Πανεπιστήμιο Δυτικής Αττικής

Σχολή Μηχανικών

Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών

ΓΟΥΡΔΟΜΙΧΑΛΗΣ ΔΗΜΗΤΡΙΟΣ

ice20390043

ΕΡΓΑΣΤΗΡΙΟ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΤΥΠΩΝ ΚΑΙ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

ΕΡΓΑΣΤΗΡΙΑΚΗ ΑΣΚΗΣΗ #2

Τμήμα 1

ΧΕΙΜΕΡΙΝΟ ΕΞΑΜΗΝΟ 2023-2024

Εισαγωγή

Για την εκπόνηση της συγκεκριμένης εργασίας χρησιμοποιήθηκαν σημειώσεις-ιδέες από τα παρακάτω αρχεία που βρίσκονται στο eclass του εργαστηριακού μαθήματος:

MatlabStatistics.pdf, StatisticsWithMatlabOctave.pdf, Sirish_Shah_CCA_Lecture_1B.pdf.

Επίσης, προηγήθηκε ανάγνωση των σημειώσεων του θεωρητικού μέρους, ώστε να γίνουν κατανοητές οι θεωρητικές έννοιες που συναντούμε στην εργασία.

Σε όλη τη διάρκεια της εργασίας χρησιμοποιήθηκε το Matlab και ειδικότερα τα toolboxes: statistics and machine learning toolbox , curve fitting toolbox .

Μέρος Α. ΤΟ ΠΕΡΙΒΑΛΛΟΝ MATLAB ΓΙΑ ΕΦΑΡΜΟΓΕΣ Α.Π.Μ.Μ.

<u>Α1. Άσκηση #2 Κλήση συναρτήσεων</u>

Η συγκεκριμένη άσκηση λύνεται με 2 τρόπους. Αρχικά, με το γνωστό τρόπο, με σύνταξη του κώδικα σε ένα αρχείο .m. Στη συνέχεια, για τη λύση του ίδιου προβλήματος, δημιουργούμε μία συνάρτηση, και τρέχουμε το αρχείο(με όποιες παραμέτρους επιλέξουμε) στο Command Window.

Συνοπτικά, διαθέτουμε 40 νομίσματα των 50 λεπτών, 40 νομίσματα του 1€, 40 νομίσματα των 2€, και 40 χαρτονομίσματα των 5€. Το πρόγραμμα υπολογίζει και εμφανίζει όλους τους δυνατούς συνδυασμούς με τους οποίους μπορούμε να χρησιμοποιήσουμε 40 συνολικά νομίσματα-χαρτονομίσματα έτσι ώστε η συνολική τους αξία να είναι 40€.

Κώδικας:

ask1_1.m

```
clear all;

clc;

close all;


% x το πλήθος των νομισμάτων με αξία 0.5 ευρώ.

% y το πλήθος των νομισμάτων με αξία 1.0 ευρώ.

% z το πλήθος των νομισμάτων με αξία 2.0 ευρώ.

% w το πλήθος των χαρτονομισμάτων με αξία 5 ευρώ.


% Αρχικοποίηση του πλήθους των συνδυασμών που ικανοποιούν τις

% συνθήκες

PS = 0;


for x = 0:40

  for y = 0:40

    for z = 0:40

      for w = 0:40


      syn_plithos_nom = x + y + z + w;

      syn_aksia_nom = x * 0.5 + y * 1 + z * 2 + w * 5;
```

```
            if syn_plithos_nom == 40 && syn_aksia_nom == 40
                PS = PS + 1;

                % Παρουσίαση συνδυασμών μέσω γραφήματος(bar)
                bar(PS, [x, y, z, w], 'stacked');
                hold on;

                disp(['Combination #' num2str(PS) ': ' ...
                    num2str(x) 'x0.5E ' ...
                    num2str(y) 'x1E ' num2str(z) 'x2E ' ...
                    num2str(w) 'x5E ']);
            end
        end
    end
end
end
disp(['Total number of combinations: ' num2str(PS)]);

title('Κατανομή νομισμάτων και χαρτονομισμάτων σε κάθε συνδυασμό');
xlabel('Συνδυασμός');
ylabel('Ποσότητα');
legend('0.5E', '1E', '2E', '5E');
```

Αποτελέσματα:

Combination #1: 0x0.5E 40x1E 0x2E 0x5E

Combination #2: 2x0.5E 37x1E 1x2E 0x5E

Combination #3: 4x0.5E 34x1E 2x2E 0x5E

Combination #4: 6x0.5E 31x1E 3x2E 0x5E

Combination #5: 8x0.5E 28x1E 4x2E 0x5E

Combination #6: 8x0.5E 31x1E 0x2E 1x5E

Combination #7: 10x0.5E 25x1E 5x2E 0x5E

Combination #8: 10x0.5E 28x1E 1x2E 1x5E

Combination #9: 12x0.5E 22x1E 6x2E 0x5E

Combination #10: 12x0.5E 25x1E 2x2E 1x5E

Combination #11: 14x0.5E 19x1E 7x2E 0x5E

Combination #12: 14x0.5E 22x1E 3x2E 1x5E

Combination #13: 16x0.5E 16x1E 8x2E 0x5E

Combination #14: 16x0.5E 19x1E 4x2E 1x5E

Combination #15: 16x0.5E 22x1E 0x2E 2x5E

Combination #16: 18x0.5E 13x1E 9x2E 0x5E

Combination #17: 18x0.5E 16x1E 5x2E 1x5E

Combination #18: 18x0.5E 19x1E 1x2E 2x5E

Combination #19: 20x0.5E 10x1E 10x2E 0x5E

Combination #20: 20x0.5E 13x1E 6x2E 1x5E

Combination #21: 20x0.5E 16x1E 2x2E 2x5E

Combination #22: 22x0.5E 7x1E 11x2E 0x5E

Combination #23: 22x0.5E 10x1E 7x2E 1x5E

Combination #24: 22x0.5E 13x1E 3x2E 2x5E

Combination #25: 24x0.5E 4x1E 12x2E 0x5E

Combination #26: 24x0.5E 7x1E 8x2E 1x5E

Combination #27: 24x0.5E 10x1E 4x2E 2x5E

Combination #28: 24x0.5E 13x1E 0x2E 3x5E

Combination #29: 26x0.5E 1x1E 13x2E 0x5E

Combination #30: 26x0.5E 4x1E 9x2E 1x5E

Combination #31: 26x0.5E 7x1E 5x2E 2x5E

Combination #32: 26x0.5E 10x1E 1x2E 3x5E

Combination #33: 28x0.5E 1x1E 10x2E 1x5E

Combination #34: 28x0.5E 4x1E 6x2E 2x5E

Combination #35: 28x0.5E 7x1E 2x2E 3x5E

Combination #36: 30x0.5E 1x1E 7x2E 2x5E

Combination #37: 30x0.5E 4x1E 3x2E 3x5E

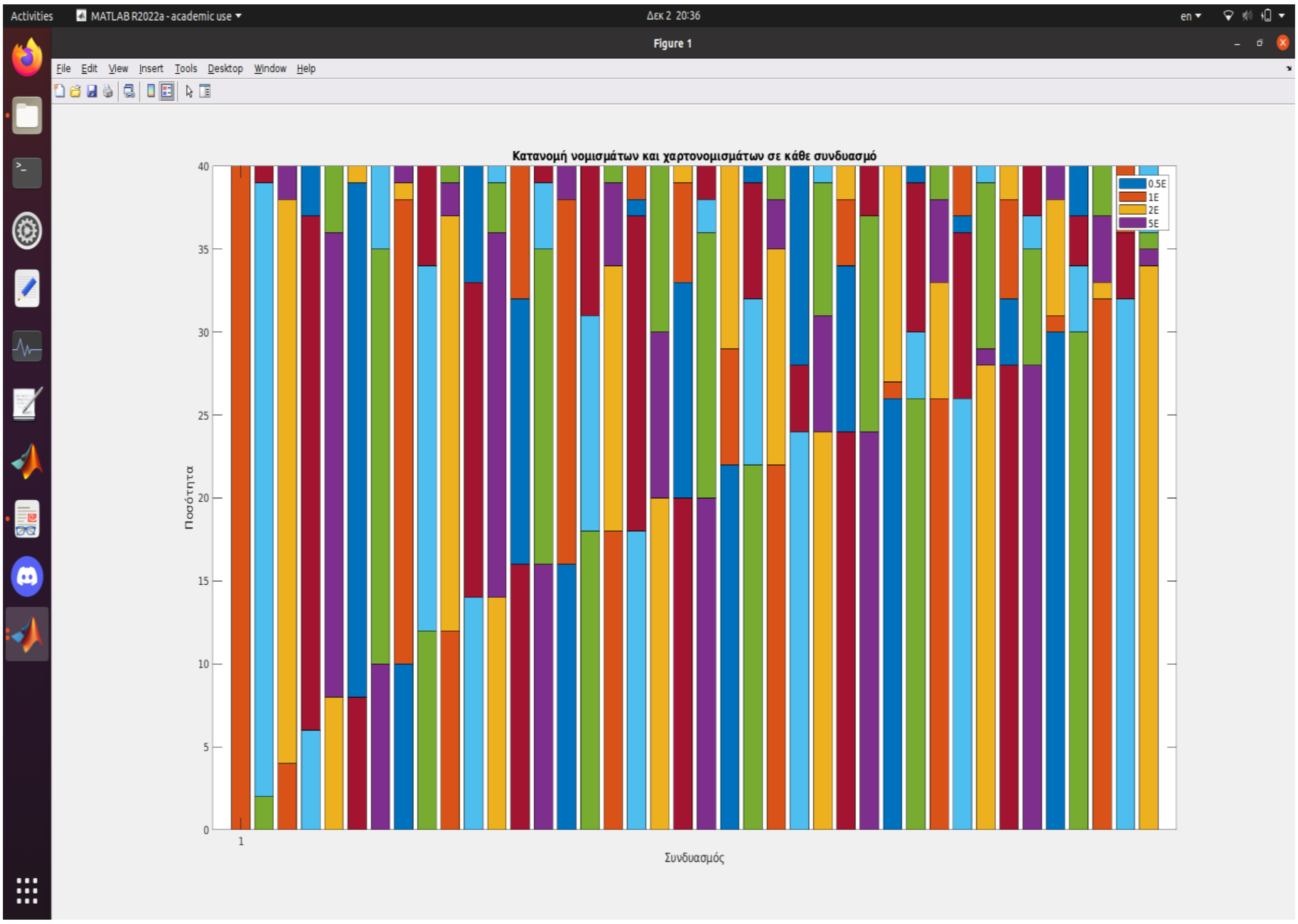
Combination #38: 32x0.5E 1x1E 4x2E 3x5E

Combination #39: 32x0.5E 4x1E 0x2E 4x5E

Combination #40: 34x0.5E 1x1E 1x2E 4x5E

Total number of combinations: 40

Γραφήματα:



Κώδικας:

ask1_1f.m

function PS=ask1_1f(val1, val2, val3, val4)

% x το πλήθος των νομισμάτων με αξία 0.5 ευρώ.

% y το πλήθος των νομισμάτων με αξία 1.0 ευρώ.

```
% z το πλήθος των νομισμάτων με αξία 2.0 ευρώ.
% w το πλήθος των χαρτονομισμάτων με αξία 5 ευρώ.

% Αρχικοποίηση του πλήθους των συνδυασμών που ικανοποιούν τις
% συνθήκες
PS = 0;

for x = 0:40
  for y = 0:40
    for z = 0:40
      for w = 0:40

      syn_plithos_nom=x+y+z+w;
      syn_aksia_nom=x*val1+y*val2+z*val3+w*val4;

        if syn_plithos_nom == 40 && syn_aksia_nom == 40
          PS = PS + 1;

          % Παρουσίαση συνδυασμών μέσω γραφήματος(bar)
          %bar(PS, [x, y, z, w], 'stacked');
          %hold on;

          disp(['Combination #' num2str(PS) ': ' ...
            num2str(x) 'x0.5E ' ...
            num2str(y) 'x1E ' num2str(z) 'x2E ' ...
            num2str(w) 'x5E ']);
        end
      end
    end
  end
end
```

```
disp(['Total number of combinations: ' num2str(PS)]);
end
```

Αποτελέσματα:

Combination #1: 0x0.5E 40x1E 0x2E 0x5E

Combination #2: 2x0.5E 37x1E 1x2E 0x5E

Combination #3: 4x0.5E 34x1E 2x2E 0x5E

Combination #4: 6x0.5E 31x1E 3x2E 0x5E

Combination #5: 8x0.5E 28x1E 4x2E 0x5E

Combination #6: 8x0.5E 31x1E 0x2E 1x5E

Combination #7: 10x0.5E 25x1E 5x2E 0x5E

Combination #8: 10x0.5E 28x1E 1x2E 1x5E

Combination #9: 12x0.5E 22x1E 6x2E 0x5E

Combination #10: 12x0.5E 25x1E 2x2E 1x5E

Combination #11: 14x0.5E 19x1E 7x2E 0x5E

Combination #12: 14x0.5E 22x1E 3x2E 1x5E

Combination #13: 16x0.5E 16x1E 8x2E 0x5E

Combination #14: 16x0.5E 19x1E 4x2E 1x5E

Combination #15: 16x0.5E 22x1E 0x2E 2x5E

Combination #16: 18x0.5E 13x1E 9x2E 0x5E

Combination #17: 18x0.5E 16x1E 5x2E 1x5E

Combination #18: 18x0.5E 19x1E 1x2E 2x5E

Combination #19: 20x0.5E 10x1E 10x2E 0x5E

Combination #20: 20x0.5E 13x1E 6x2E 1x5E

Combination #21: 20x0.5E 16x1E 2x2E 2x5E

Combination #22: 22x0.5E 7x1E 11x2E 0x5E

Combination #23: 22x0.5E 10x1E 7x2E 1x5E

Combination #24: 22x0.5E 13x1E 3x2E 2x5E

Combination #25: 24x0.5E 4x1E 12x2E 0x5E

Combination #26: 24x0.5E 7x1E 8x2E 1x5E

Combination #27: 24x0.5E 10x1E 4x2E 2x5E

Combination #28: 24x0.5E 13x1E 0x2E 3x5E

Combination #29: 26x0.5E 1x1E 13x2E 0x5E

Combination #30: 26x0.5E 4x1E 9x2E 1x5E

Combination #31: 26x0.5E 7x1E 5x2E 2x5E

Combination #32: 26x0.5E 10x1E 1x2E 3x5E

Combination #33: 28x0.5E 1x1E 10x2E 1x5E

Combination #34: 28x0.5E 4x1E 6x2E 2x5E

Combination #35: 28x0.5E 7x1E 2x2E 3x5E

Combination #36: 30x0.5E 1x1E 7x2E 2x5E

Combination #37: 30x0.5E 4x1E 3x2E 3x5E

Combination #38: 32x0.5E 1x1E 4x2E 3x5E

Combination #39: 32x0.5E 4x1E 0x2E 4x5E

Combination #40: 34x0.5E 1x1E 1x2E 4x5E

Total number of combinations: 40

A2. Χειρισμός διανυσμάτων και πινάκων.

Κώδικας:

Για το συγκεκριμένο παράδειγμα συμβουλευόμαστε αρχικά το αρχείο «Εισαγωγή στο Matlab». Στον κώδικα, επιλέγουμε τις βασικές πράξεις-λειτουργίες των πινάκων και παρουσιάζουμε τα αποτελέσματα της επεξεργασίας τους. Επιπροσθέτως, παράγουμε μία ομάδα τυχαίων αριθμών. Ταξινομούμε το συγκεκριμένο σύνολο και υπολογίζουμε τα εξής: Μέση τιμή, ενδιάμεση τιμή, Διασπορά και Τυπική Απόκλιση.

ask1_2.m

```
% Χειρισμός - Πράξεις Πινάκων

A=[1 2 3; 5 7 9];

B=[-1 0 1; 4 6 0];

J=[-1 0; 6 0];


disp('Πίνακας A:');

disp(A);


disp('Πίνακας B:');
```

```
disp(B);

C = A + B;
disp('Πρόσθεση πινάκων - Πίνακας C:');
disp(C);

D = A .* B;
disp('Πολλαπλασιασμός πινάκων - Πίνακας D: ');
disp(D);

E = B';
disp('Αντιστροφή του πίνακα B - Πίνακας E :');
disp(E);

disp('Πίνακας J:');
disp(J);

Y = fliplr(D);
disp('Πίνακας Y - Αντίστροφος του Πίνακα J :');
disp(Y);

disp(['Αποτέλεσμα της det(J): ' num2str(det(J))]);

disp(['Διαστάσεις του πίνακα C: ' num2str(size(C))]);

disp(['Έλεγχος ύπαρξης στοιχείων στον πίνακα E: ' num2str(isempty(E))]);

% Χρήση βασικών συναρτήσεων για την κατανόηση εννοιών:
% Παραγωγή τυχαίων αριθμών
random_data = randn(20,1);
disp('Τυχαίο σύνολο δεδομένων:');
disp(random_data);
```

```
disp('Ταξινόμηση συνόλου δεδομένων:');
disp(sort(random_data));

disp('Εύρεση μέσης τιμής:');
disp(median(random_data));

disp('Εύρεση ενδιάμεσης τιμής:');
disp(mean(random_data));

variance = var(random_data);
disp('Υπολογισμός διασποράς:');
disp(variance);

ta = std(random_data);
disp('Υπολογισμός τυπικής απόκλισης:');
disp(ta);
```

Αποτελέσματα:

```
Πίνακας A:
     1     2     3
     5     7     9

Πίνακας B:
    -1     0     1
     4     6     0

Πρόσθεση πινάκων - Πίνακας C:
     0     2     4
     9    13     9

Πολλαπλασιασμός πινάκων - Πίνακας D:
```

-1 0 3
20 42 0

Αντιστροφή του πίνακα B - Πίνακας E :

-1 4
0 6
1 0

Πίνακας J:

-1 0
6 0

Πίνακας Y - Αντίστροφος του Πίνακα J :

3 0 -1
0 42 20

Αποτέλεσμα της det(J): 0

Διαστάσεις του πίνακα C: 2 3

Έλεγχος ύπαρξης στοιχείων στον πίνακα E: 0

Τυχαίο σύνολο δεδομένων:

1.2008
1.0902
-0.3587
-0.1299
0.7337
0.1203
1.1363
-0.6868
0.4717
0.2883
1.3919
-1.3455

0.0007

0.0535

-2.3419

1.2481

2.8092

-0.2322

0.2870

-0.4646

Ταξινόμηση συνόλου δεδομένων:

-2.3419

-1.3455

-0.6868

-0.4646

-0.3587

-0.2322

-0.1299

0.0007

0.0535

0.1203

0.2870

0.2883

0.4717

0.7337

1.0902

1.1363

1.2008

1.2481

1.3919

2.8092

Εύρεση μέσης τιμής:

0.2037

Εύρεση ενδιάμεσης τιμής:

0.2636

Υπολογισμός διασποράς:

1.2129

Υπολογισμός τυπικής απόκλισης:

1.1013

Α3. Άσκηση #2 Παραγωγή τυχαίων αριθμών.

Παράγουμε τυχαία δεδομένα από μία διακριτή κατανομή πιθανότητας(για παράδειγμα ρίξιμο ζαριού) και από μία συνεχή(για παράδειγμα παρατήρηση της διάρκειας ζωής μιας λάμπας).

Διακριτή κατανομή: διωνυμική με χρήση της συνάρτησης randi.

Συνεχής κατανομή: κανονική με χρήση της συνάρτησης randn.

Τα σχήματα των δύο κατανομών παρουσιάζονται σε κοινό γράφημα 2 παραθύρων.

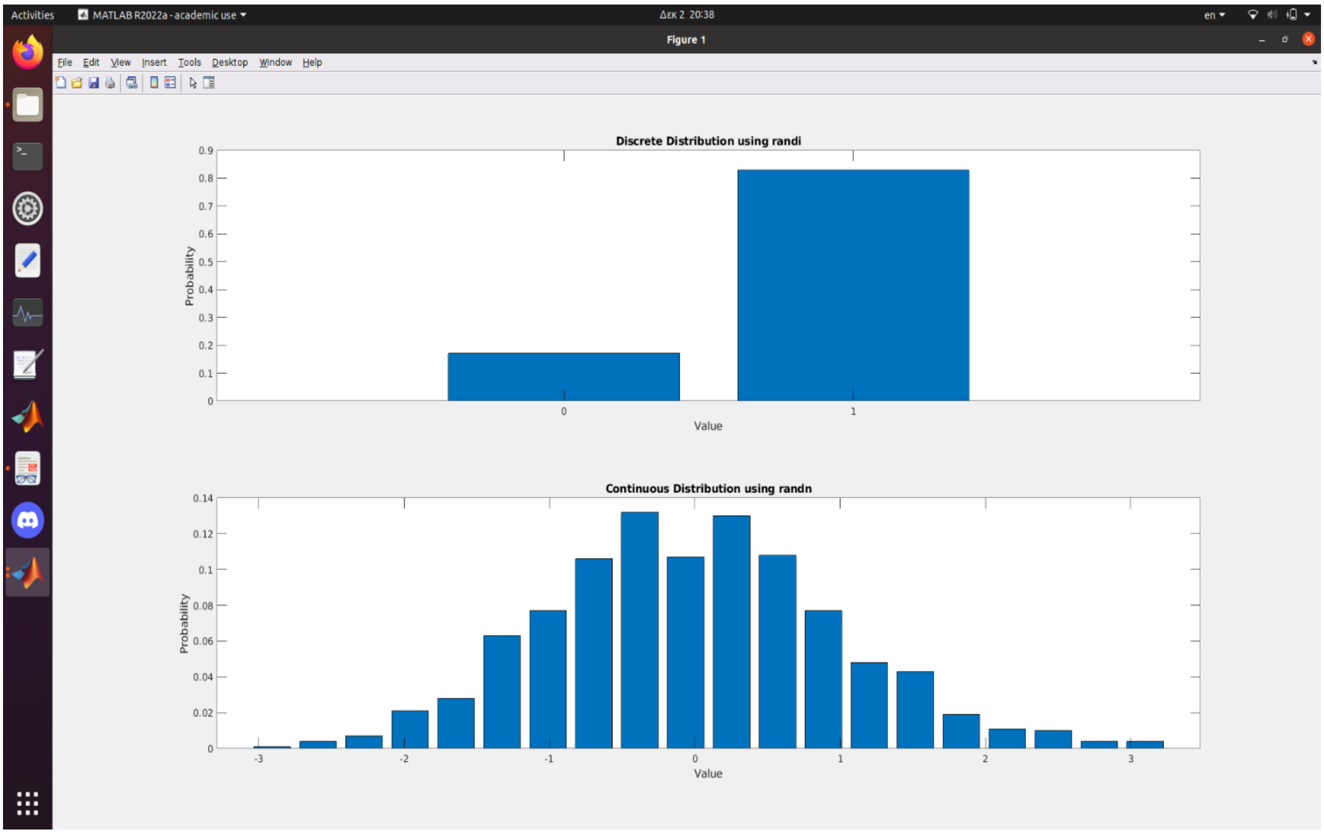
Κώδικας:

ask1_3.m

```
% Παραγωγή τυχαίων δεδομένων

% Παραμέτροι για τη διακριτή κατανομή (διωνυμική)

n = 9;

p = 0.8;

N = 1000;


% Δημιουργία τυχαίων δεδομένων από διωνυμική κατανομή με χρήση της randi

data_discrete = randi([0, n], N, 1);

data_discrete = sum(data_discrete <= n * p, 2);


% Παραμέτροι για τη συνεχή κατανομή (κανονική)

m = 0;
```

```
s = 1;
N_continuous = 1000;

% Δημιουργία τυχαίων δεδομένων από κανονική κατανομή με χρήση της randn
data_continuous = m + s * randn(N_continuous, 1);

% Ιστόγραμμα για τη διακριτή κατανομή
[height_discrete, centers_discrete] = hist(data_discrete, unique(data_discrete));

% Ιστόγραμμα για τη συνεχή κατανομή
% Χρήση 20 bins
[height_continuous, centers_continuous] = hist(data_continuous, 20);

figure;
% Διακριτή κατανομή
subplot(2, 1, 1);
bar(centers_discrete, height_discrete / sum(height_discrete));
xlabel('Value');
ylabel('Probability'); using
title('Discrete Distribution randi');

% Συνεχής κατανομή
subplot(2, 1, 2);
bar(centers_continuous, height_continuous / sum(height_continuous));
xlabel('Value');
ylabel('Probability');
title('Continuous Distribution using randn');
```

Γραφήματα:



A4. Άσκηση #2 Data Sets.

