diluted wines
- 13) Proline

In a classification context, this is a well posed problem with "well behaved" class structures. A good data set for first testing of a new classifier, but not very challenging.

SHOW LESS

Has Missing Values?
No

Keywords
Chemistry

Creators
Stefan Aeberhard
M. Forina

DOI
10.24432/C5PC7J

License
This dataset is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.


This allows for the sharing and adaptation of the datasets for any purpose, provided that the appropriate credit is given.

BREAST CANCER WISCONSIN (DIAGNOSTIC) DATASET

Το τρίτο σύνολο δεδομένων που χρησιμοποιήθηκε ως βάση για αυτήν την υλοποίηση είναι το σύνολο δεδομένων breast cancer.

Μερικά χαρακτηριστικά για το σύνολο δεδομένων breast cancer.:

- Όπως το σύνολο δεδομένων Iris, το σύνολο δεδομένων breast cancer είναι ένα πρόβλημα ταξινόμησης. Περιλαμβάνει την πρόβλεψη εάν ένας όγκος καρκίνου του μαστού είναι καλοήθης ή κακοήθης
- Υπάρχουν 2 κατηγορίες, καλοήθεις και κακοήθεις.
- Περιλαμβάνει 30 χαρακτηριστικά ανά παρουσία.
- Δεν υπάρχουν κενές τιμές.

**Breast Cancer Wisconsin (Diagnostic)**
Donated on 10/31/1995

Diagnostic Wisconsin Breast Cancer Database.

Dataset Characteristics	Subject Area	Associated Tasks
Multivariate	Health and Medicine	Classification
Feature Type	# Instances	# Features
Real	569	30

Dataset Information

Additional Information

Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image. A few of the images can be found at <http://www.cs.wisc.edu/~street/images/>


Separating plane described above was obtained using Multisurface Method-Tree (MSM-T) [K. P. Bennett, "Decision Tree Construction Via Linear Programming," Proceedings of the 4th Midwest Artificial Intelligence and Cognitive Science Society, pp. 97-101, 1992], a classification method which uses linear programming to construct a decision tree. Relevant features were selected using an exhaustive search in the space of 1-4 features and 1-3 separating planes.

The actual linear program used to obtain the separating plane in the 3-dimensional space is that described in: [K. P. Bennett and O. L. Mangasarian: "Robust Linear Programming Discrimination of Two Linearly Inseparable Sets", Optimization Methods and Software 1, 1992, 23-34].

This database is also available through the UW CS ftp server:
ftp ftp.cs.wisc.edu
cd math-prog/cpo-dataset/machine-learn/WDBC/
[SHOW LESS](#)

Has Missing Values?
No

DOWNLOAD

 **IMPORT IN PYTHON**

CITE

37 citations
256355 views

Keywords
[health](#) [cancer](#)

Creators

- William Wolberg
- Olvi Mangasarian
- Nick Street
- W. Street

DOI
[10.24432/CSDW2B](https://doi.org/10.24432/CSDW2B)

License

This dataset is licensed under a [Creative Commons Attribution 4.0 International](#) (CC BY 4.0) license.

This allows for the sharing and adaptation of the datasets for any purpose, provided that the appropriate credit is given.

ΑΠΟΤΕΛΕΣΜΑΤΑ ΥΛΟΠΟΙΗΣΗΣ C-K-Means (MATLAB)

IRIS DATA SET

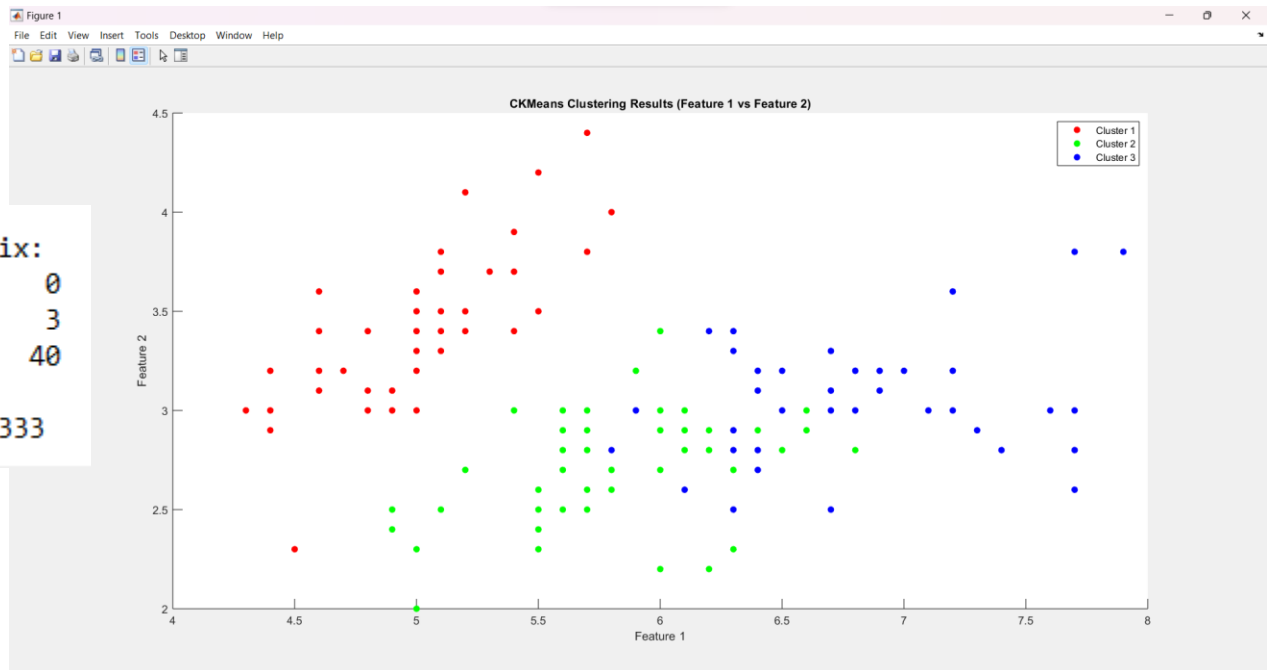
Ακολουθούν τα αποτελέσματα του αλγορίθμου CK-Means του συνόλου δεδομένων Iris:

Τα δείγματα ίριδας διαχωρίστηκαν επιτυχώς σε συστάδες από τον αλγόριθμο. Σε αυτή τη συγκεκριμένη εκτέλεση παρήγαγε τρεις συστάδες, οι οποίες αντιστοιχούν στα τρία είδη ανθών ίριδας που είναι γνωστό ότι υπάρχουν. Τα περισσότερα δείγματα ίριδας ταξινομήθηκαν κατάλληλα στις αντίστοιχες συστάδες τους, σύμφωνα με τον πίνακα σύγχυσης. Η υψηλή βαθμολογία ακρίβειας του αλγορίθμου, περίπου 91,33%, υποδηλώνει ότι μπορεί να εντοπίσει σημαντικά μοτίβα στα δεδομένα.

Confusion Matrix:

50	0	0
0	47	3
0	10	40

Accuracy: 0.91333



WINE DATA SET

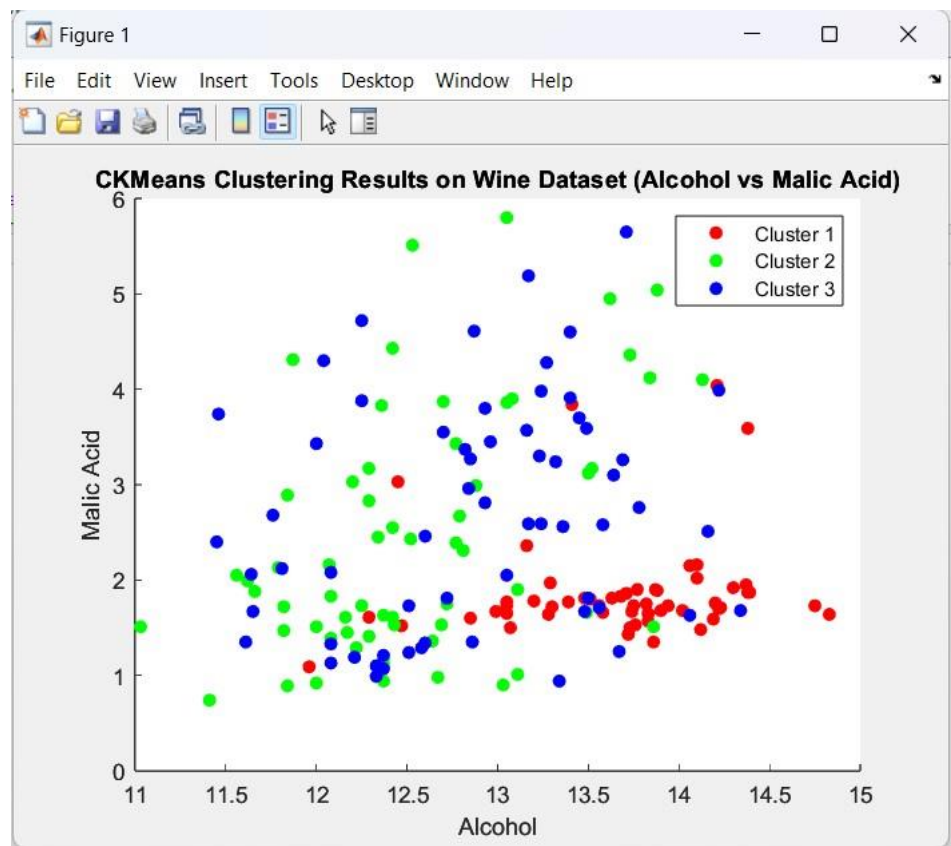
Ακολουθούν τα αποτελέσματα του αλγορίθμου CK-Means του συνόλου δεδομένων wine :

Τα δείγματα κρασιού διαχωρίστηκαν επιτυχώς σε συστάδες από τον αλγόριθμο. Σε αυτή τη συγκεκριμένη εκτέλεση παράγαγε τρεις συστάδες, οι οποίες αντιστοιχούν στις 3 κατηγορίες κρασιού. Η βαθμολογία ακρίβειας του αλγορίθμου, περίπου 70,78%, υποδηλώνει ότι μπορεί να εντοπίσει σημαντικά μοτίβα στα δεδομένα.

Confusion Matrix:

50	0	9
4	45	22
1	16	31

Accuracy: 0.70787



Ακολουθούν τα αποτελέσματα του αλγορίθμου CK-Means του συνόλου δεδομένων breast cancer :

Τα δείγματα από το σύνολο δεδομένων breast cancer διαχωρίστηκαν επιτυχώς σε συστάδες από τον αλγόριθμο. Σε αυτή τη συγκεκριμένη εκτέλεση παράγαγε δύο συστάδες, οι οποίες αντιστοιχούν στις 2 κατηγορίες - καλοήθειες και κακοήθειες. Η βαθμολογία ακρίβειας του αλγορίθμου, περίπου 90%, υποδηλώνει ότι μπορεί να εντοπίσει σημαντικά μοτίβα στα δεδομένα.

Confusion Matrix:

351	6
51	161

Accuracy: 0.89982