Για την εκπόνηση της συγκεκριμένης άσκησης χρησιμοποιήσαμε 4 data sets, από το σύνδεσμο: [MATLAB Example Data Sets - MATLAB & Simulink (mathworks.com)](), με διαφορετικά χαρακτηριστικά ως προς τον τύπο και τον αριθμό μεταβλητών. Η επεξεργασία των παραπάνω data sets επιτεύχθηκε με τη βοήθεια των εργαλείων διαχείρισης δεδομένων του Matlab.

Οι συγκεκριμένες συλλογές δεδομένων χρησιμοποιούνται για την αξιολόγηση μοντέλων μηχανικής μάθησης.

Ειδικότερα, επιλέξαμε τα εξής data sets:

1. **Geographic Data: tsunamis.xlsx**
2. **Observational Data: outages.csv**
3. **Image Data: flujet.mat**
4. **Video and Audio Data: train.mat**

1. **Geographic Data: tsunamis.xlsx**

Κώδικας για φόρτωση των δεδομένων:

ask1_4_2.m

T2 = readtable('tsunamis.xlsx');

geobubble(T.Latitude,T.Longitude,T.MaxHeight)

% Αρχεία που δημιουργούνται(auto-generated files):ClassificationLearnerSession_tsunamis.mat ,

% trainClassifier2.m ,

% session2.sfit ,

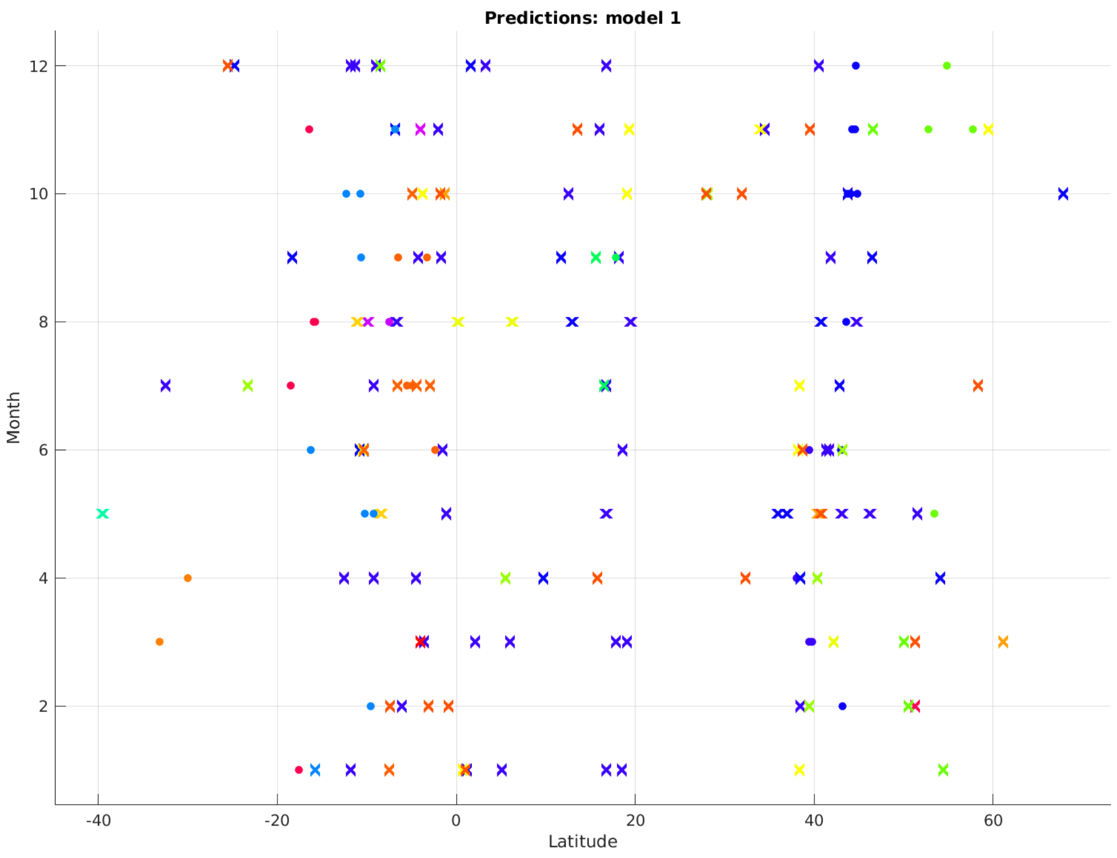
% createFittsunamis.m

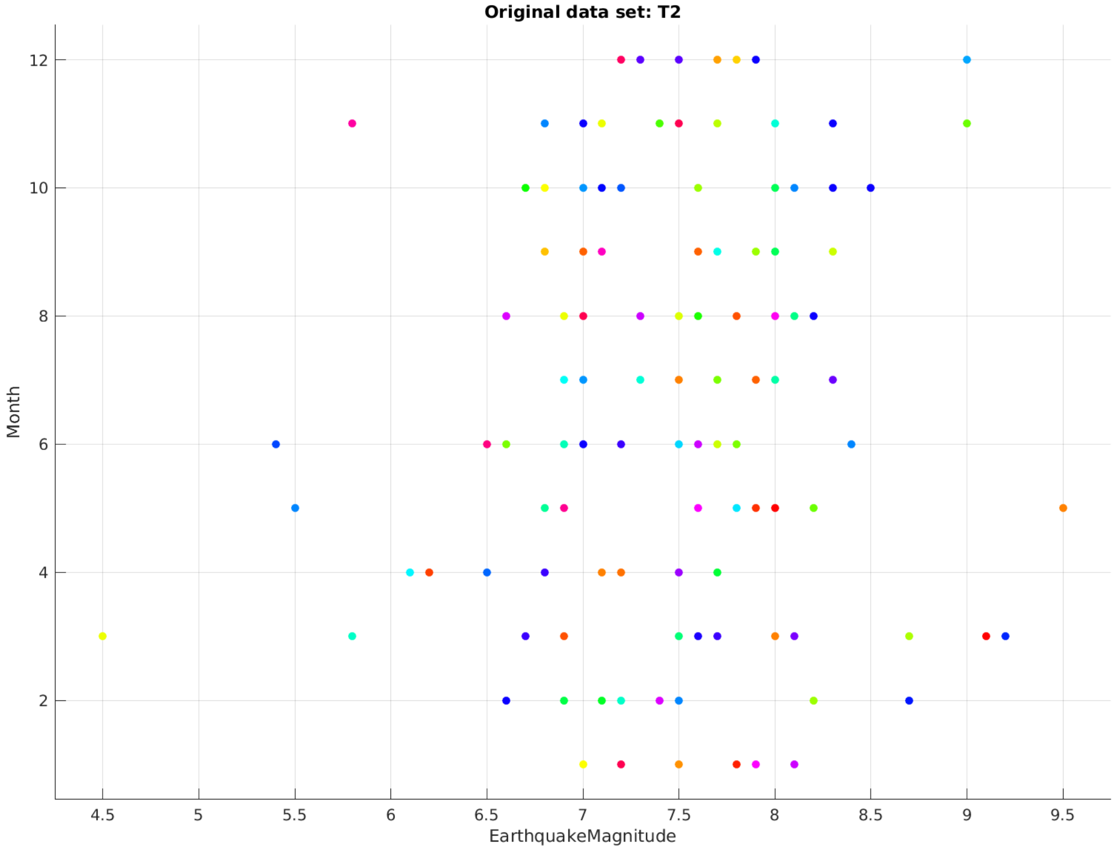
Βήματα:

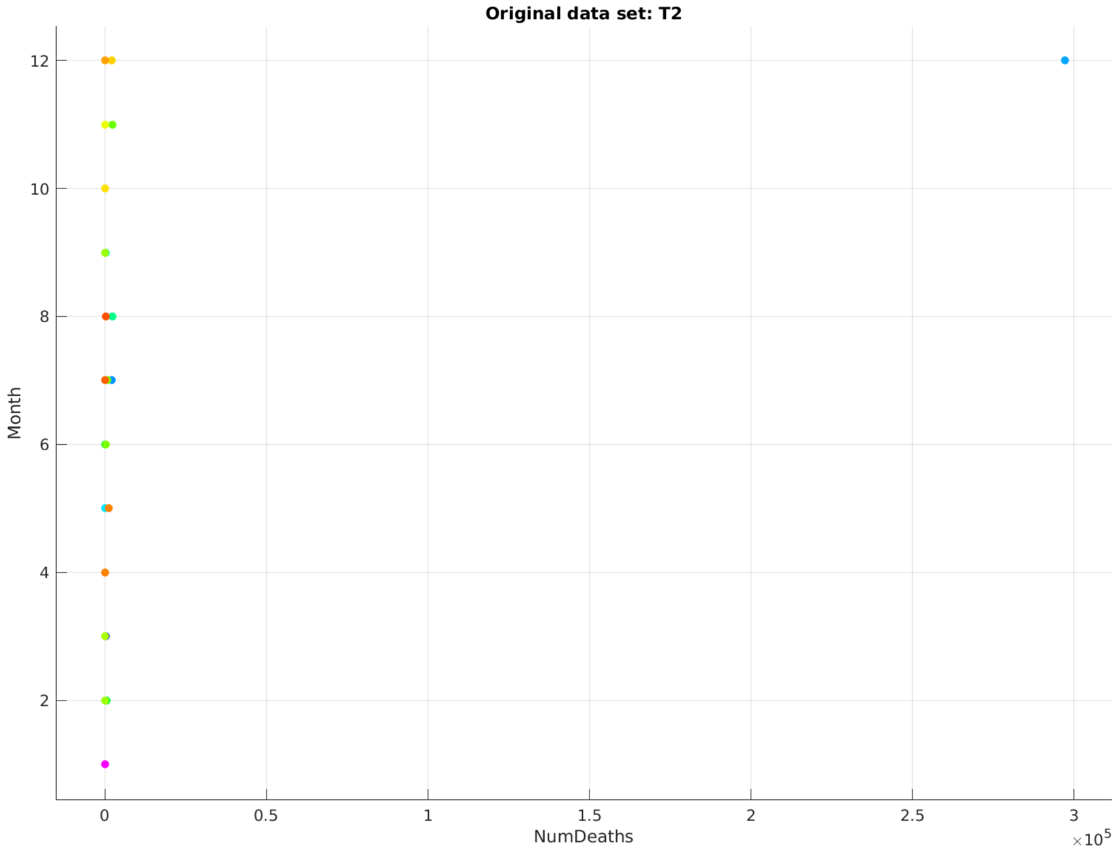
Μετά τη φόρτωση των δεδομένων στο Matlab, ακολουθούν τα εξής:

Apps → **Classification Learner**(εργαλείο επεξεργασίας datasets που δείξαμε και κατά τη διάρκεια του εργαστηριακού μαθήματος) → open → ClassificationLearnerSession_tsunamis.mat → train all(εμφάνιση των μοντέλων στην αριστερή πλευρά της οθόνης και του accuracy του κάθε μοντέλου για τα συγκεκριμένα δεδομένα) → Tree → Scatter plots →

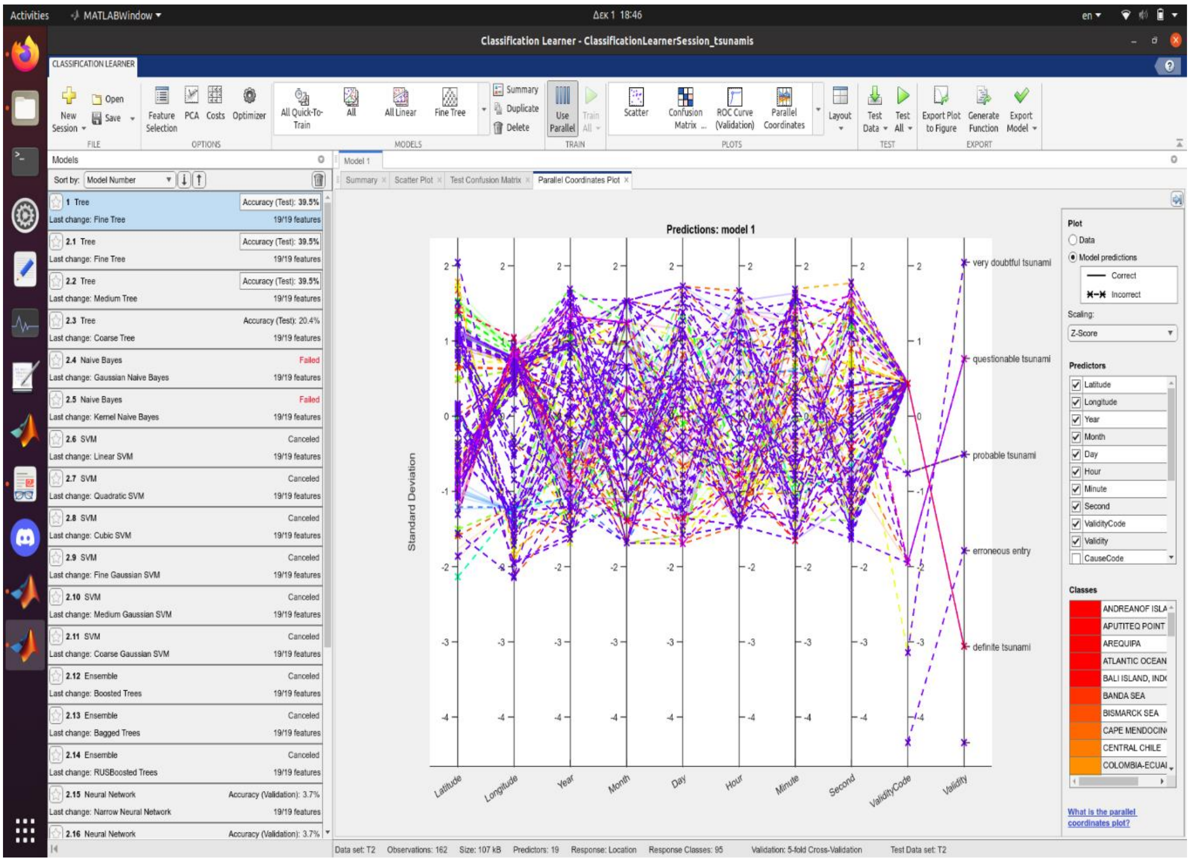

Original data set: T2
Month
Latitude
12
10
8
6
4
2
-40
-20
0
20
40
60


Predictions: model 1
Month
Latitude
12
10
8
6
4
2
-40
-20
0
20
40
60

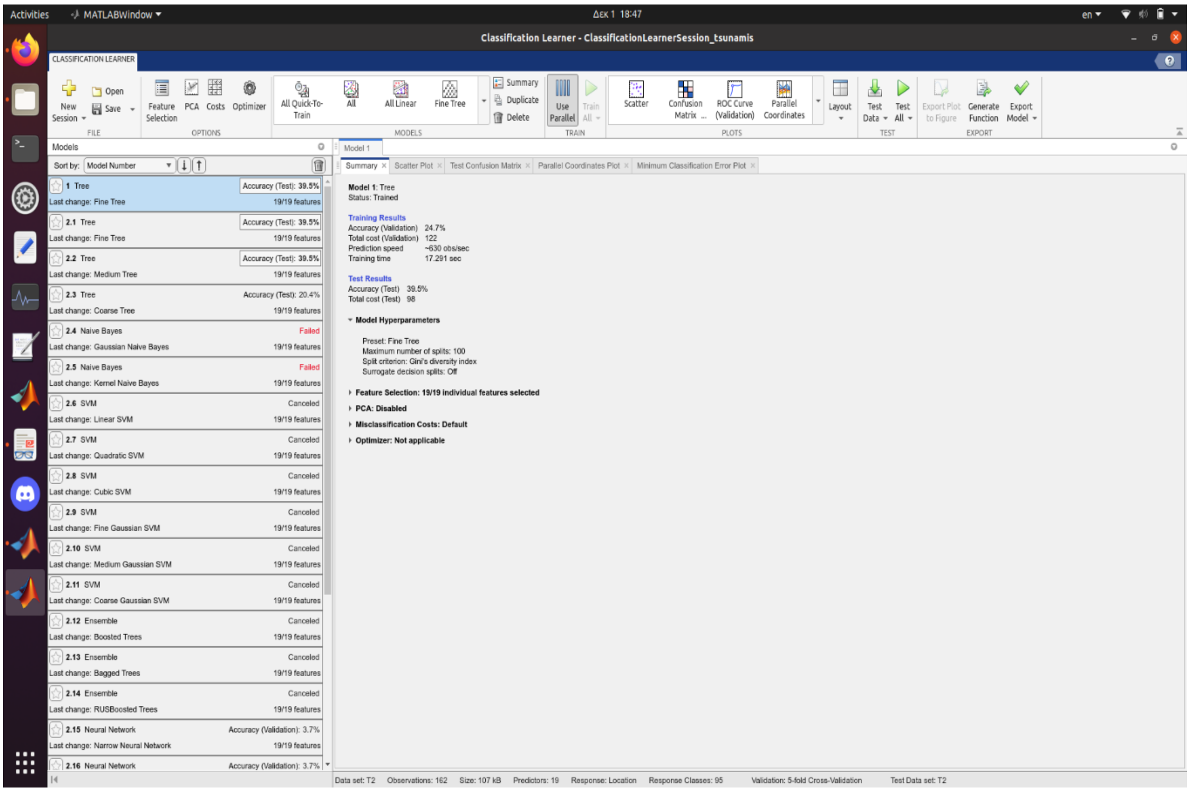

Original data set: T2
Month
EarthquakeMagnitude
12
10
8
6
4
2
4.5
5
5.5
6
6.5
7
7.5
8
8.5
9
9.5


Original data set: T2
Month
NumDeaths
12
10
8
6
4
2
0
0.5
1
1.5
2
2.5
3
×10^5

→ Parallel Coordinates Plot →



→ Train Results →



→ Generate Function → Παράγεται το αρχείο trainClassifier2.m

Apps → **Curve Fitter** → Select Data → open → session2.sfit

→ Select Fitting Data:

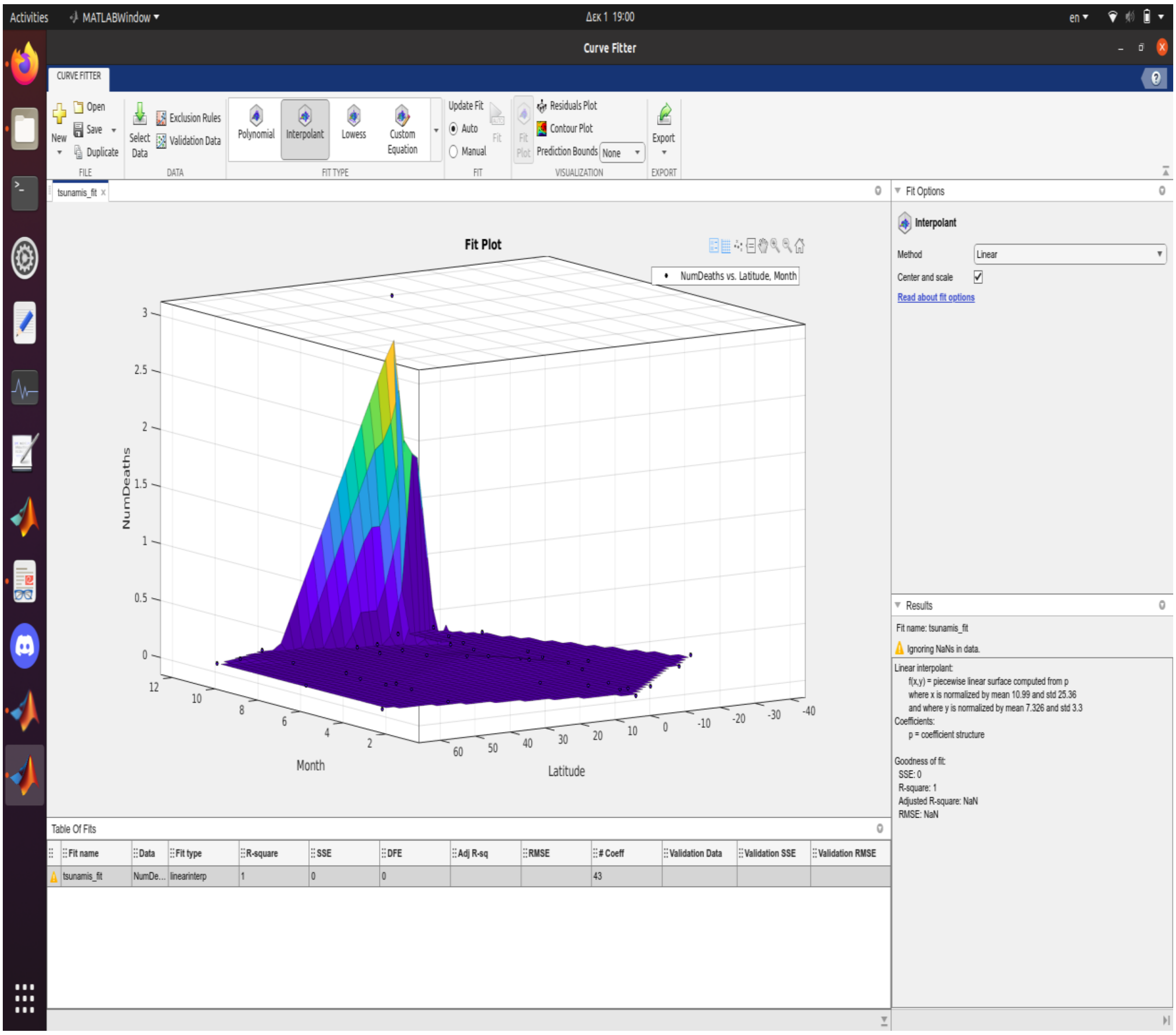
Fit name: tsunamis_fit

X Data:T2, select variable: Latitude

Y Data:T2, select variable: Month

Z Data:T2, select variable: NumDeaths

→ Export → Generate Code : Παράγεται το αρχείο createFittsunamis.m →

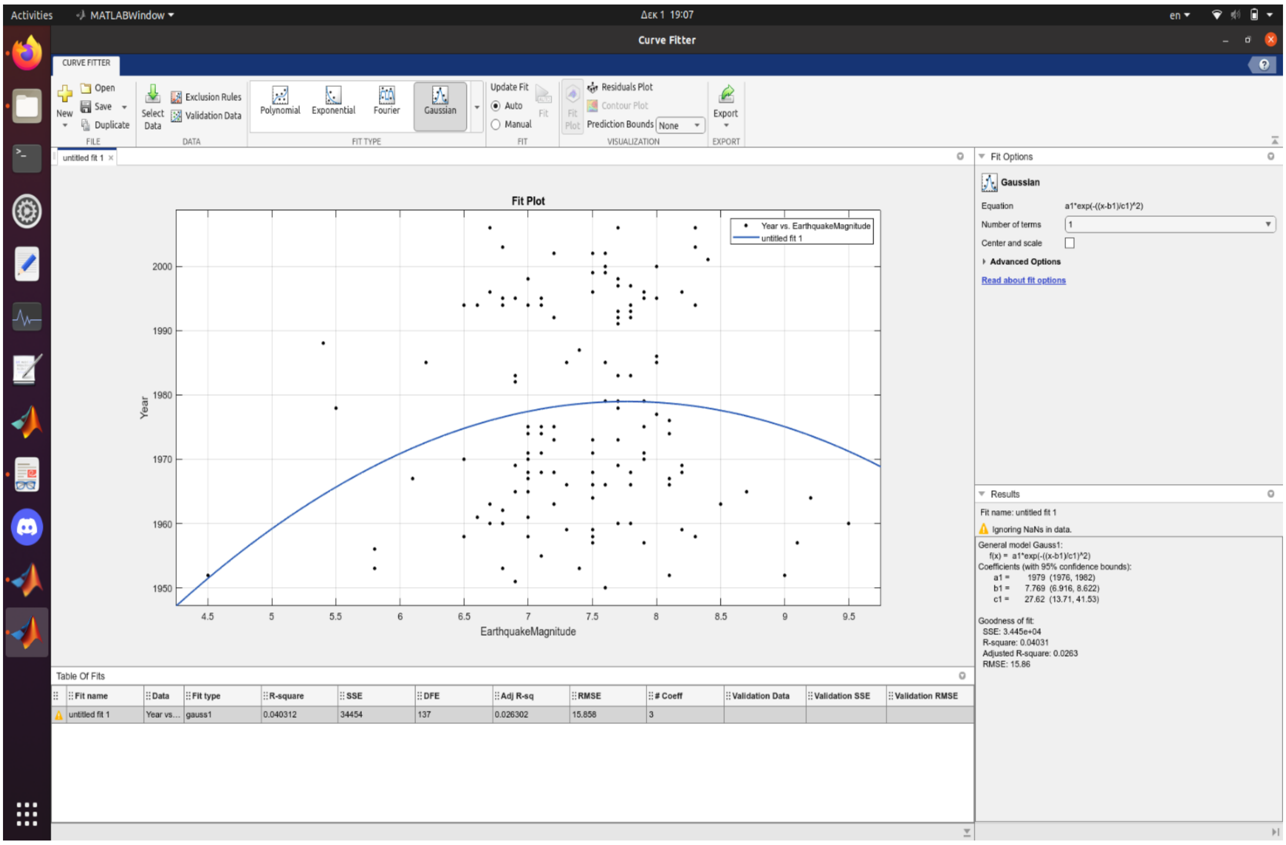


→ Validation Data:

X: EarthquakeMagnitude

Y: Year

Fit type: Gaussian →



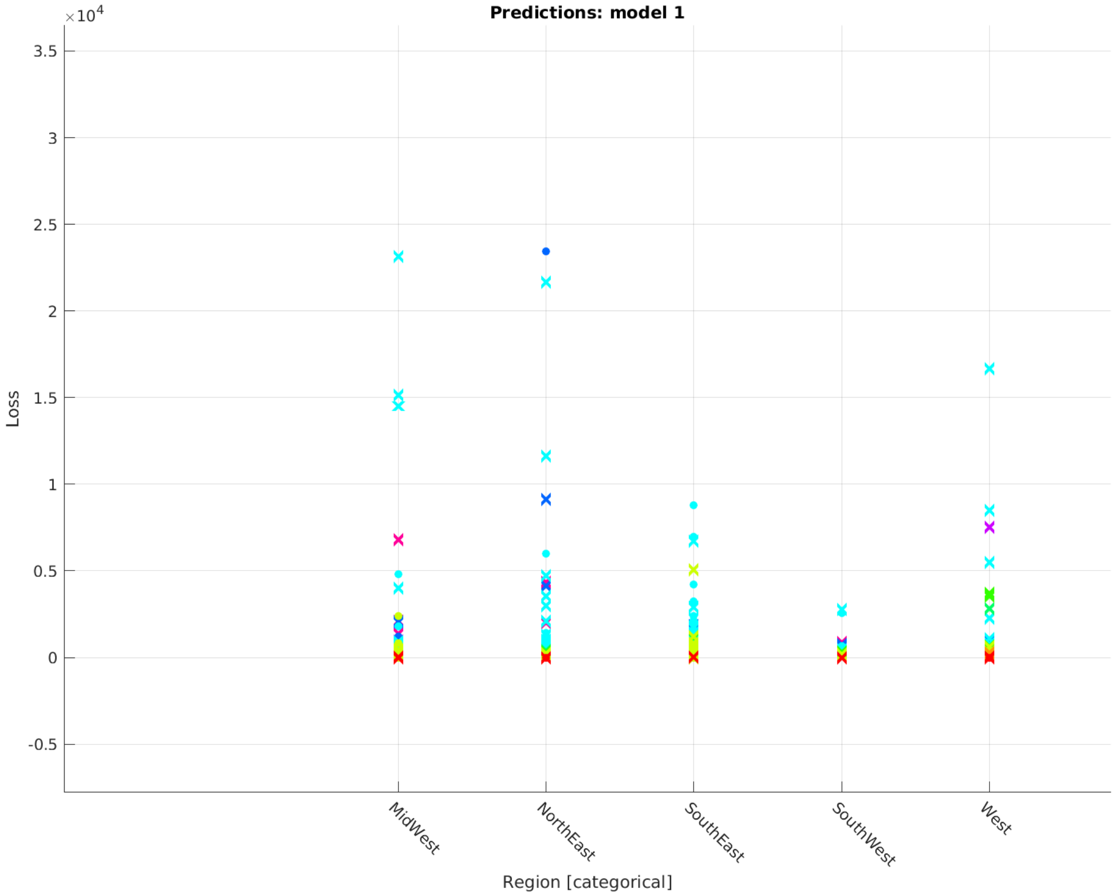
2. Observational Data: outages.csv

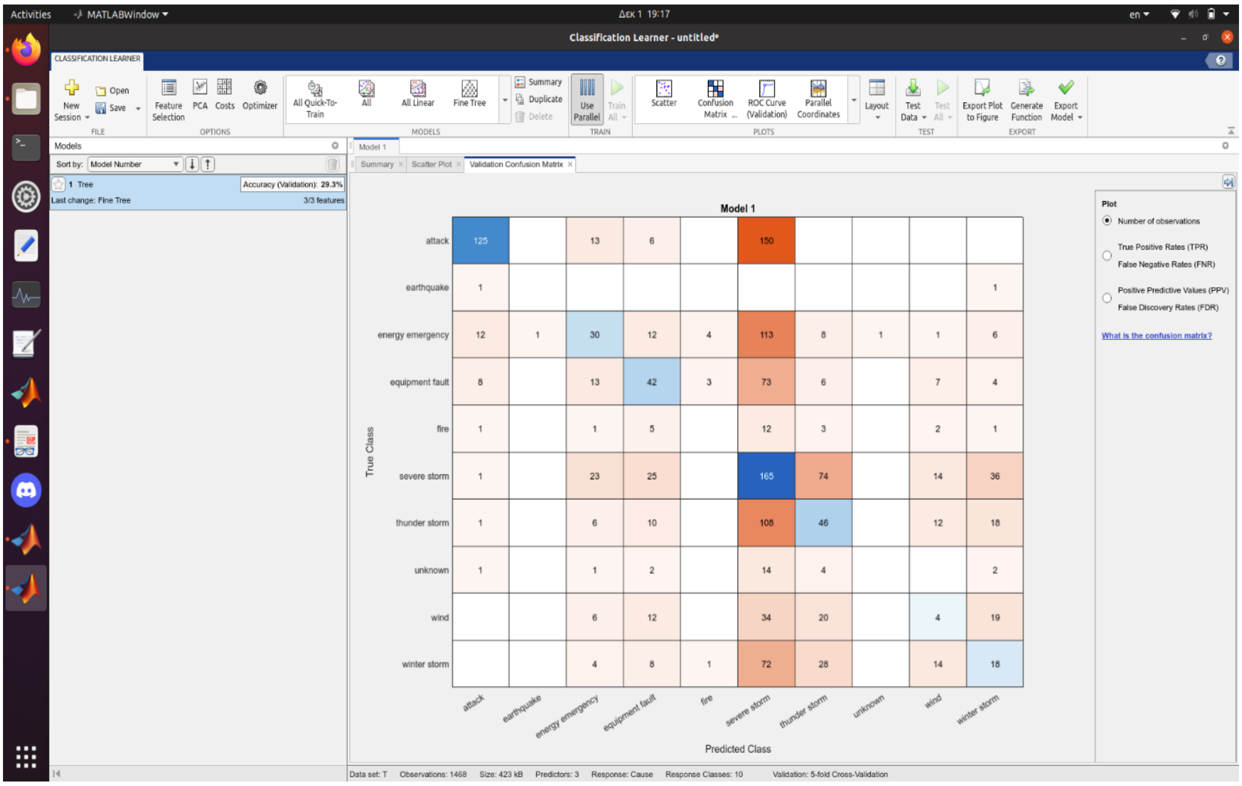
Κώδικας:

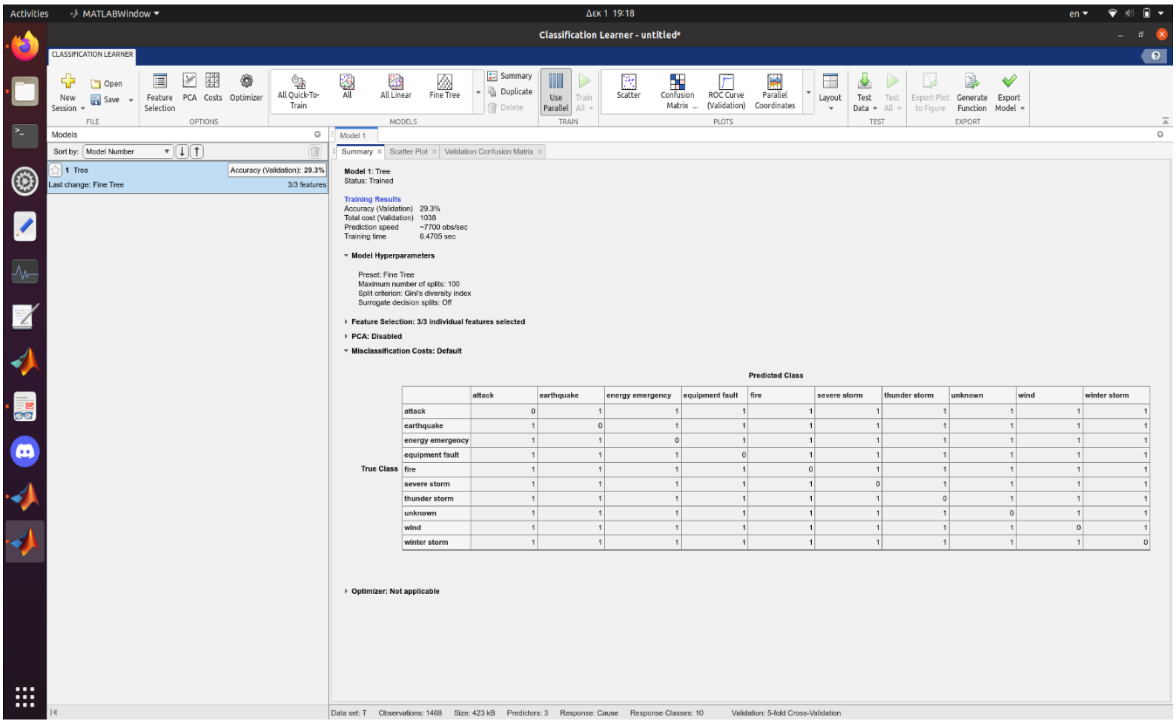
ask1_4.m

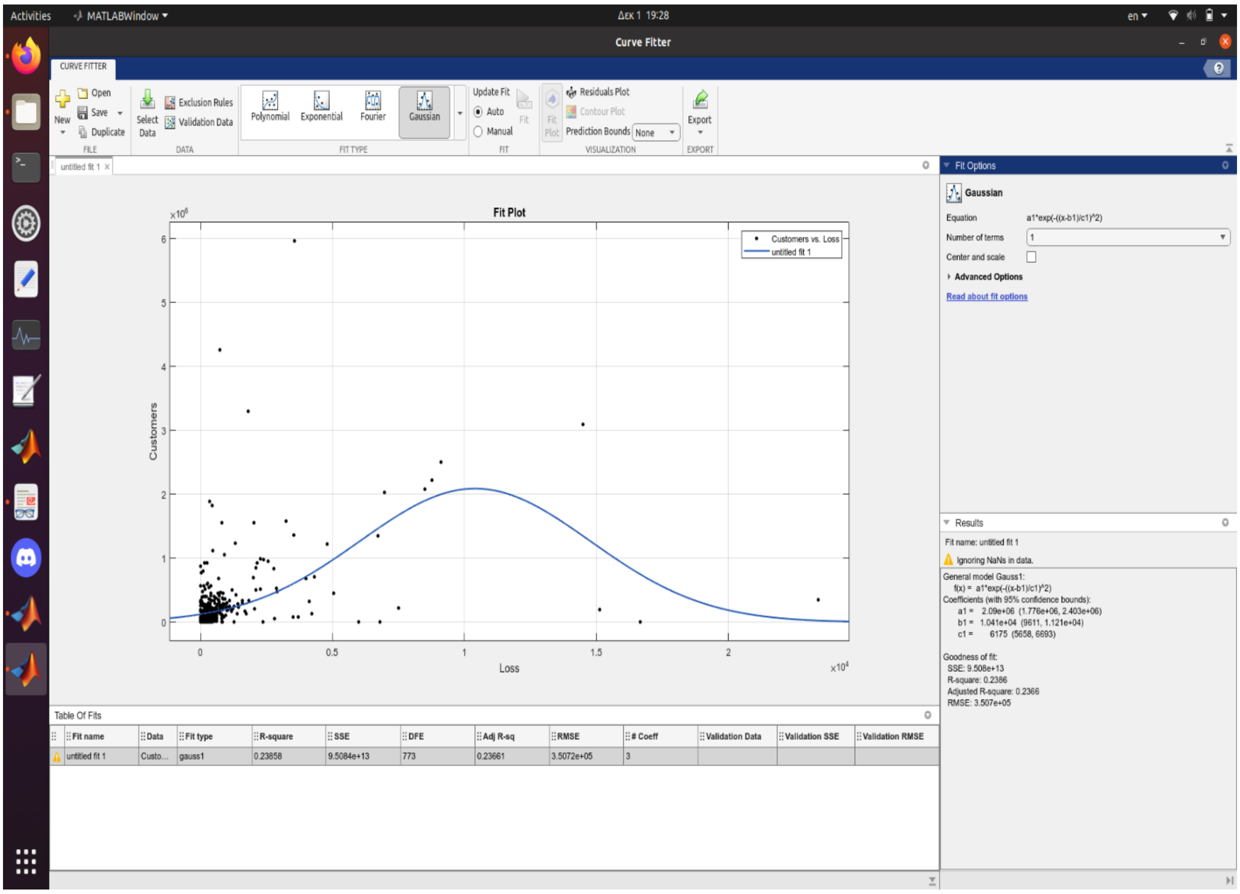
T = readtable('outages.csv');
% Αρχεία που δημιουργούνται(auto-generated files):ClassificationLearnerSession_outages.mat και
% trainClassifier.m

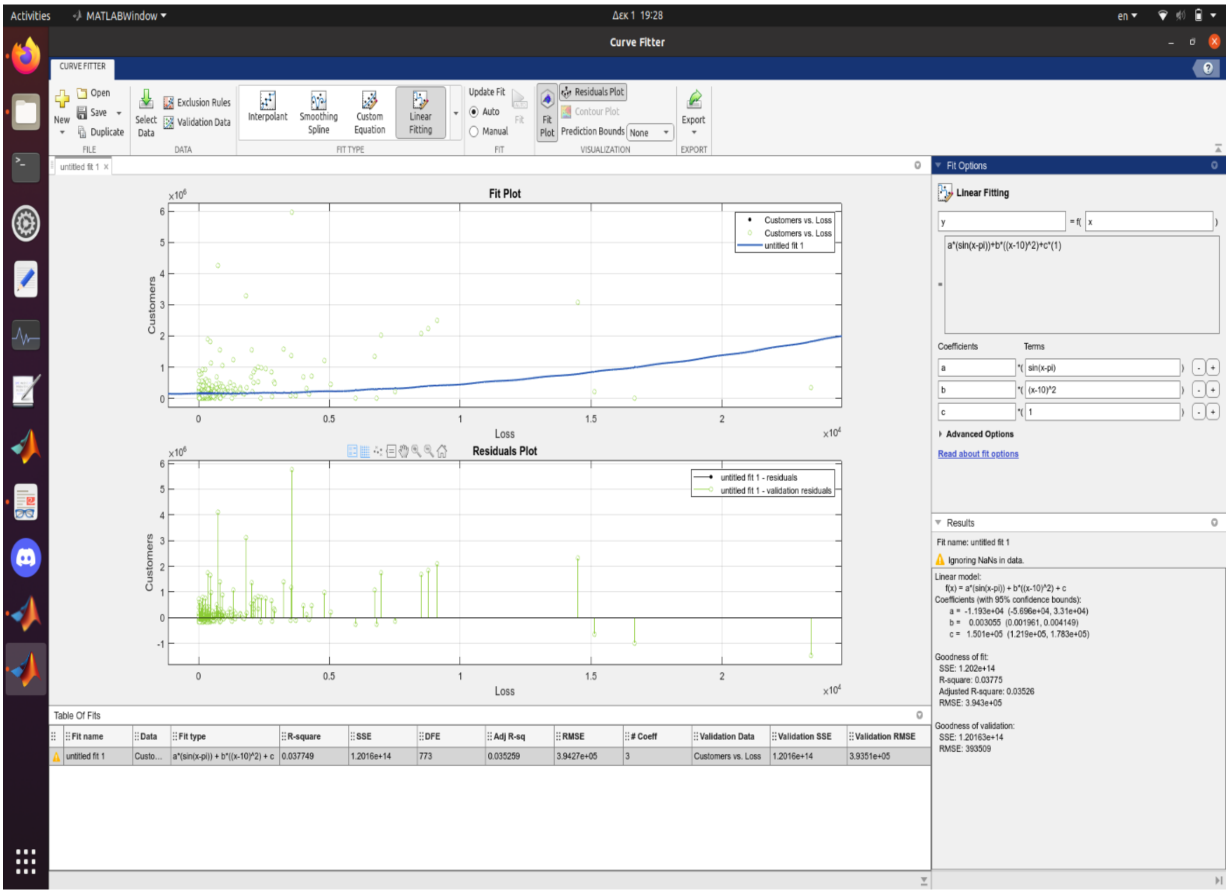
Επανάληψη των παραπάνω βημάτων:


Predictions: model 1
Loss
Region [categorical]
MidWest
NorthEast
SouthEast
SouthWest
West


Classification Learner - untitled*
Model 1
True Class
Predicted Class
attack
earthquake
energy emergency
equipment fault
fire
severe storm
thunder storm
unknown
wind
winter storm


Classification Learner - untitled*
CLASSIFICATION LEARNER
Model 1: Tree
Status: Trained
Training Results
Accuracy (Validation) 29.3%
Total cost (Validation) 1038
Prediction speed ~7700 obs/sec
Training time 6.4705 sec
Model Hyperparameters
Preset: Fine Tree
Maximum number of splits: 100
Split criterion: Gini's diversity index
Surrogate decision splits: Off
Feature Selection: 3/3 individual features selected
PCA: Disabled
Misclassification Costs: Default
Predicted Class
True Class
attack
earthquake
energy emergency
equipment fault
fire
severe storm
thunder storm
unknown
wind
winter storm
Optimizer: Not applicable
Data set: T Observations: 1468 Size: 423 kB Predictors: 3 Response: Cause Response Classes: 10 Validation: 5-fold Cross-Validation


Curve Fitter
CURVE FITTER
Fit Plot
Customers
Loss
Customers vs. Loss
untitled fit 1
Fit Options
Gaussian
Equation a1*exp(-((x-b1)/c1)^2)
Number of terms 1
Center and scale
Advanced Options
Read about fit options
Results
Fit name: untitled fit 1
Ignoring NaNs in data.
General model Gauss1:
f(x) = a1*exp(-((x-b1)/c1)^2)
Coefficients (with 95% confidence bounds):
a1 = 2.09e+06 (1.776e+06, 2.403e+06)
b1 = 1.041e+04 (9611, 1.121e+04)
c1 = 6175 (5658, 6693)
Goodness of fit:
SSE: 9.508e+13
R-square: 0.2386
Adjusted R-square: 0.2366
RMSE: 3.507e+05
Table Of Fits
Fit name Data Fit type R-square SSE DFE Adj R-sq RMSE # Coeff Validation Data Validation SSE Validation RMSE
untitled fit 1 Custo... gauss1 0.23858 9.5084e+13 773 0.23661 3.5072e+05 3



3. Image Data: flujet.mat

Εργαλεία επεξεργασίας δεδομένων: Curve Fitter , Classification Learner

Κώδικας:

ask1_4_3.m

```
load flujet.mat
imshow(X,map)

disp('Εύρεση μέσης τιμής:');
disp(median(map));

disp('Εύρεση ενδιάμεσης τιμής:');
disp(mean(map));

variance = var(map);
```

```
disp('Υπολογισμός διασποράς:');
disp(variance);


ta = std(map);
disp('Υπολογισμός τυπικής απόκλισης:');
disp(ta);
```

Αποτελέσματα:

Εύρεση μέσης τιμής:

    0.5312   0.5312   0.5312

Εύρεση ενδιάμεσης τιμής:

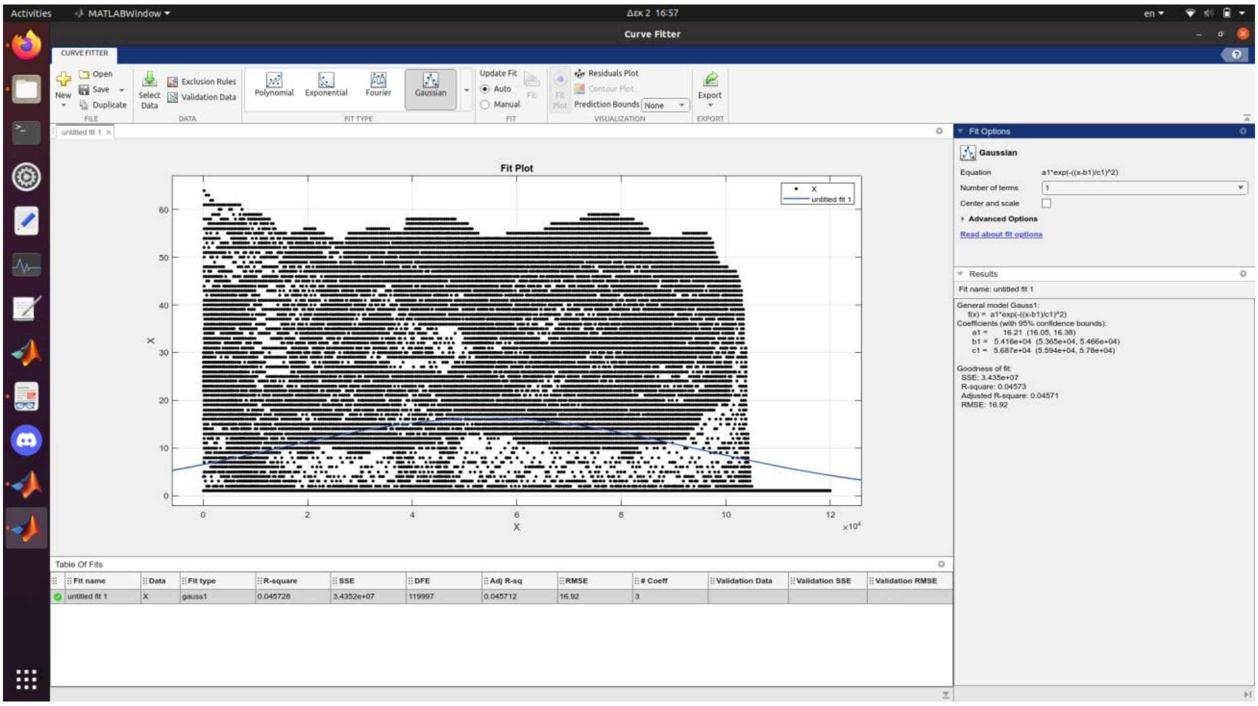
    0.4805   0.5156   0.4805

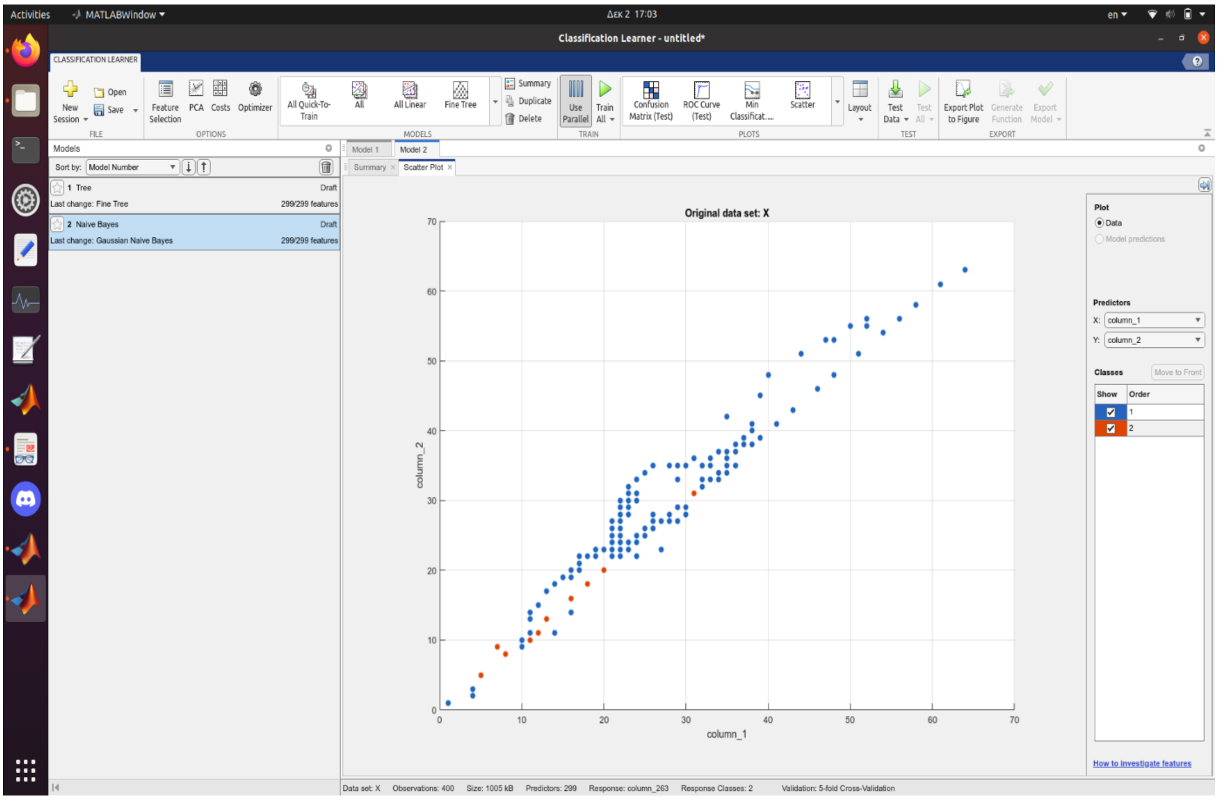
Υπολογισμός διασποράς:

    0.1923   0.1694   0.1923

Υπολογισμός τυπικής απόκλισης:

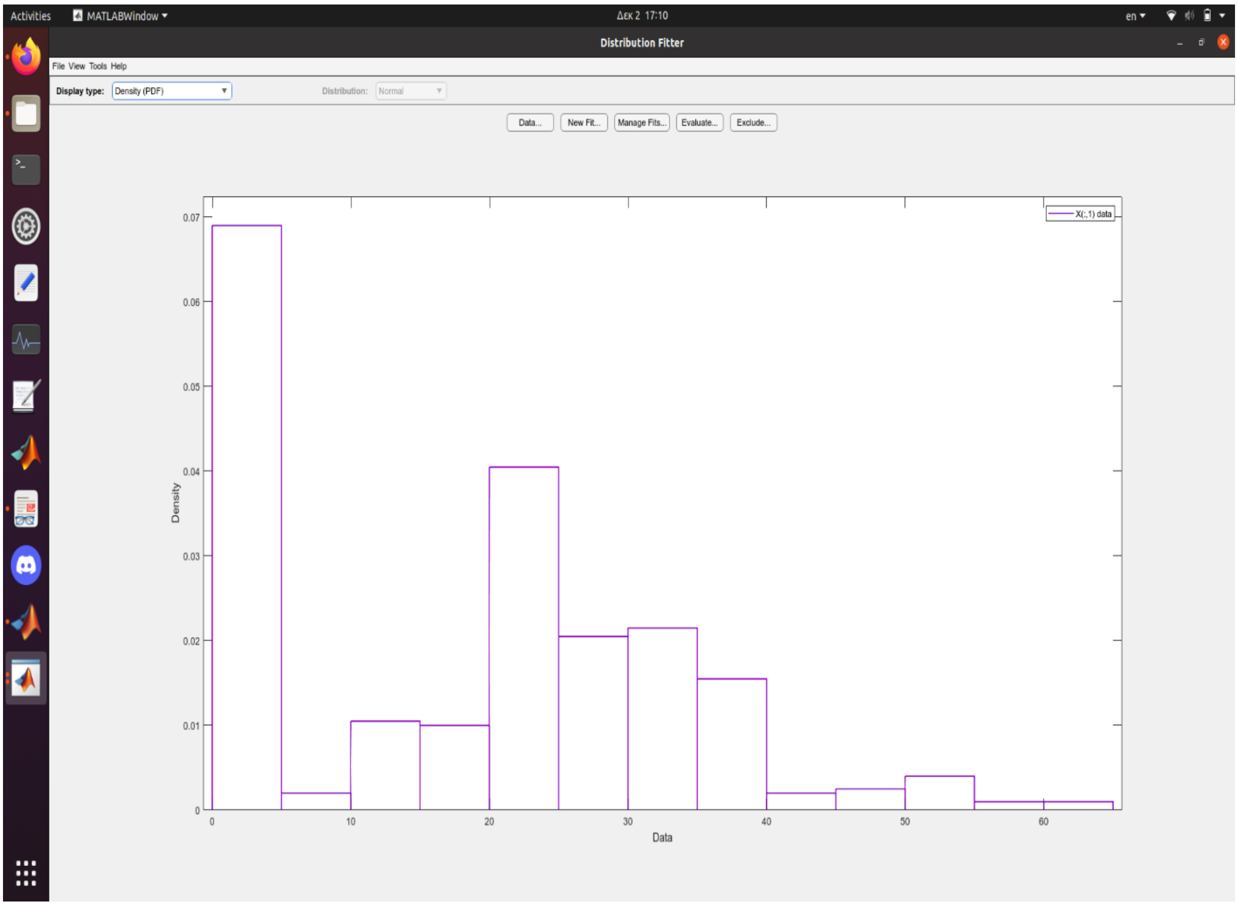
    0.4385   0.4116   0.4385





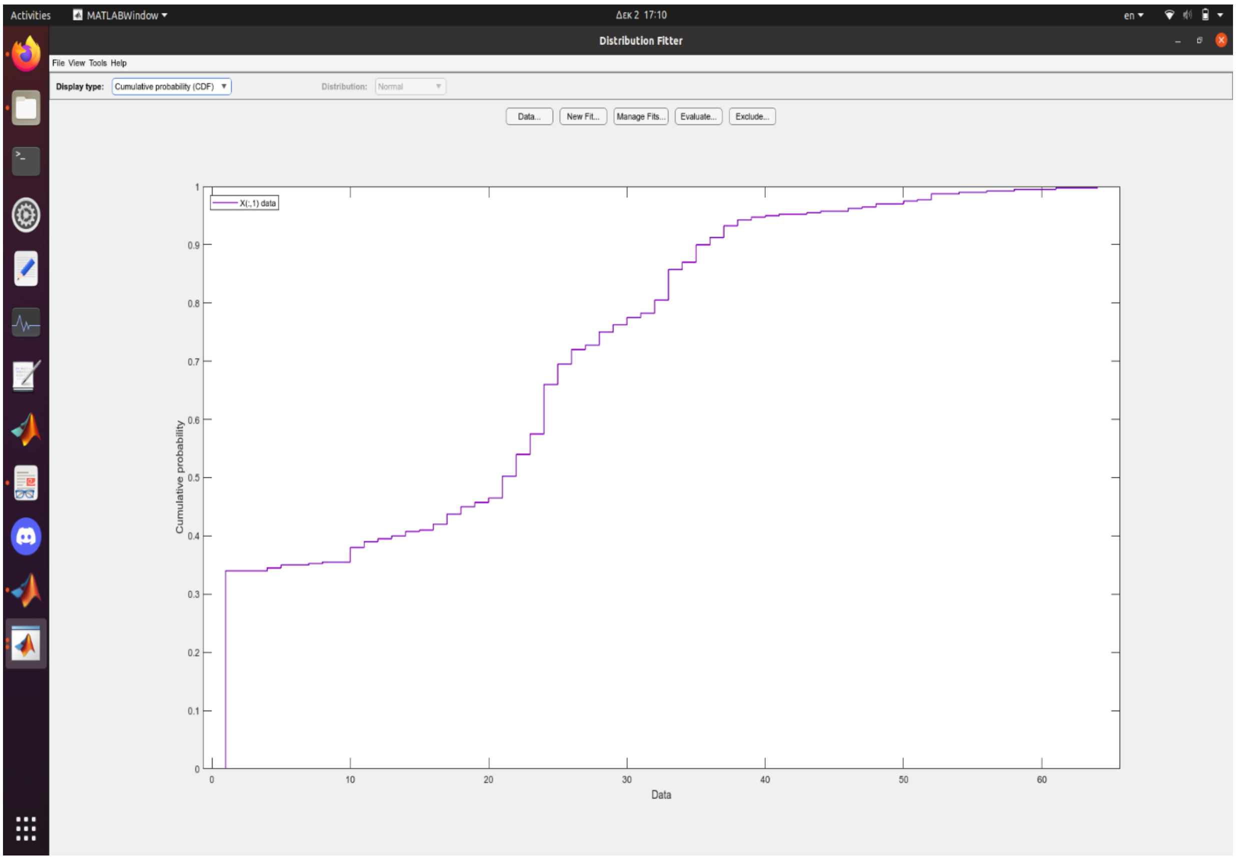
Classification Learner - untitled*
Original data set: X
column_1
column_2
Data set: X
Observations: 400
Predictors: 299
Response: column_263
Response Classes: 2
Validation: 5-fold Cross-Validation

Pdf:

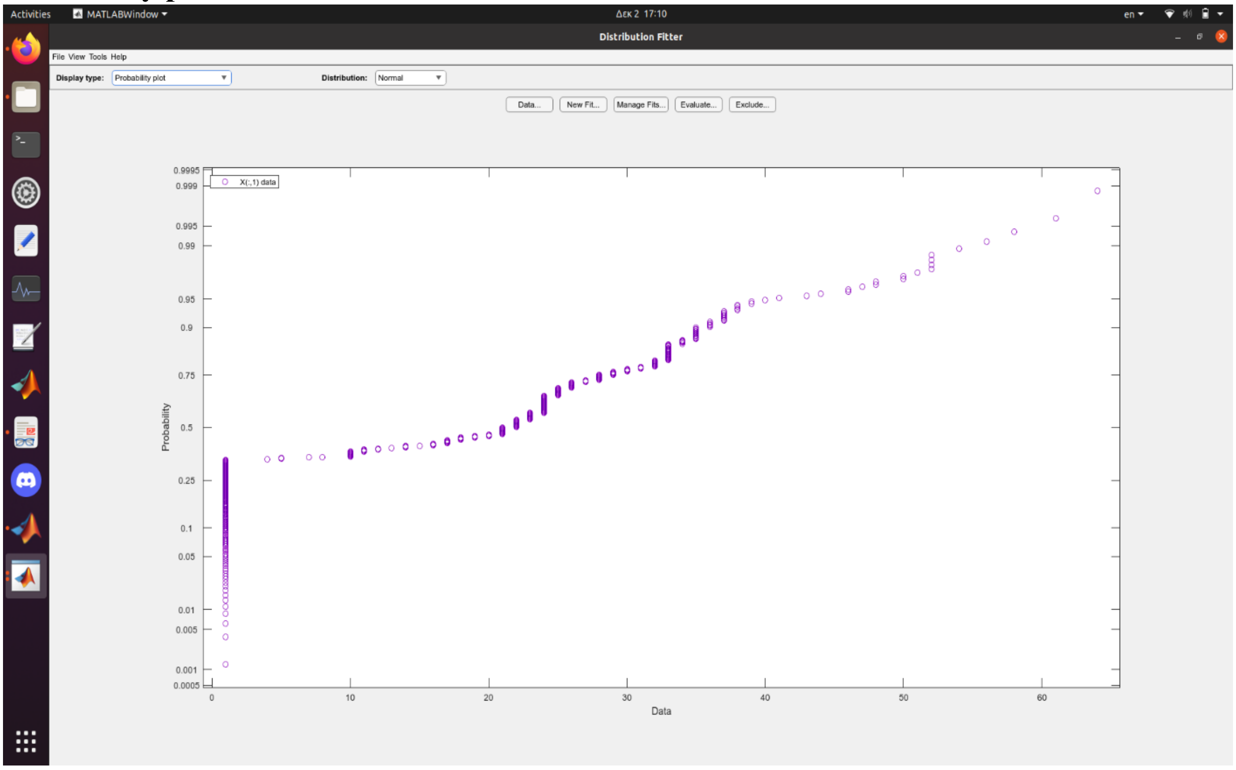


Distribution Fitter
Display type:
Density (PDF)
X(:,1) data
Density
Data

Cdf:



Probability plot:



4. Video and Audio Data: train.mat

Εργαλεία επεξεργασίας δεδομένων: Curve Fitter , Distribution Fitter

Κώδικας:

ask1_4_4.m

```
load train.mat

sound(y,Fs)

% όπου Fs:ο ρυθμός δειγματοληψίας


disp('Εύρεση μέσης τιμής:');

disp(median(y));


disp('Εύρεση ενδιάμεσης τιμής:');

disp(mean(y));


variance = var(y);

disp('Υπολογισμός διασποράς:');

disp(variance);


ta = std(y);

disp('Υπολογισμός τυπικής απόκλισης:');

disp(ta);


%https://www.mathworks.com/matlabcentral/answers/1651580-spectrogram-and-spectrum-of-an-audio-file

spectrogram(y, hamming(256), 250, 256, Fs, 'yaxis');

title('Spectrogram of the Audio Signal');
```

Αποτελέσματα:

Εύρεση μέσης τιμής:

-0.0023

Εύρεση ενδιάμεσης τιμής:

5.7202e-04

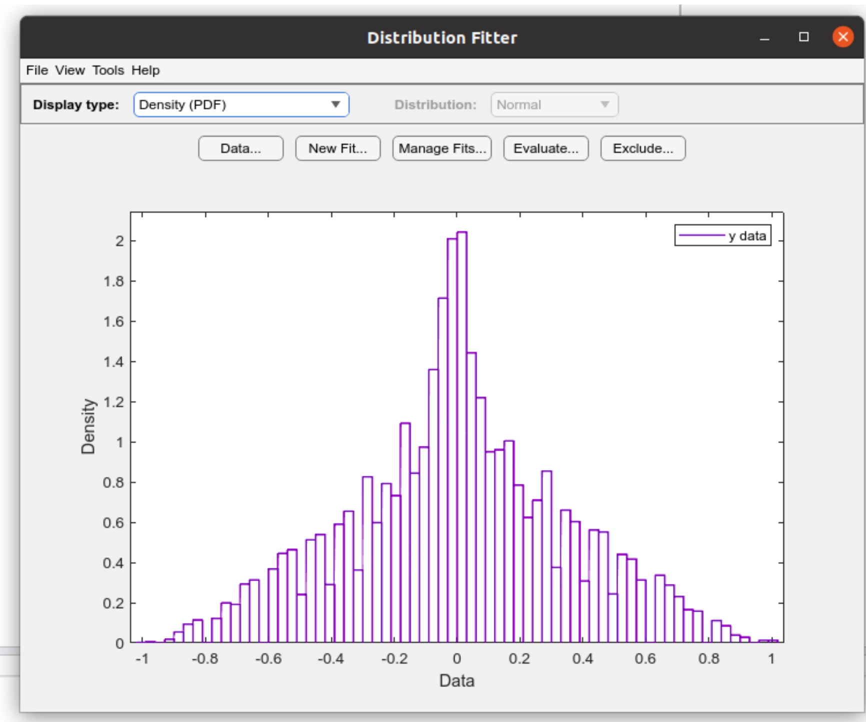
Υπολογισμός διασποράς:

0.1144

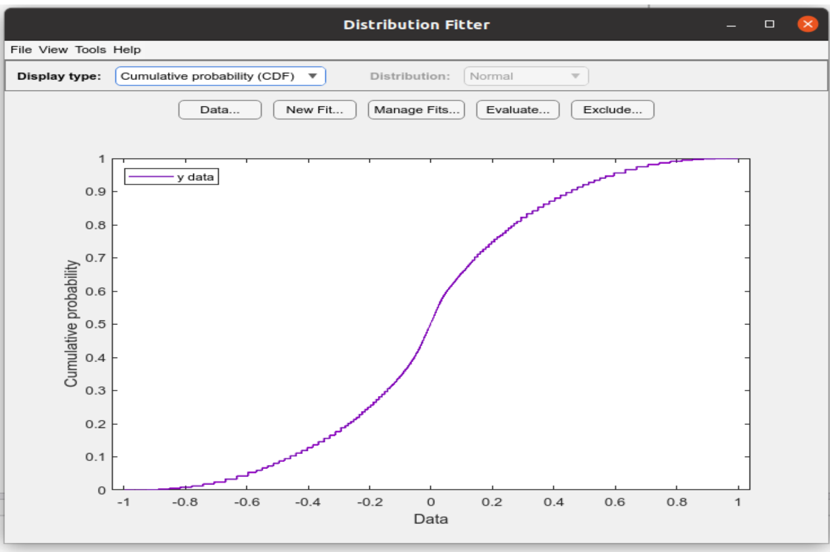
Υπολογισμός τυπικής απόκλισης:

0.3382

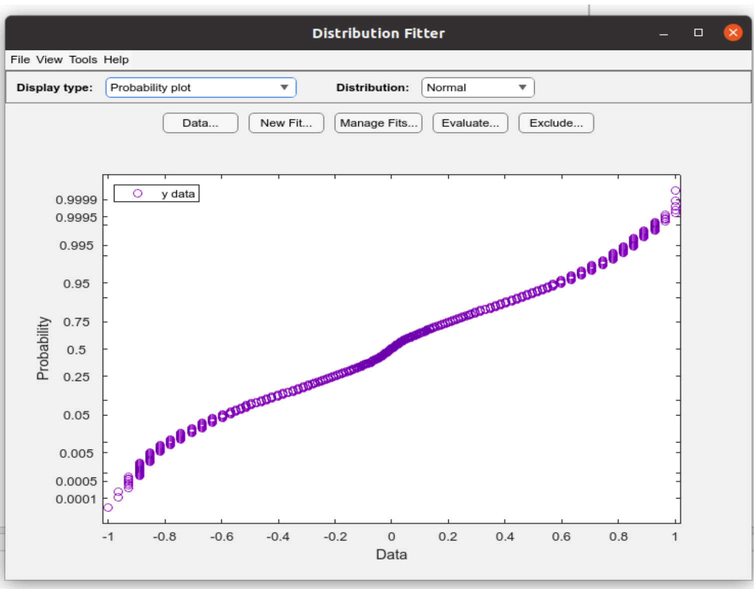
Pdf



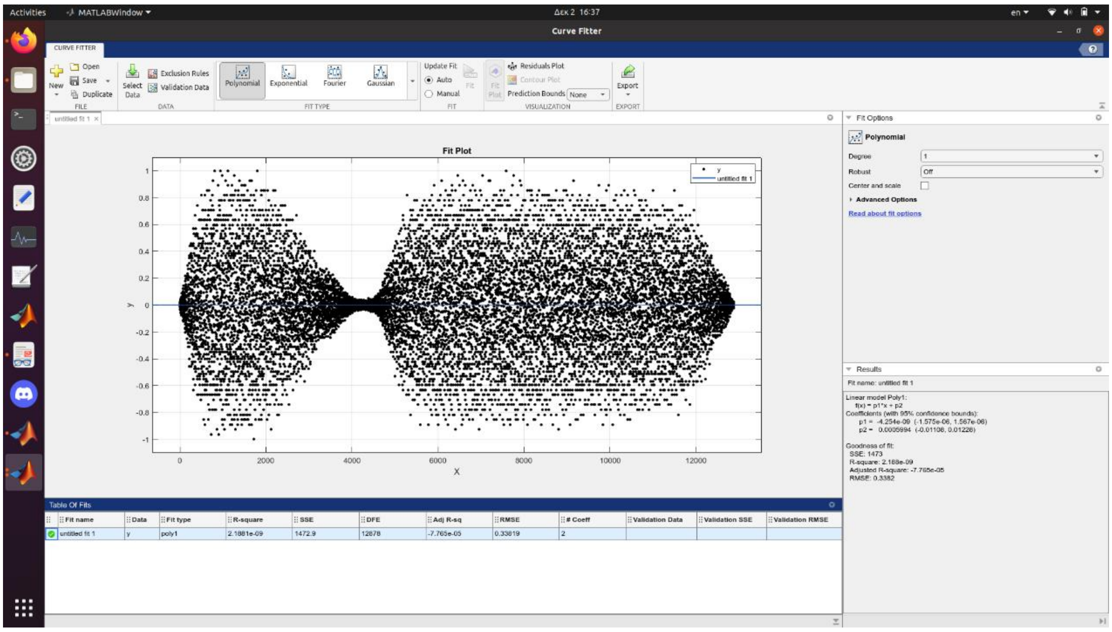
Cdf


Distribution Fitter
File View Tools Help
Display type: Cumulative probability (CDF)
Distribution: Normal
Data... New Fit... Manage Fits... Evaluate... Exclude...
y data
Cumulative probability
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1
Data

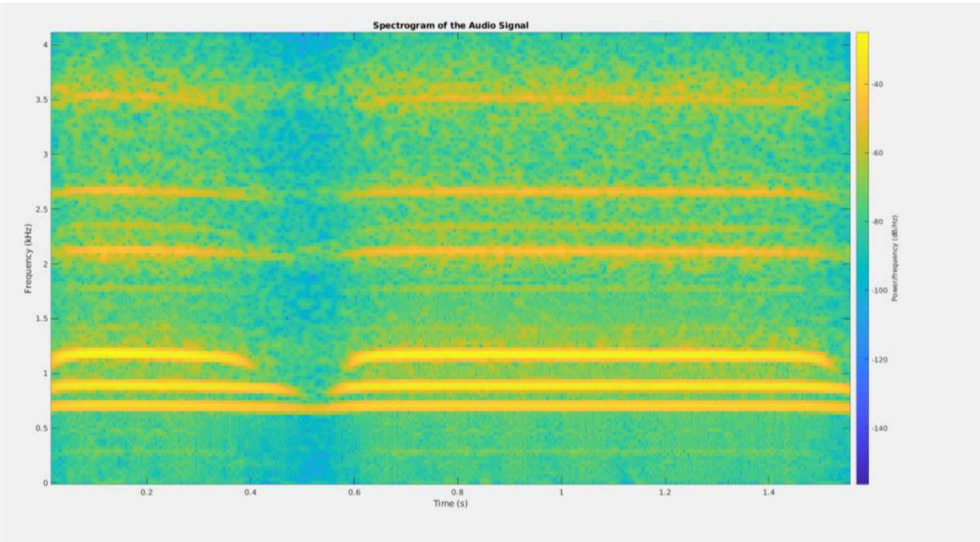
Prob


Distribution Fitter
File View Tools Help
Display type: Probability plot
Distribution: Normal
Data... New Fit... Manage Fits... Evaluate... Exclude...
y data
Probability
0.9999
0.9995
0.995
0.95
0.75
0.5
0.25
0.05
0.005
0.0005
0.0001
-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1
Data

Waveform:



Spectogram:



Μέρος Β. Άσκηση #1 Πιθανότητες & Τυχαίες μεταβλητές.

Β.1. Άσκηση #2 Υπολογισμός μέσης τιμής, διασποράς και σφαλμάτων για διαφορετικές κατανομές.

Αρχεία: ask2_1.m → Κανονική Κατανομή και ask2_2.m → Γεωμετρική Κατανομή

Κώδικας:

ask2_1.m

```
clear all; close all; clc;
mu = 4;
sigma = 2;
iter = [100 1000 10000]; % Αριθμός δειγμάτων

for k = 1:length(iter)
  for i = 1:100
    % Δημιουργία δειγμάτων κανονικής κατανομής
    samples = normrnd(mu, sigma, iter(k), 1);

    % Δημιουργία πυκνότητας πιθανότητας(pdf) της κανονικής κατανομής
    x = mu - 4 * sigma:0.01:mu + 4 * sigma;
    y = normpdf(x, mu, sigma);

    % Δημιουργία και κανονικοποίηση ιστογράμματος
    [n, xout] = hist(samples, 30);
    bw = xout(2) - xout(1);
    n1 = n / sum(n .* bw);

    % Σχεδίαση ιστογράμματος(με μπάρες) και pdf(κόκκινη καμπύλη)
    bar(xout, n1)
    hold on
    plot(x, y, '-r', 'Linewidth', 3);
    hold off
% ------------------------------------------------------------------

    % Υπολογισμός μέσης τιμής και διασποράς
    Ey(i, k) = mean(samples);
```

```
        s(i, k) = var(samples);

        % Υπολογισμός μέσης τιμής και διασποράς βάσει ml
        Eyy(i, k) = (1 / iter(k)) * sum(samples);
        ss(i, k) = (1 / iter(k)) * sum((samples - Eyy(i, k)).^2);
    end

% -------------------------------------------------------------------

    % Υπολογισμός σφάλματος μέσης τιμής και διασποράς
    ermt(:, k) = (1 / 100) * ((Ey(:, k) - Eyy(:, k)).^2);
    errmt(k) = mean(ermt(:, k));

    erd(:, k) = (1 / 100) * ((s(:, k) - ss(:, k)).^2);
    errd(k) = mean(erd(:, k));
end

% -------------------------------------------------------------------
% extra διαγράμματα
% Μέση τιμή
figure;
loglog(Ey, errmt, '-o', 'LineWidth', 2); %
https://www.mathworks.com/help/matlab/ref/loglog.html
title('Μέση τιμή');
xlabel('Αριθμός Δειγμάτων');
ylabel('Μέση τιμή');
grid on;

% ------------------------------------------------------------------------------------------------
```

ask2_1_2.m

```
% Γεωμετρική Κατανομή
clear all;
close all;
clc;

mu = 4;
p = 1/6;
iter = [100 1000 10000];

k = 1;

for k = 1:length(iter)
   for i = 1:100
      % Δημιουργία δειγμάτων
      samples = geornd(p, iter(k), 1);

      % Δημιουργία pdf της γεωμετρικής κατανομής
      x = 1:max(samples);
      y = geopdf(x, p);

      % Δημιουργία και κανονικοποίηση ιστογράμματος
      [n, xout] = hist(samples, 30);
      bw = xout(2) - xout(1);
      n1 = n / sum(n .* bw);

      % Σχεδίαση ιστογράμματος(με μπάρες) και pdf(κόκκινη καμπύλη)
      bar(xout, n1)
      hold on
      plot(x, y, '-r', 'Linewidth', 3);
      hold off

      % Μέση τιμή και διασπορά
```

```
        Ey(i, k) = mean(samples);
        s(i, k) = var(samples);
    end;

    % Σφάλμα μέσης τιμής
    ermt(:, k) = (1/100) * ((Ey(:, k) - 1/p).^2);
    errmt(k) = mean(ermt(:, k));

    % Σφάλμα διασποράς
    erd(:, k) = (1/100) * ((s(:, k) - (1-p)/p^2).^2);
    errd(k) = mean(erd(:, k));

    k = k + 1;
end

% -------------------------------------------------------------------
% extra διαγράμματα
% Μέση τιμή
figure;
loglog(Ey, errmt, '-o', 'LineWidth', 2); %
https://www.mathworks.com/help/matlab/ref/loglog.html
title('Μέση τιμή');
xlabel('Αριθμός Δειγμάτων');
ylabel('Μέση τιμή');
grid on;


% Διασπορά
figure;
loglog(iter, s, '-o', 'LineWidth', 2);
title('Διασπορά');
xlabel('Αριθμός Δειγμάτων');
ylabel('Διασπορά');
```

```
grid on;


% Σφάλμα της μέσης τιμής
figure;
loglog(iter, errmt, '-o', 'LineWidth', 2);
title('Σφάλμα Μέσης Τιμής');
xlabel('Αριθμός Δειγμάτων');
ylabel('Σφάλμα');
grid on;


% Σφάλμα της διασποράς
figure;
loglog(iter, errd, '-o', 'LineWidth', 2);
title('Σφάλμα Διασποράς');
xlabel('Αριθμός Δειγμάτων');
ylabel('Σφάλμα');
grid on;
```

Ask2_1

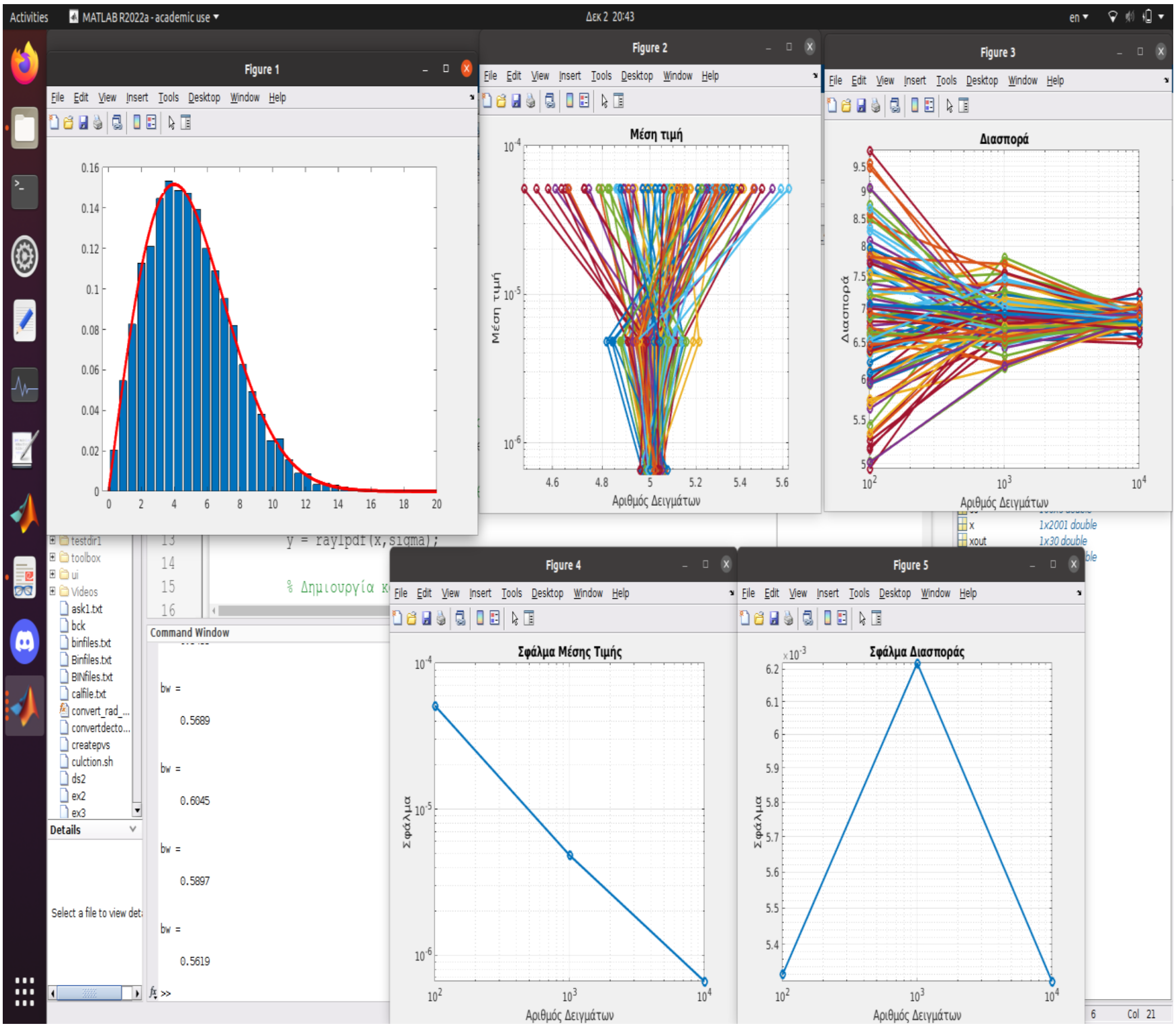
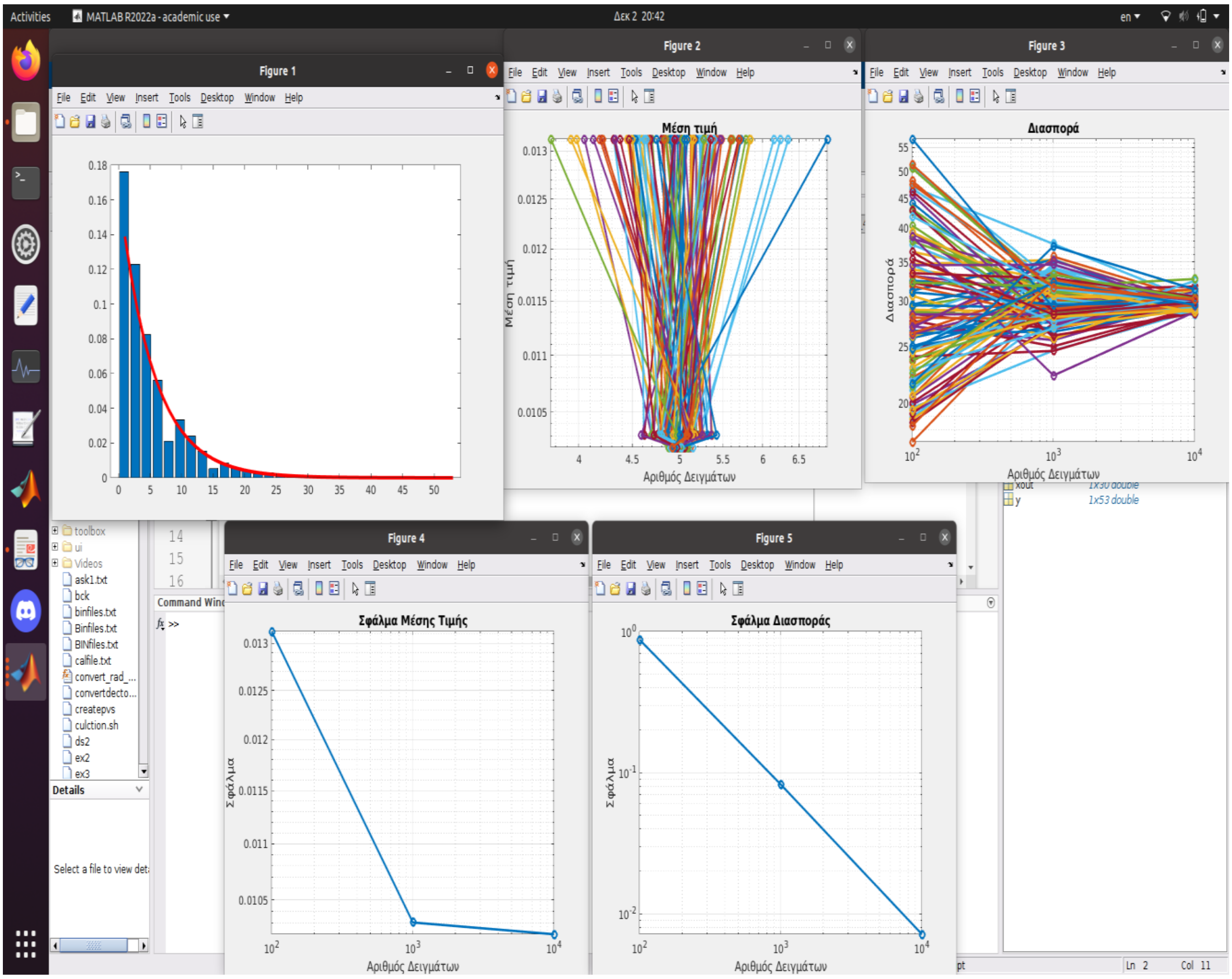


Figure 1
Figure 2
Μέση τιμή
Μέση τιμή
Αριθμός Δειγμάτων
Figure 3
Διασπορά
Διασπορά
Αριθμός Δειγμάτων
Figure 4
Σφάλμα Μέσης Τιμής
Σφάλμα
Αριθμός Δειγμάτων
Figure 5
Σφάλμα Διασποράς
Σφάλμα
Αριθμός Δειγμάτων
Command Window
bw =
0.5689
bw =
0.6045
bw =
0.5897
bw =
0.5619

Ask2_2



Σημείωση: Υπολογίζουμε τη μέση τιμή, τη διασπορά και τα σφάλματα για διαφορετικές κατανομές. Ο δοσμένος κώδικας επεξεργάστηκε, ώστε να γίνει απλούστερος στην ανάγνωση και την κατανόηση. Επίσης, προστέθηκαν κάποια διαγράμματα, με στόχο την πλήρη κατανόηση του προβλήματος.

Για τα διαγράμματα χρησιμοποιείται η συνάρτηση loglog, για να εμφανίζονται τα δεδομένα σε λογαριθμική κλίμακα.

Συμπεράσματα:

Η μέση τιμή και η διασπορά υπολογίζονται για κάθε σειρά δειγμάτων. Η κανονική κατανομή είναι συνεχής, ενώ η γεωμετρική κατανομή είναι διακριτή.

- Όσον αφορά την κανονική κατανομή, το σφάλμα μέσης τιμής μειώνεται γραμμικά μέχρι τα 10^3 δείγματα. Μετά από αυτό το όριο, αυξάνεται γραμμικά κατά την αύξηση του αριθμού των δειγμάτων. Ενώ, το σφάλμα διασποράς, μειώνεται όσο αυξάνεται ο αριθμός των δειγμάτων. Η μέση τιμή λαμβάνει τιμές από 6 έως και 13*10^-34. Η διασπορά λαμβάνει τιμές από 0 έως και 6.

- Όσον αφορά τη γεωμετρική κατανομή, το σφάλμα μέσης τιμής μειώνεται γραμμικά μέχρι τα 10^3 δείγματα. Έπειτα, έχει μία πολύ μικρή αύξηση, και με <<γυμνό μάτι>> φαίνεται ότι το σφάλμα βαίνει στο 0.Το σφάλμα διασποράς, μειώνεται όσο αυξάνεται ο αριθμός των δειγμάτων. Η μέση τιμή λαμβάνει τιμές από 0 έως και 0.014. Η διασπορά λαμβάνει τιμές από 0 έως και 74.

B.2. Άσκηση #2 Υπολογισμός μέσης τιμής, διασποράς και σφαλμάτων για διαφορετικές κατανομές.

Αρχεία: ask2_2.m → Κανονική Rayleigh και ask2_2_2.m → Κατανομή Poisson

Κώδικας:

ask2_2.m

```
clear all; close all; clc;

sigma=4;

iter = [100 1000 10000]; % αριθμός δειγμάτων

k=1; % βοηθητική μεταβλητή


for k=1:length(iter)

   for i=1:100

      % Δημιουργία δειγμάτων κατανομής Rayleigh

      samples = raylrnd(sigma,iter(k),1);


      % Δημιουργία πυκνότητας πιθανότητας(pdf) της κατανομής Rayleigh

      x =0:0.01:20;

      y = raylpdf(x,sigma);


      % Δημιουργία και κανονικοποίηση ιστογράμματος

      [n,xout] = hist(samples,30);

      bw = xout(2)-xout(1)

      n1 = n/sum(n.*bw);


      % Σχεδίαση ιστογράμματος(με μπάρες) και pdf(κόκκινη καμπύλη)

      bar(xout,n1)

      hold on
```

```
        plot(x,y,'-r','Linewidth',3);
        hold off

       % --------------------------------------------------------------------

        % Υπολογισμός μέσης τιμής και διασποράς
        Ey(i,k)=mean(samples);
        s(i,k)=var(samples);

        % Υπολογισμός μέσης τιμής και διασποράς βάσει ml
        sigmaa(i,k)=sqrt((1/(2*iter(k)))*sum(samples.^2));
        Eyy(i,k)=sigmaa(i,k)*sqrt((pi/2));
        ss(i,k)=((4-pi)/2)*(sigmaa(i).^2);
    end;

       % --------------------------------------------------------------------

        % Υπολογισμός σφάλματος μέσης τιμής και διασποράς
        ermt(:,k)=(1/100)*((Ey(:,k)-Eyy(:,k)).^2);
        errmt(k)=mean(ermt(:,k));

        erd(:,k)=(1/100)*((s(:,k)-ss(:,k)).^2);
        errd(k)=mean(erd(:,k));
        k=k+1;
end;

% --------------------------------------------------------------------
% extra διαγράμματα

% Μέση τιμή
figure;
loglog(Ey, errmt, '-o', 'LineWidth', 2); %
https://www.mathworks.com/help/matlab/ref/loglog.html
```

```
title('Μέση τιμή');
xlabel('Αριθμός Δειγμάτων');
ylabel('Μέση τιμή');
grid on;


% Διασπορά
figure;
loglog(iter, s, '-o', 'LineWidth', 2);
title('Διασπορά');
xlabel('Αριθμός Δειγμάτων');
ylabel('Διασπορά');
grid on;


% Σφάλμα της μέσης τιμής
figure;
loglog(iter, errmt, '-o', 'LineWidth', 2);
title('Σφάλμα Μέσης Τιμής');
xlabel('Αριθμός Δειγμάτων');
ylabel('Σφάλμα');
grid on;


% Σφάλμα της διασποράς
figure;
loglog(iter, errd, '-o', 'LineWidth', 2);
title('Σφάλμα Διασποράς');
xlabel('Αριθμός Δειγμάτων');
ylabel('Σφάλμα');
grid on;
```

% --

ask2_2_2.m

```
clear all;
close all;
clc;

lambda = 4; % Παράμετρος της κατανομής Poisson
iter = [100 1000 10000]; % Αριθμός δειγμάτων

for k = 1:length(iter)
  for i = 1:100
      % Δημιουργία δειγμάτων κατανομής Poisson
      samples = poissrnd(lambda, iter(k), 1);

      % Δημιουργία πυκνότητας πιθανότητας(pdf) της κατανομής Poisson
      x = 0:max(samples);
      y = poisspdf(x, lambda);

      % Δημιουργία και κανονικοποίηση ιστογράμματος
      [n, xout] = hist(samples, 30);
      bw = xout(2) - xout(1);
      n1 = n / sum(n .* bw);

      % Σχεδίαση ιστογράμματος(με μπάρες) και pdf(κόκκινη καμπύλη)
      bar(xout, n1)
      hold on
      plot(x, y, '-r', 'Linewidth', 3);
      hold off

 % --------------------------------------------------------------------
```

```
        % Υπολογισμός μέσης τιμής και διασποράς
        Ey(i, k) = mean(samples);
        s(i, k) = var(samples);

        % Υπολογισμός μέσης τιμής και διασποράς βάσει ml
        lambdaa(i, k) = (1 / iter(k)) * sum(samples);
        Eyy(i, k) = lambdaa(i, k);
        ss(i, k) = (1 / iter(k)) * sum((samples - Eyy(i, k)).^2);
    end
  % ----------------------------------------------------------------------

    % Υπολογισμός σφάλματος μέσης τιμής και διασποράς
    ermt(:, k) = (1 / 100) * ((Ey(:, k) - Eyy(:, k)).^2);
    errmt(k) = mean(ermt(:, k));

    erd(:, k) = (1 / 100) * ((s(:, k) - ss(:, k)).^2);
    errd(k) = mean(erd(:, k));
end

% ----------------------------------------------------------------------
% extra διαγράμματα

% Μέση τιμή
figure;
loglog(Ey, errmt, '-o', 'LineWidth', 2);
title('Μέση τιμή');
xlabel('Αριθμός Δειγμάτων');
ylabel('Μέση τιμή');
grid on;

% Διασπορά
figure;
```

```
loglog(iter, s, '-o', 'LineWidth', 2);
title('Διασπορά');
xlabel('Αριθμός Δειγμάτων');
ylabel('Διασπορά');
grid on;

% Σφάλμα της μέσης τιμής
figure;
loglog(iter, errmt, '-o', 'LineWidth', 2);
title('Σφάλμα Μέσης Τιμής');
xlabel('Αριθμός Δειγμάτων');
ylabel('Σφάλμα');
grid on;

% Σφάλμα της διασποράς
figure;
loglog(iter, errd, '-o', 'LineWidth', 2);
title('Σφάλμα Διασποράς');
xlabel('Αριθμός Δειγμάτων');
ylabel('Σφάλμα');
grid on;
```

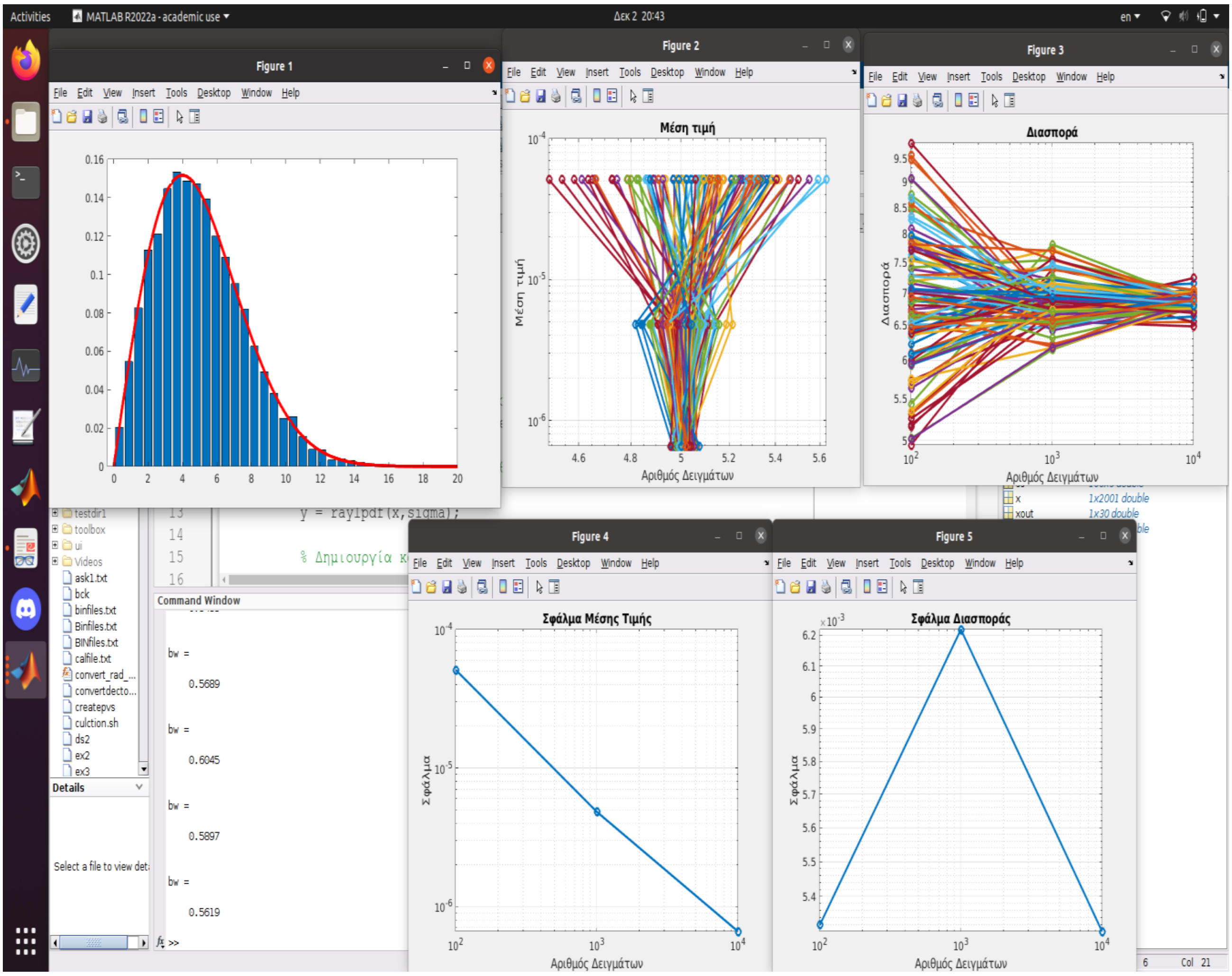
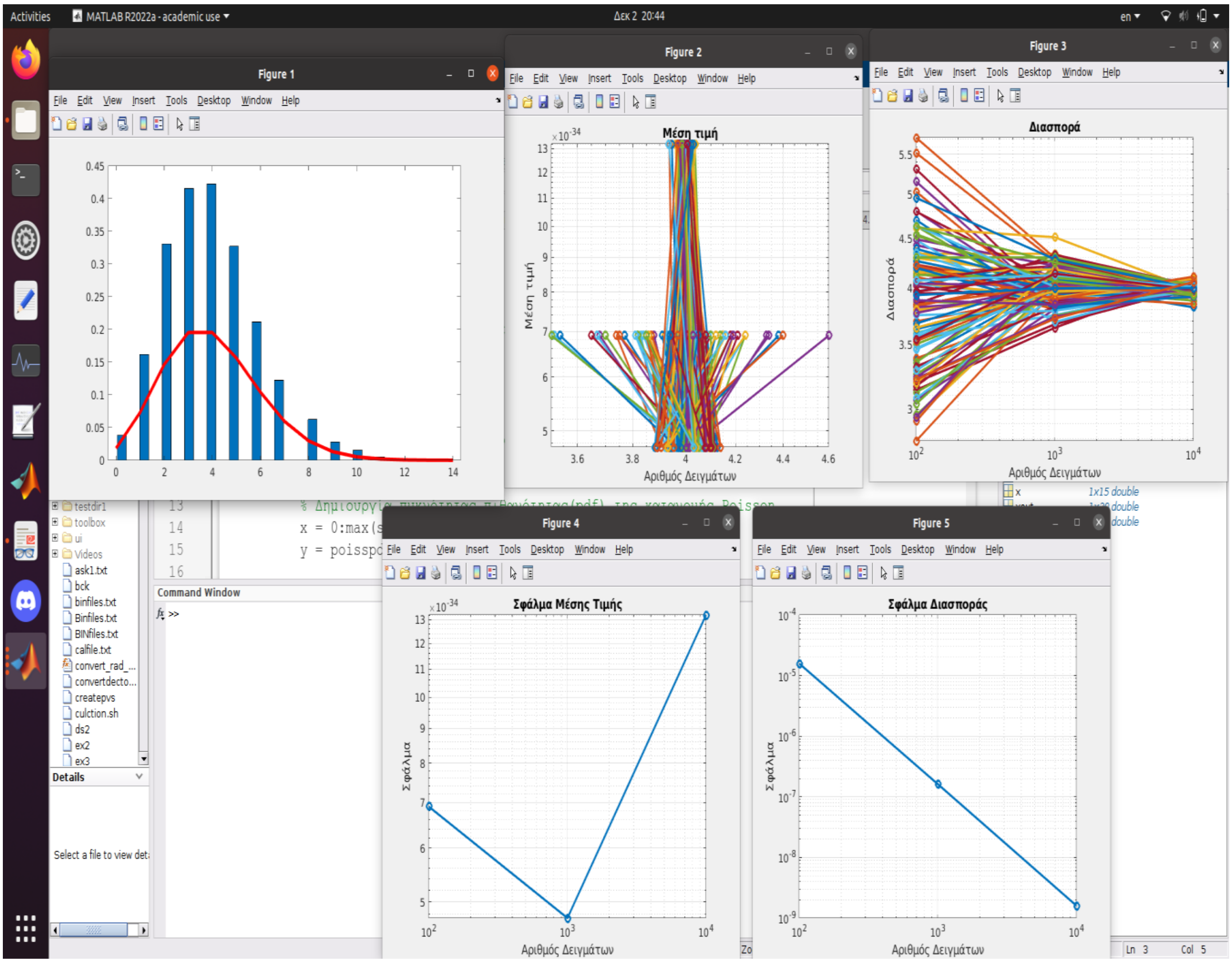


Figure 1
Figure 2
Μέση τιμή
Αριθμός Δειγμάτων
Figure 3
Διασπορά
Αριθμός Δειγμάτων
Figure 4
Σφάλμα Μέσης Τιμής
Σφάλμα
Αριθμός Δειγμάτων
Figure 5
Σφάλμα Διασποράς
Σφάλμα
Αριθμός Δειγμάτων
Command Window
bw =
0.5689
bw =
0.6045
bw =
0.5897
bw =
0.5619



Συμπεράσματα:

Η κατανομή Rayleigh είναι συνεχής πιθανοτική κατανομή, ενώ ή Poisson διακριτή πιθανοτική.

- Όσον αφορά την κατανομή Rayleigh, μοιάζει οπτικά στο πρώτο μισό της με την κανονική κατανομή. Ωστόσο, στο δεύτερο μισό της η καμπύλη διαγράφει φθίνουσα πορεία, μέχρι να μηδενιστεί. Η μέση τιμή λαμβάνει τιμές από 0 έως και 7*10^-5. Η διασπορά λαμβάνει τιμές από 5 έως και 10. Το σφάλμα μέσης τιμής μειώνεται γραμμικά, όσο αυξάνεται ο αριθμός των δειγμάτων. Τέλος, το σφάλμα διασποράς, μειώνεται με την αύξηση του αριθμού των δειγμάτων. Οφείλει να τονιστεί το γεγονός ότι μετά από 10^3 δείγματα η πτώση της τιμής του σφάλματος διασποράς είναι μεγαλύτερη, δηλαδή πιο απότομη.

- Όσον αφορά την κατανομή Poisson, μοιάζει οπτικά με την κατανομή Rayleigh. Και οι δύο ανήκουν στην κατηγορία των πιθανοτικών κατανομών.

  Μέση τιμή: 0 – 7.5*10^-34
  Διασπορά: 0 – 5.6

Σφάλμα μέσης τιμής: Αυξάνεται γραμμικά, με την αύξηση του αριθμού των δειγμάτων, σε αντίθεση με την κατανομή Rayleigh.
Σφάλμα διασποράς: Μειώνεται γραμμικά, με την αύξηση του αριθμού των δειγμάτων.

Β.3. Άσκηση #2 Πολυδιάστατες κατανομές.

Κώδικας:

ask2_3.m

```
randn('seed',0) % initialization
m = [0 0]';
N = 500;

S1 = [1 0; 0 1];
X1 = mvnrnd(m, S1, N)'; % μητρώο που περιέχει τα διανύσματα στις στήλες
του

figure(1), plot(X1(1,:), X1(2,:),'.');
figure(1), axis equal
figure(1), axis ([-7 7 -7 7])

% ----------------------------------------------------------------------
S2 = [0.2 0; 0 0.2];
X2 = mvnrnd(m, S2, N)';

figure(2), plot(X2(1,:), X2(2,:),'.');
figure(2), axis equal
figure(2), axis ([-7 7 -7 7])

% ----------------------------------------------------------------------
S3 = [2 0; 0 2];
X3 = mvnrnd(m, S3, N)';

figure(3), plot(X3(1,:), X3(2,:),'.');
figure(3), axis equal
figure(3), axis ([-7 7 -7 7])

% ----------------------------------------------------------------------
S4 = [0.2 0; 0 2];
X4 = mvnrnd(m, S4, N)';

figure(4), plot(X4(1,:), X4(2,:),'.');
figure(4), axis equal
figure(4), axis ([-7 7 -7 7])
```

```
% -----------------------------------------------------------------------
S5 = [2 0; 0 0.2];
X5 = mvnrnd(m, S5, N)';

figure(5), plot(X5(1,:), X5(2,:),'.');
figure(5), axis equal
figure(5), axis ([-7 7 -7 7])

% -----------------------------------------------------------------------
S6 = [1 0.5; 0.5 1];
X6 = mvnrnd(m, S6, N)';

figure(6), plot(X6(1,:), X6(2,:),'.');
figure(6), axis equal
figure(6), axis ([-7 7 -7 7])

% -----------------------------------------------------------------------
S7 = [0.3 0.5; 0.5 2];
X7 = mvnrnd(m, S7, N)';

figure(7), plot(X7(1,:), X7(2,:),'.');
figure(7), axis equal
figure(7), axis ([-7 7 -7 7])

% -----------------------------------------------------------------------
S8 = [0.3 -0.5; -0.5 2];
X8 = mvnrnd(m, S8, N)';

figure(8), plot(X8(1,:), X8(2,:),'.');
figure(8), axis equal
figure(8), axis ([-7 7 -7 7])
```

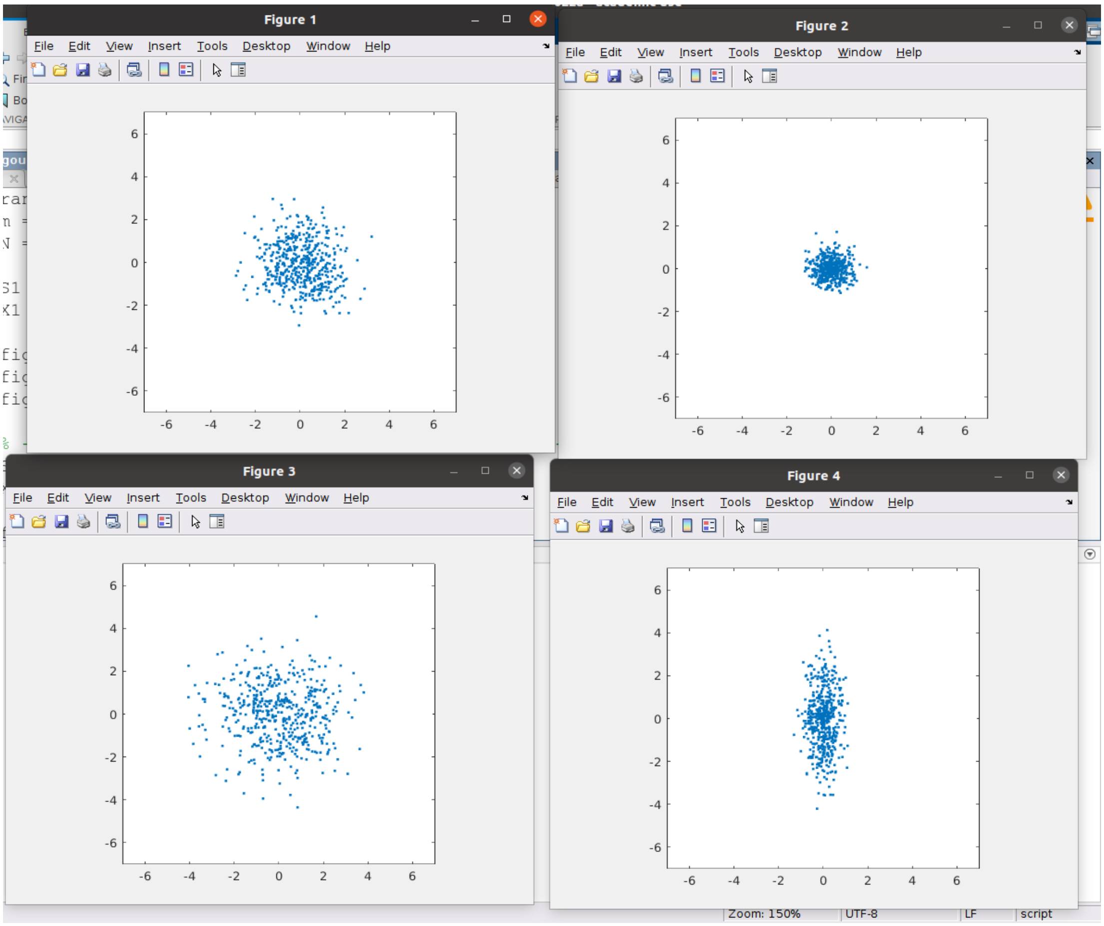


Figure 1
Figure 2
Figure 3
Figure 4
Zoom: 150%
UTF-8
LF
script

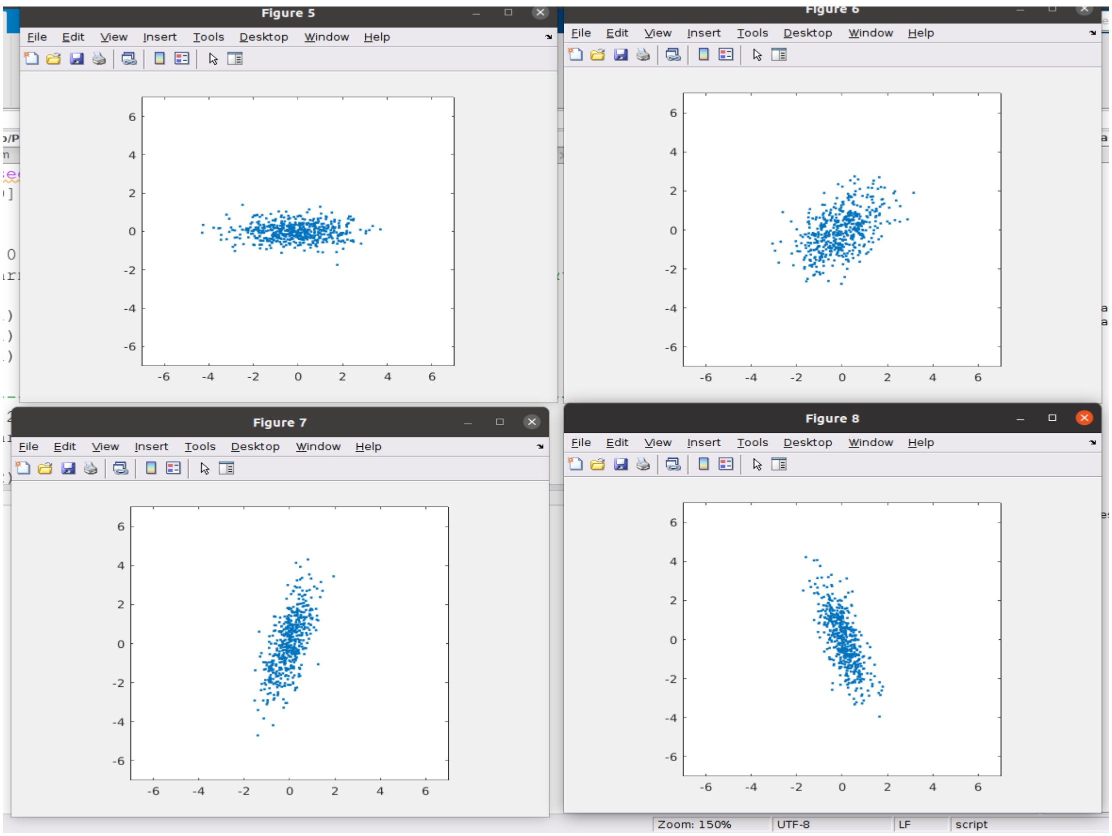


Figure 5
Figure 6
Figure 7
Figure 8
Zoom: 150%
UTF-8
LF
script

Συμπεράσματα:
Κανονική κατανομή N(m,S). Δίνονται η μέση τιμή και ο αριθμός των δισδιάστατων σημείων. Δίνονται επίσης, οι τιμές του πίνακα συνδιασποράς S. Συγκεκριμένα δίνονται 8 περιπτώσεις για τον πίνακα. Δηλαδή, παράγονται 8 σύνολα δεδομένων, έπειτα απεικονίζονται γραφικά και τέλος ακολουθεί σχολιασμός των αποτελεσμάτων. Για κάθε μία από τις 8 περιπτώσεις με βάση τον εκάστοτε πίνακα S, υπολογίζουμε το X=mvnrnd(m,S,N)'; .

Από τη θεωρία γνωρίζουμε ότι ο πίνακας συνδιασποράς S δίνει τη διασπορά των χαρακτηριστικών στην κύρια διαγώνιο. Επιπροσθέτως, παρέχει τη συνδιακύμανση των χαρακτηριστικών xi, xj εκτός διαγωνίου. Γενικά, αν σij=0 τα xi, xj είναι ανεξάρτητα. Αν σij>0, τότε έχουν θετική συσχέτιση. Τέλος, αν σij<0, τότε έχουν αρνητική συσχέτιση.

Ισχύει ότι σij=ρ*σi*σj, με το ρ να ονομάζεται συντελεστής συσχέτισης και να εκτείνεται από -1 έως και 1. Αν ρ περίπου ίση με 0, τότε το scatter plot έχει μορφή κύκλου. Αν 0<ρ<1 , τότε έχουμε θετική συσχέτιση. Δηλαδή, το scatter plot των δύο χαρακτηριστικών έχει μορφή έλλειψης με πρωτεύοντα άξονα που σχηματίζει θετική γωνία με τον οριζόντιο άξονα. Ενώ, αν -1<ρ<0, τότε έχουμε αρνητική συσχέτιση.

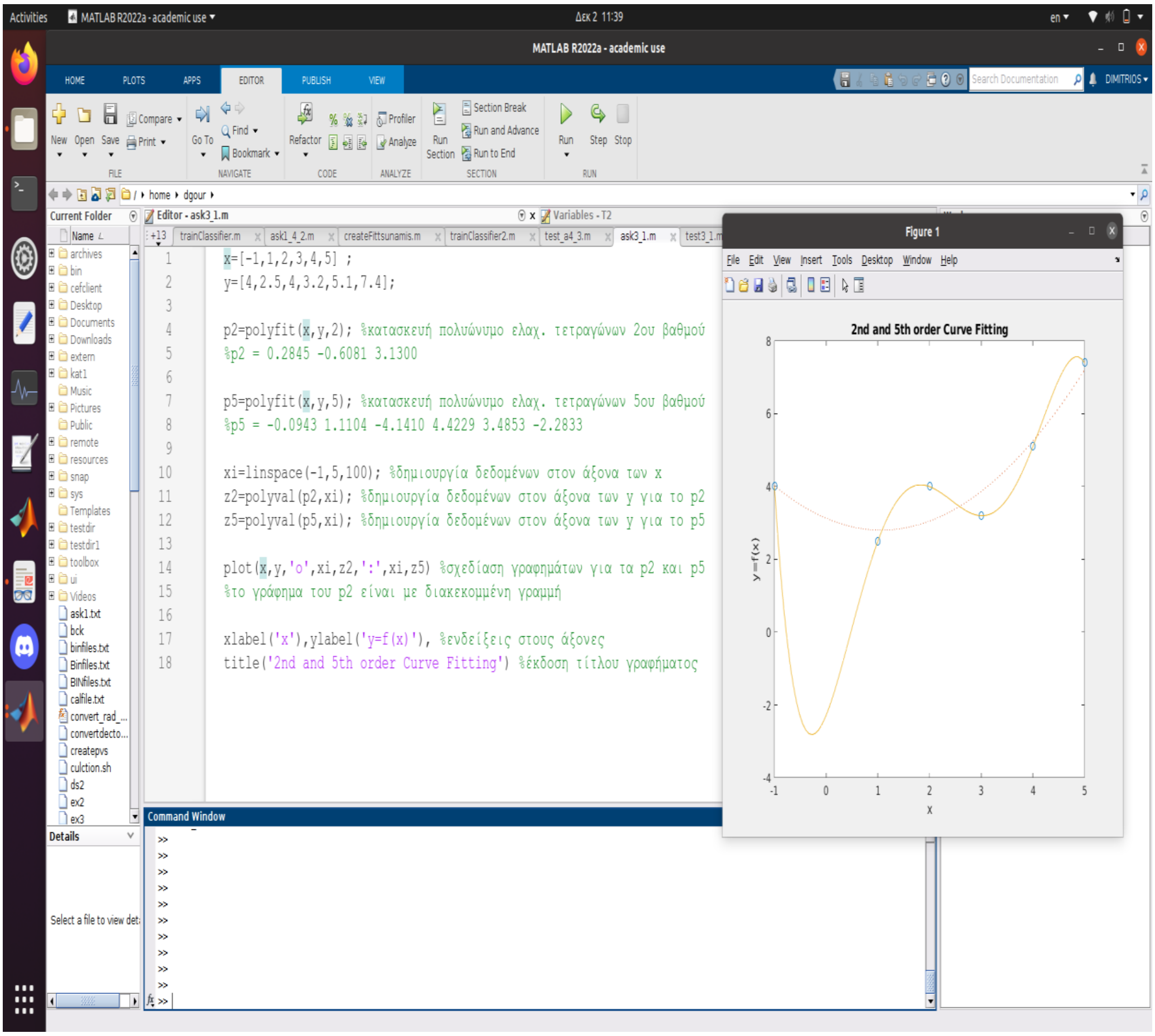
Σε αυτό το σημείο θα γίνει ανάλυση των scatter plots που προκύπτουν, για κάθε μία από τις 8 περιπτώσεις του πίνακα συνδιασποράς. Αρχικά, για την 1η περίπτωση, παρατηρούμε ότι το σύνολο δεδομένων παρουσιάζει κυκλική μορφή, με τα δεδομένα να κατανέμονται γύρω από το σημείο (0,0).
Η μορφή των δεδομένων στη 2η περίπτωση μοιάζει αρχικά με την 1η . Ωστόσο, τα δεδομένα εμφανίζουν μικρότερη διασπορά. Το συγκεκριμένο συμπέρασμα προκύπτει, καθώς τα δείγματα βρίσκονται πιο συμπυκνωμένα γύρω από το σημείο (0,0).
Σε κυκλική μορφή παρουσιάζονται και τα δεδομένα της 3ης περίπτωσης. Όμως, σε αντίθεση με τη 2η , τα δεδομένα εμφανίζουν τη μέχρι τώρα μεγαλύτερη διασπορά. Είναι τοποθετημένα αρκετά αραιά.
Η 4η περίπτωση διαφέρει από τις προηγούμενες 3. Δηλαδή, η διασπορά διαφέρει στις 2 κατευθύνσεις. Το σύνολο των δεδομένων απεικονίζεται ως έλλειψη, που βρίσκεται πάνω στον άξονα y.
Και στο 5ο παράδειγμα παρουσιάζεται ελλειπτική η μορφή των δεδομένων. Σε αντίθεση με το προηγούμενο παράδειγμα, η έλλειψη βρίσκεται στον άξονα x. Το συγκεκριμένο γεγονός αιτιολογείται, καθώς σ1^2>σ2^2.
Στο 6ο παράδειγμα επιστρέφουμε στη συνθήκη σ1^2=σ2^2. Όμως, είναι η πρώτη περίπτωση όπου σ12=0.5 και όχι 0. Η κατανομή δεν είναι συμμετρική, και γέρνει προς τα δεξιά.
Για τις 2 τελευταίες περιπτώσεις ισχύει ότι σ1^2=0.3 και σ2^2=2. Ωστόσο, στην 7η σ12=0.5 και στην 8η σ12=-0.5. Οι 2 κατανομές εμφανίζουν ελλειπτική μορφή και μικρή διασπορά. Η διαφορά τους είναι η εξής: Στην 7η περίπτωση η διασπορά είναι μεγαλύτερη στον άξονα x σε σύγκριση με τον άξονα y. Στην 8η περίπτωση ισχύει το αντίθετο. Οπτικά, εάν τις

τοποθετήσουμε σε έναν κύκλο, το 7° σχήμα έχει δεδομένα στο 1° και το 3° τεταρτημόριο, ενώ το 8° σχήμα, στο 2° και το 4° τεταρτημόριο.

Γ. Άσκηση #1 Παραμετρική προσαρμογή δεδομένων και στατιστικά μοντέλα. Γ.1.

Γ.1. Άσκηση #2 Προσαρμογή καμπυλών. Μέθοδος ελαχίστων τετραγώνων.

Παράδειγμα εργαστηρίου:



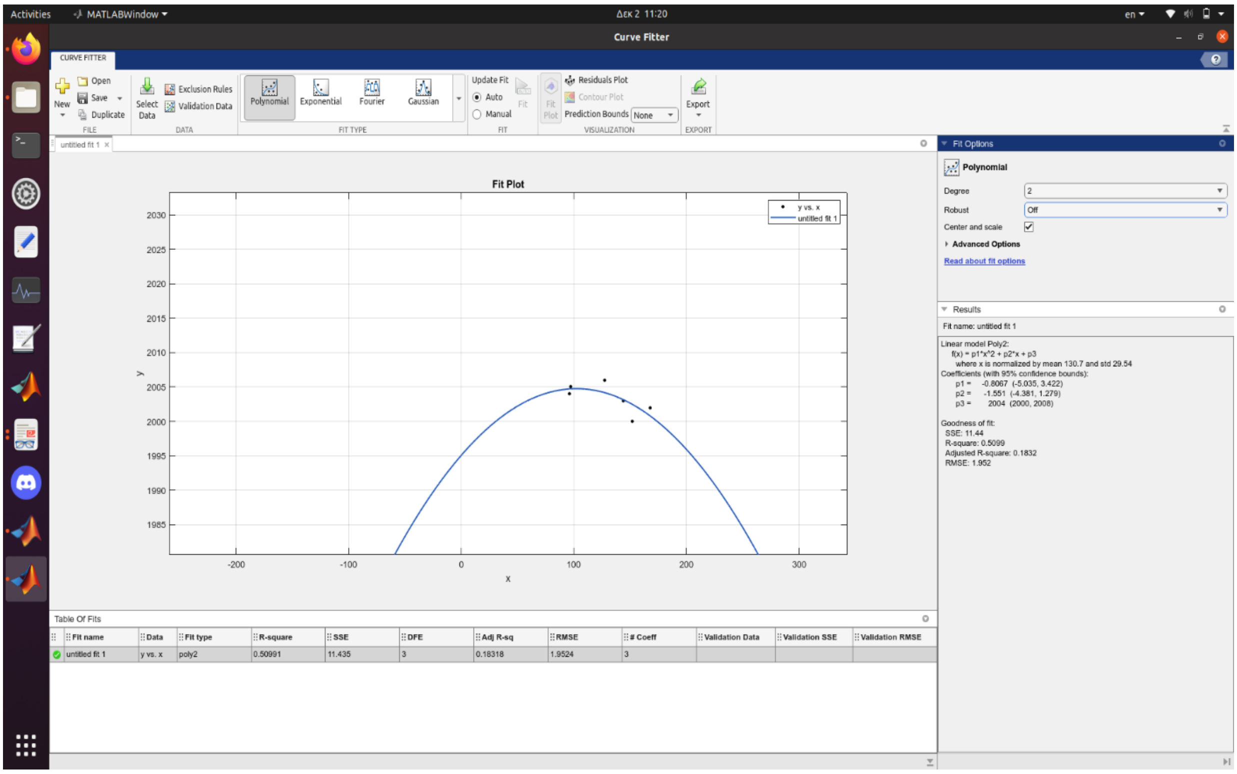
Χρήση του εργαλείου Curve Fitter.

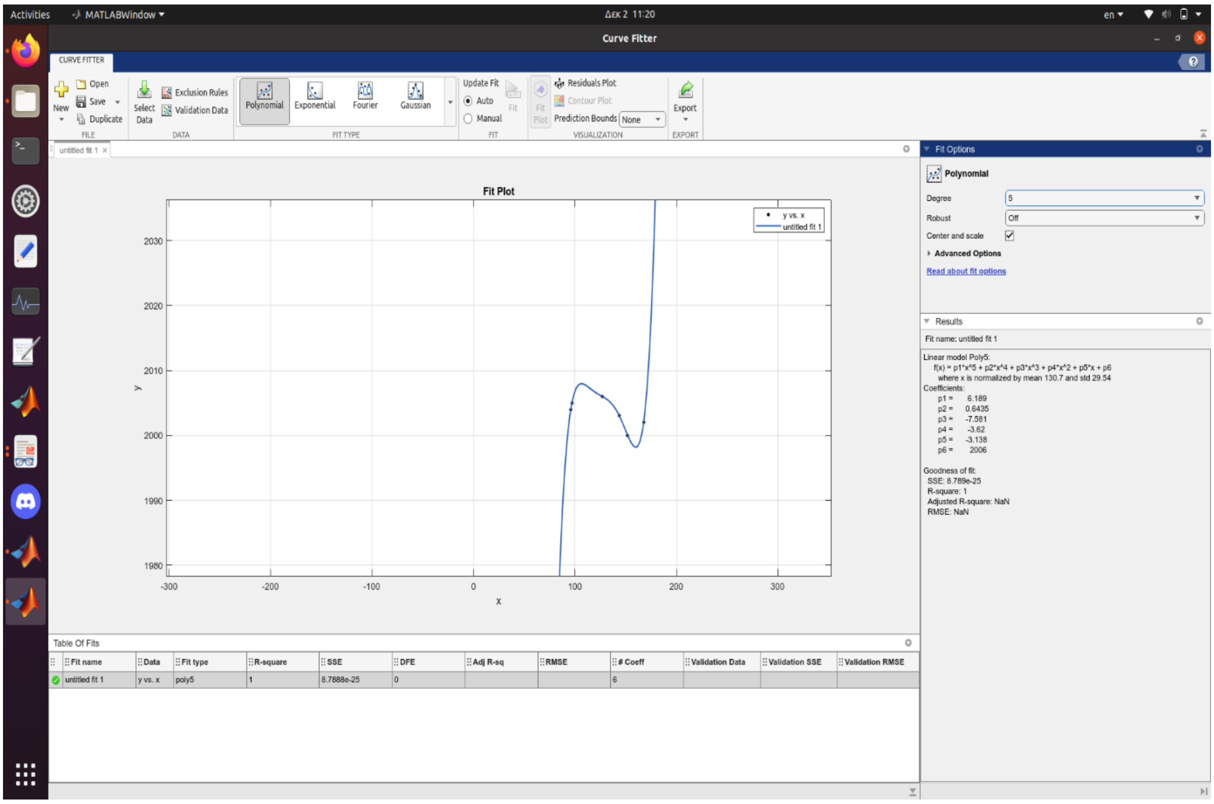
Κώδικας:
ask3_1.m

%Δεδομένα που προκύπτουν από το tsunamis.xlsx:

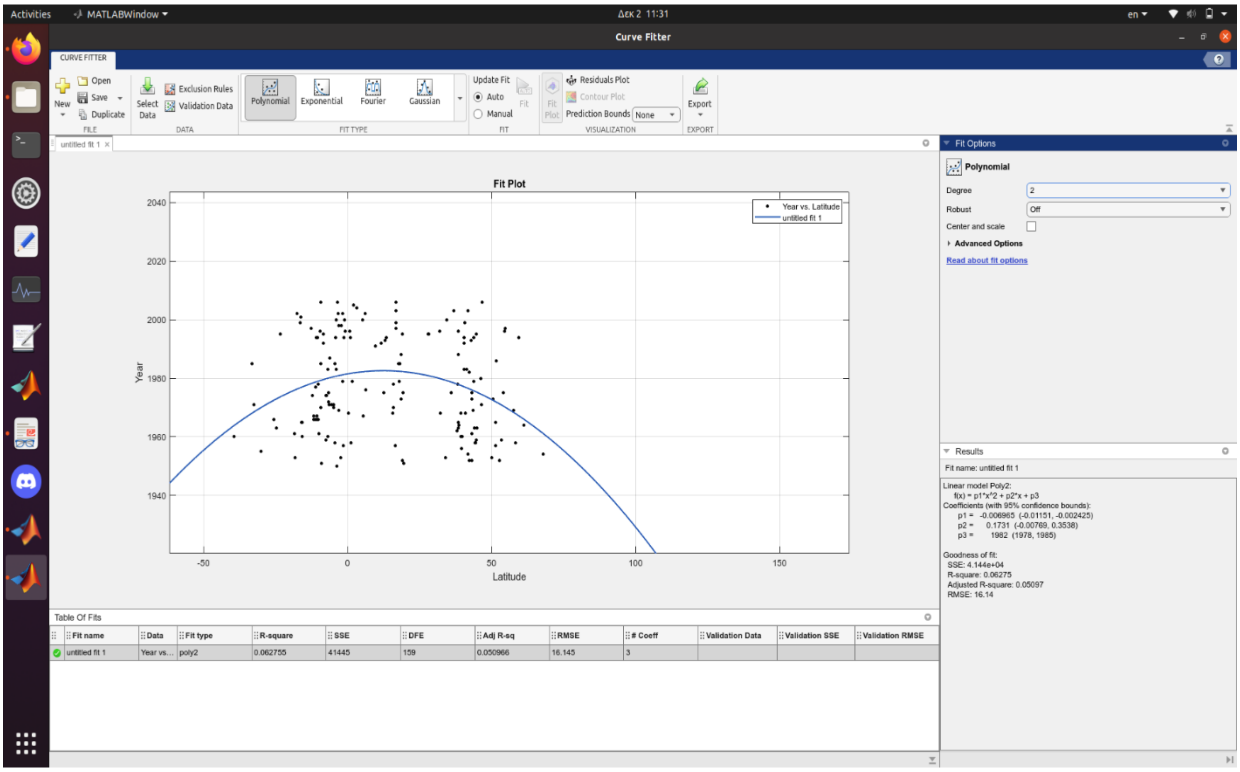
```
%Latitude       Year
%152.169000000000    2000
%167.856000000000    2002
%143.910000000000    2003
%95.9820000000000    2004
%97.1080000000000    2005
%127.214000000000    2006

%Η πολυωνυμική προσαρμογή ελαχίστων τετραγώνων υλοποιείται από τη
συνάρτηση polyfit:
%p=polyfit(x,y,n)
%η οποία βρίσκει τους συντελεστές του πολυωνύμου ελαχίστων τετραγώνων p(x)
βαθμού n
%που προσαρμόζεται στα σημεία (x(i), y(i)). Το αποτέλεσμα είναι ένα διάνυσμα-
γραμμή με
%τους συντελεστές του πολυωνύμου σε φθίνουσα διάταξη.
%Για να δώσουμε ένα γράφημα της παρεμβολής, πρέπει να παράγουμε σημεία στον
άξονα
%των x και με βάση αυτά να παράγουμε τις τιμές του p(x) με τη συνάρτηση polyval.

x=[152.169000000000,167.856000000000,143.910000000000,95.9820000000000,97
.1080000000000,127.214000000000];
y=[2000,2002,2003,2004,2005,2006];

p2=polyfit(x,y,2); %κατασκευή πολυώνυμο ελαχ. τετραγώνων 2ου βαθμού

p5=polyfit(x,y,5); %κατασκευή πολυώνυμο ελαχ. τετραγώνων 5ου βαθμού

xi=linspace(-1,5,100); %δημιουργία δεδομένων στον άξονα των x
z2=polyval(p2,xi); %δημιουργία δεδομένων στον άξονα των y για το p2
z5=polyval(p5,xi); %δημιουργία δεδομένων στον άξονα των y για το p5

plot(x,y,'o',xi,z2,':',xi,z5) %σχεδίαση γραφημάτων για τα p2 και p5
%το γράφημα του p2 είναι με διακεκομμένη γραμμή

xlabel('x'),ylabel('y=f(x)'), %ενδείξεις στους άξονες
title('2nd and 5th order Curve Fitting') %έκδοση τίτλου γραφήματος
```

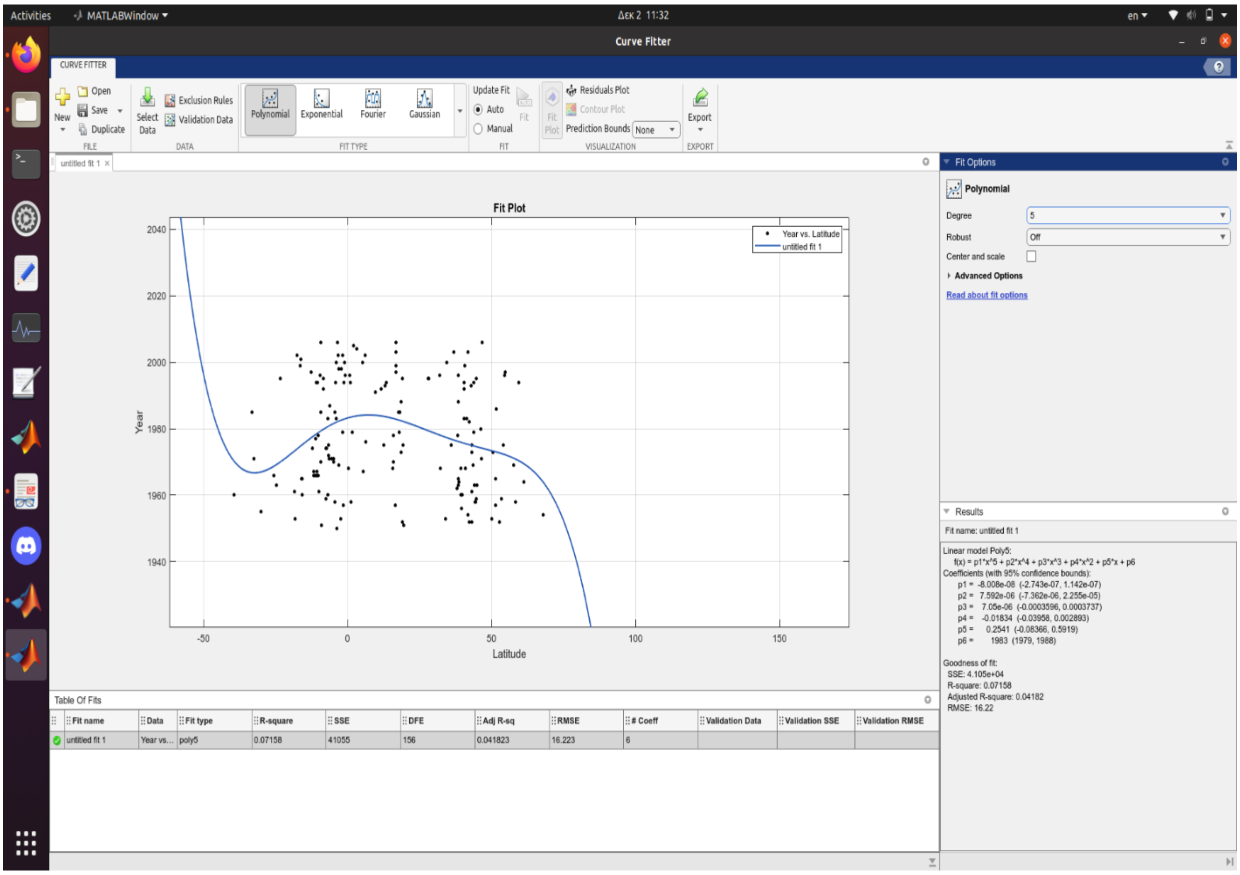
6 δεδομένα Latitude, Year





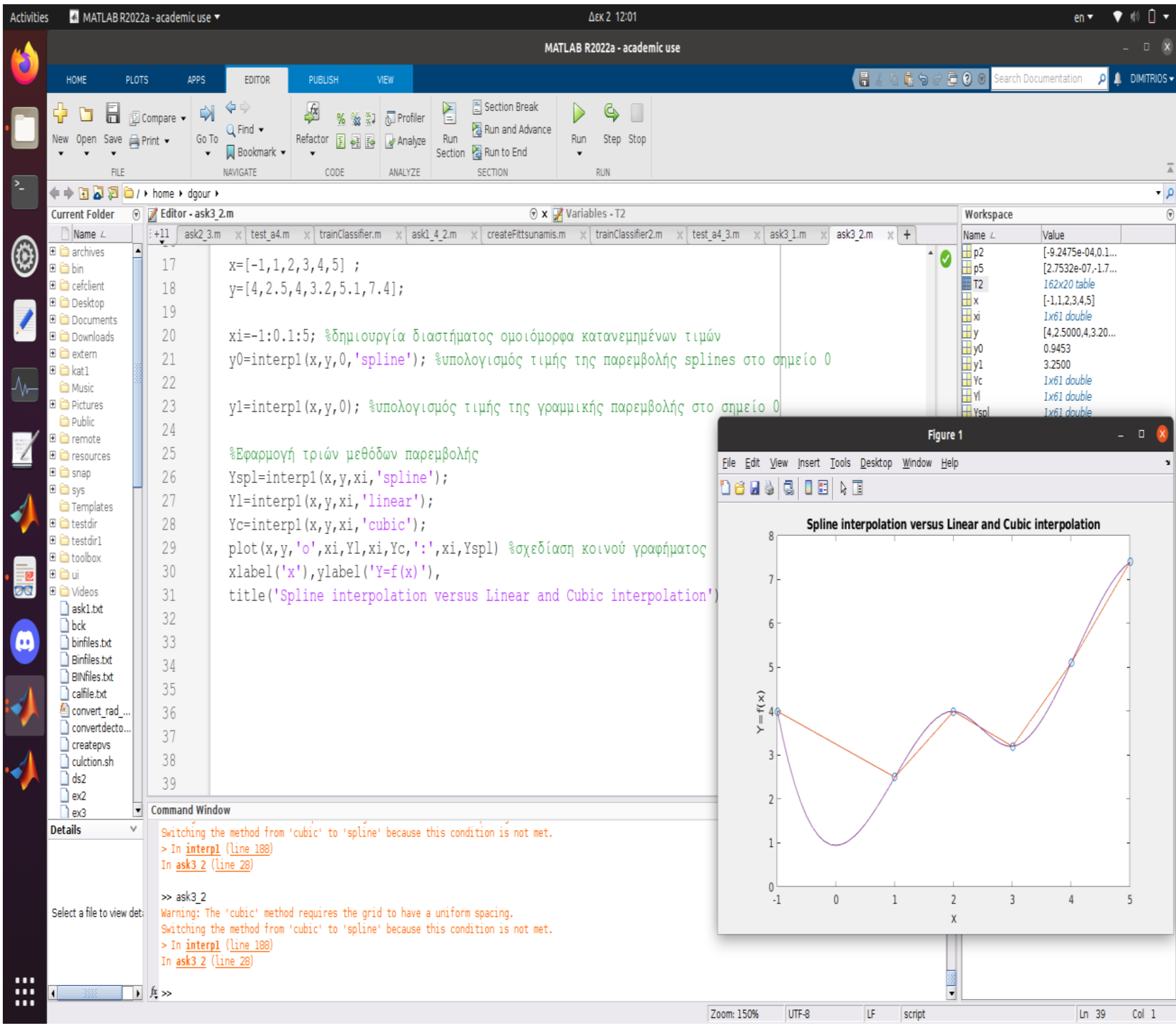
Όλα τα δεδομένα Latitude, Year





Γ.2. Άσκηση #2 Παρεμβολή.

Παράδειγμα εργαστηρίου:



Κώδικας:

ask3_2.m

%Η παρεμβολή μονοδιάστατων δεδομένων υλοποιείται από τη εντολή interp1. Τα
%σημεία (xi, yi) που θα παρεμβληθούν πρέπει να αποθηκευθούν σε δύο διανύσματα
x
%(για τα xi)και y (για τα yi).
%• yi=interp1(x,y,xi): υπολογίζει στο διάνυσμα yi τις τιμές που αντιστοιχούν στα
%δεδομένα του διανύσματος xi και οι οποίες καθορίζονται έπειτα από εφαρμογή
%γραμμικής παρεμβολής στα σημεία (xi,yi) των διανυσμάτων x και y.

%• yi=interp(x,y,xi, %method’): προσδιορίζει την μέθοδο παρεμβολής που θα
%εφαρμοσθεί, όπου method
%μπορεί να είναι:
%-%linear’ για γραμμική παρεμβολή
%-%spline’ για παρεμβολή με splines
%-%cubic’ για κυβική παρεμβολή. Εδώ απαιτείται οι τιμές του x να είναι
ομοιόμορφα
%κατανεμημένες.
%Σε όλες τις περιπτώσεις, οι τιμές του διανύσματος x θα πρέπει να δίνονται
%διατεταγμένες.

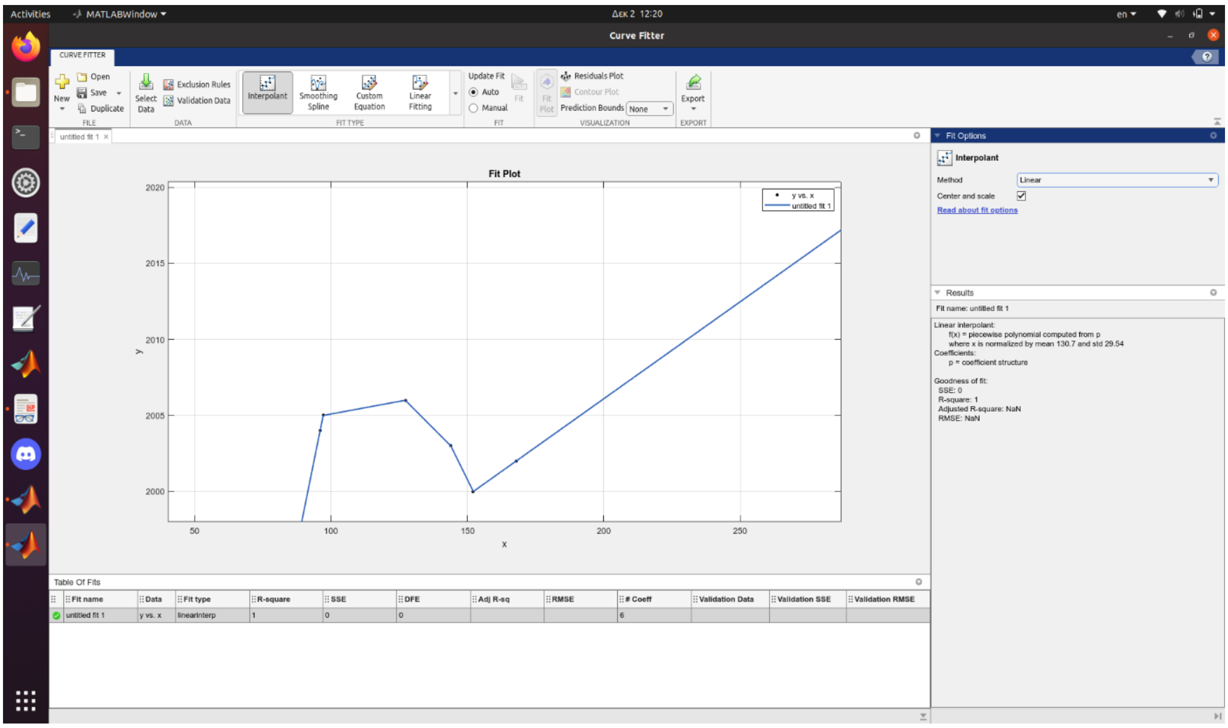
x=[152.169000000000,167.856000000000,143.910000000000,95.9820000000000,97
.1080000000000,127.214000000000];
y=[2000,2002,2003,2004,2005,2006];

xi=-1:0.1:5; %δημιουργία διαστήματος ομοιόμορφα κατανεμημένων τιμών
y0=interp1(x,y,0,'spline'); %υπολογισμός τιμής της παρεμβολής splines στο σημείο 0

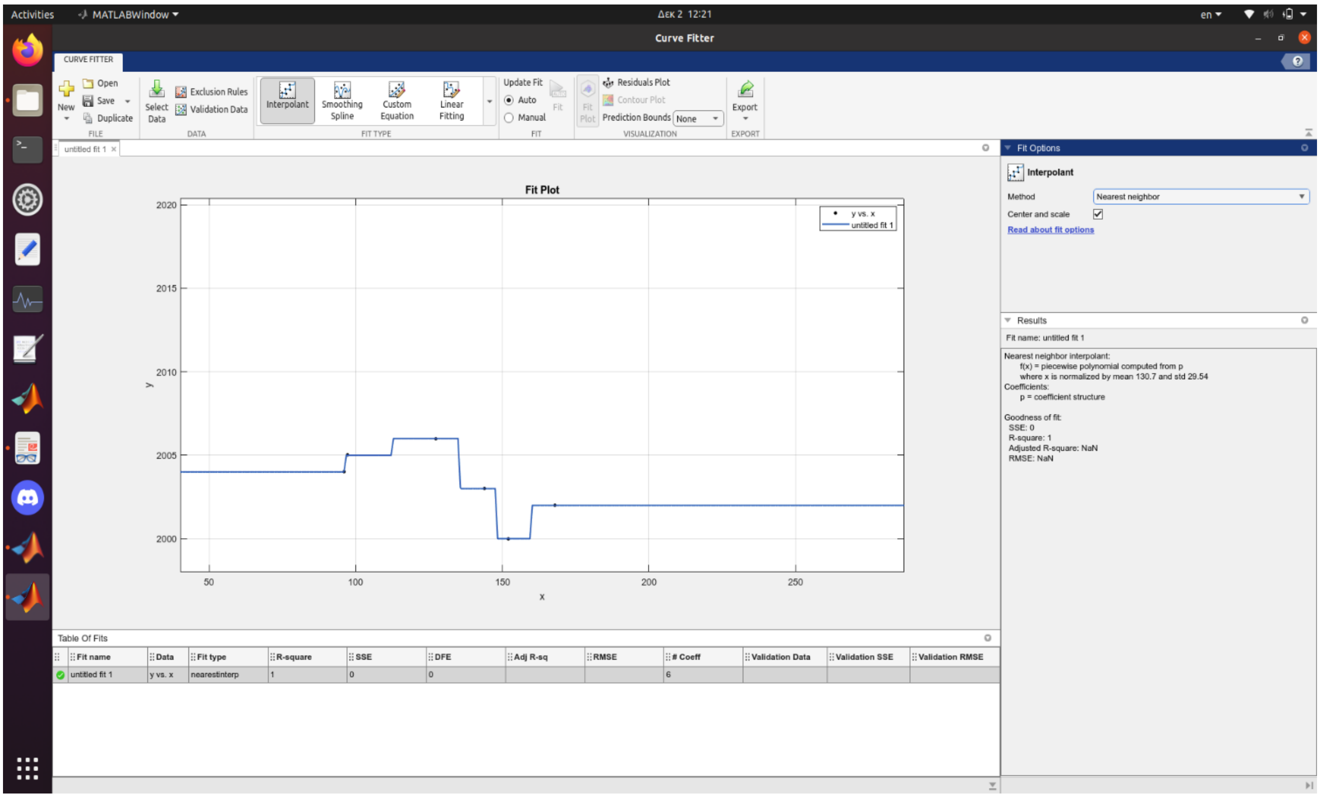
y1=interp1(x,y,0); %υπολογισμός τιμής της γραμμικής παρεμβολής στο σημείο 0

%Εφαρμογή τριών μεθόδων παρεμβολής
Yspl=interp1(x,y,xi,'spline');
Yl=interp1(x,y,xi,'linear');
Yc=interp1(x,y,xi,'cubic');
plot(x,y,'o',xi,Yl,xi,Yc,':',xi,Yspl) %σχεδίαση κοινού γραφήματος
xlabel('x'),ylabel('Y=f(x)'),
title('Spline interpolation versus Linear and Cubic interpolation')

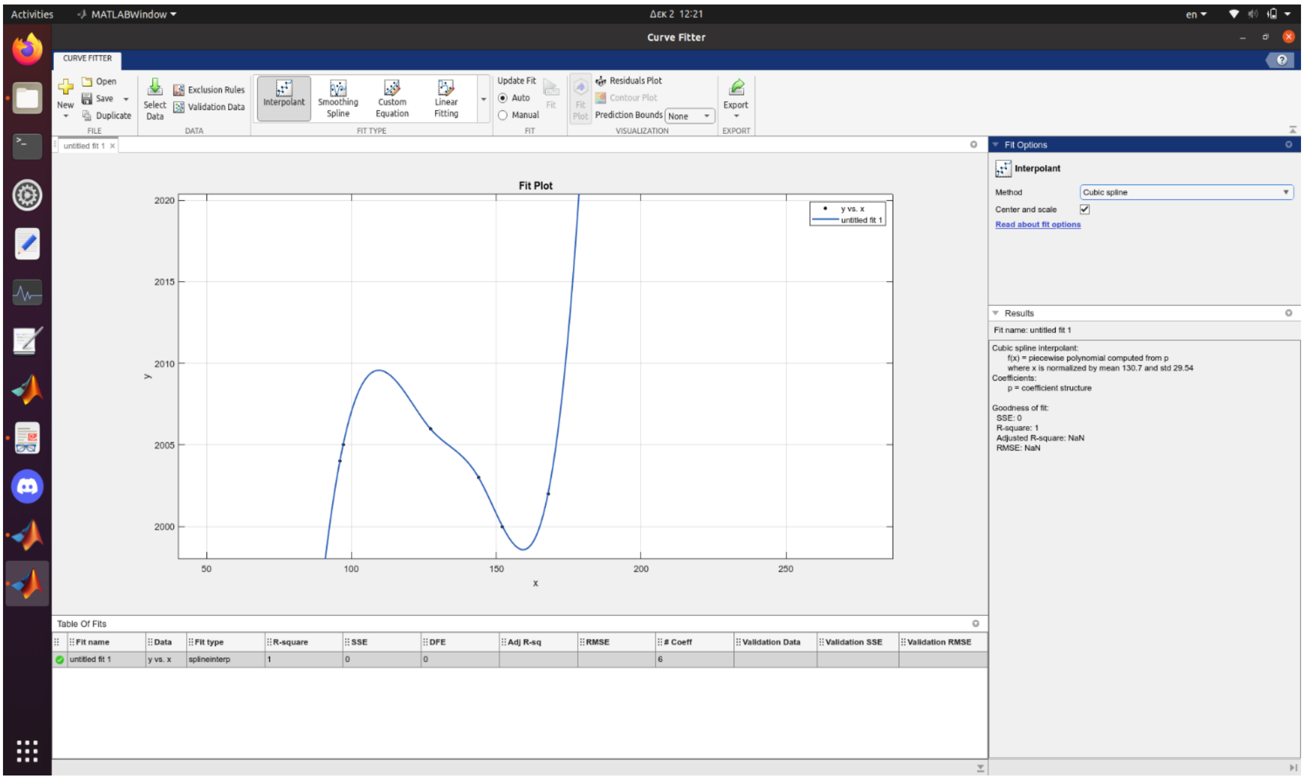
6 δεδομένα Latitude, Year
Linear



Nearest Neighbor



Cublic Spline



Γ.3. Άσκηση #2 Οπτικοποίηση και επεξεργασία πολυδιάστατων δεδομένων.

Κώδικας:

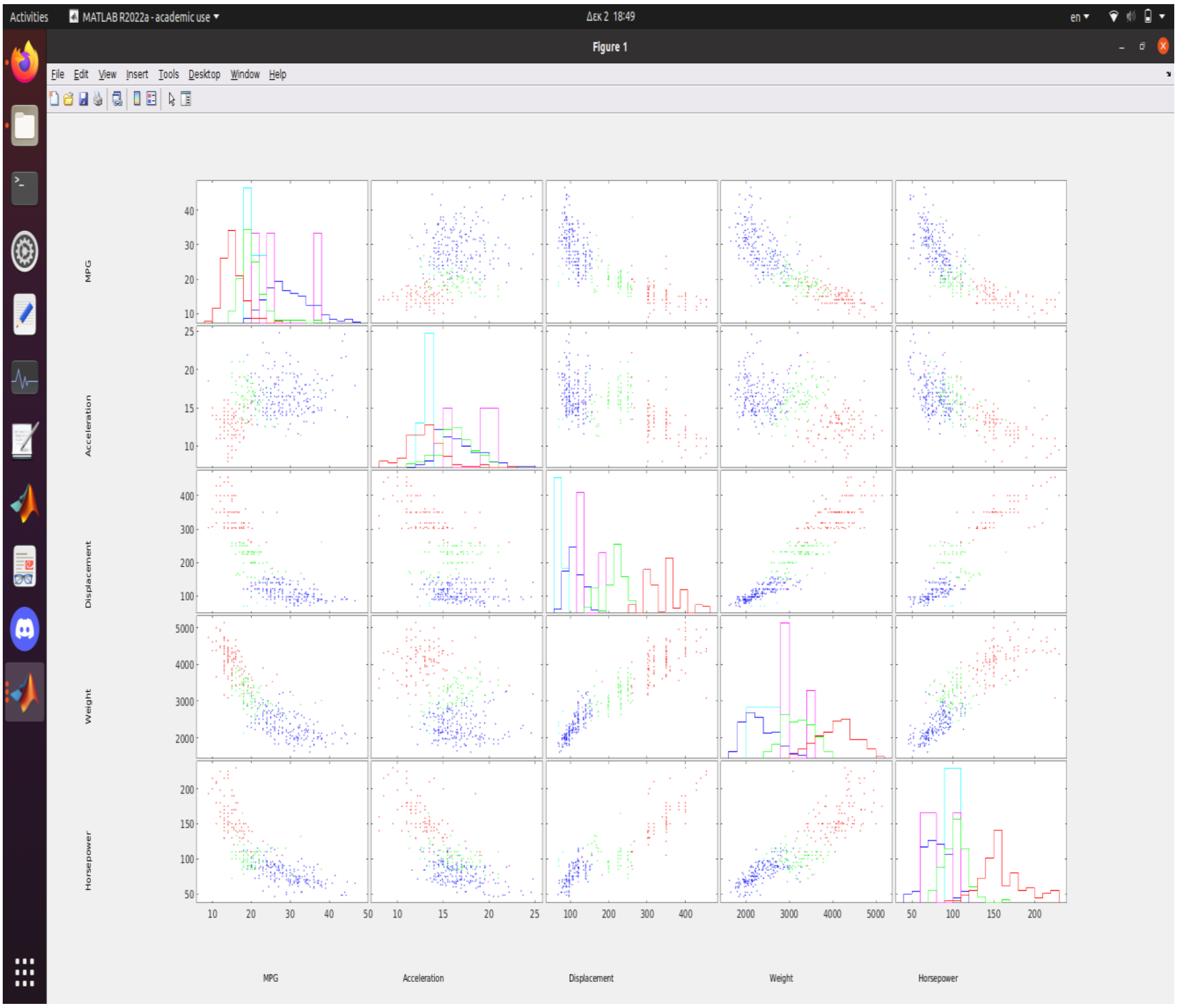
ask3_3.m

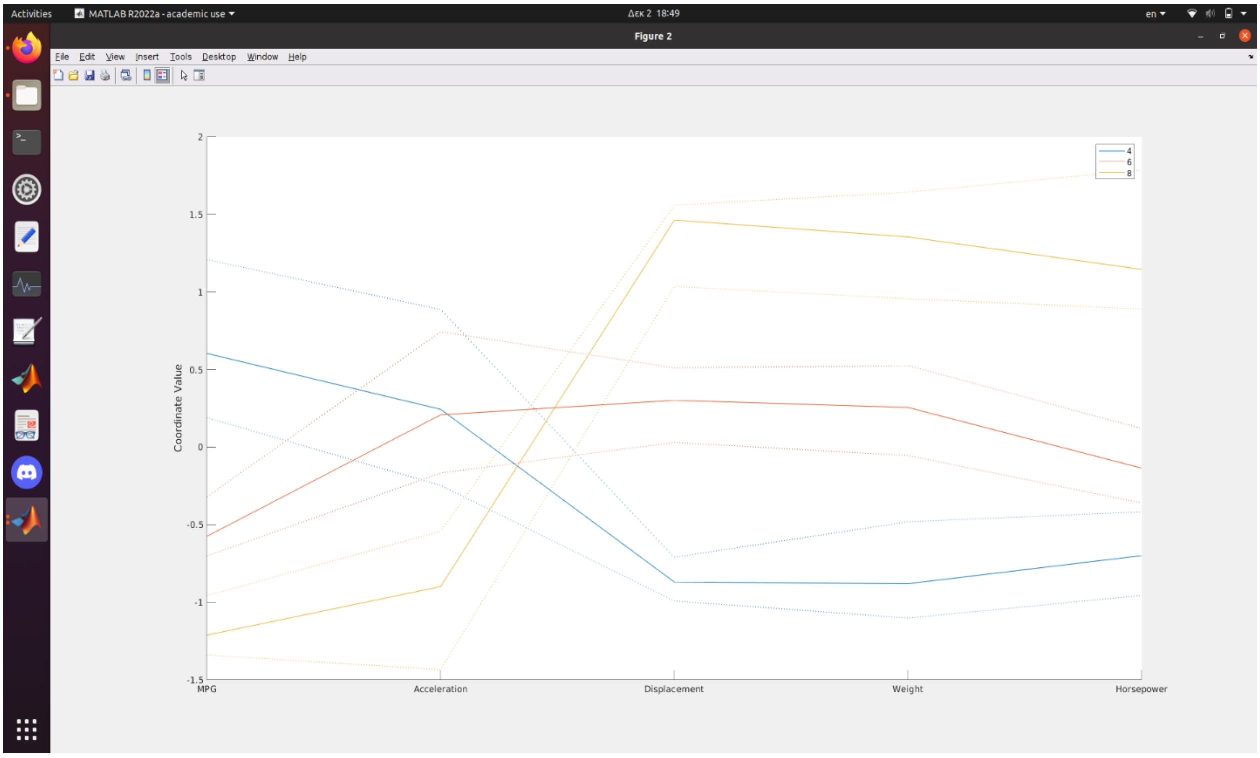
```
load carbig

X = [MPG,Acceleration,Displacement,Weight,Horsepower];
varNames = {'MPG'; 'Acceleration'; 'Displacement'; 'Weight'; 'Horsepower'};

figure(1)
gplotmatrix(X,[],Cylinders,['c' 'b' 'm' 'g' 'r'],[],[],false);
text([.08 .24 .43 .66 .83], repmat(-.1,1,5), varNames, 'FontSize',8);
text(repmat(-.12,1,5), [.86 .62 .41 .25 .02], varNames, 'FontSize',8,'Rotation',90);

figure(2)
Cyl468 = ismember(Cylinders,[4 6 8]);
parallelcoords(X(Cyl468,:), 'group',Cylinders(Cyl468), ...
'standardize','on', 'labels',varNames)

parallelcoords(X(Cyl468,:), 'group',Cylinders(Cyl468), ...
'standardize','on', 'labels',varNames, 'quantile',.25)

% ----------------------------------------------------------------------------------------------------
```

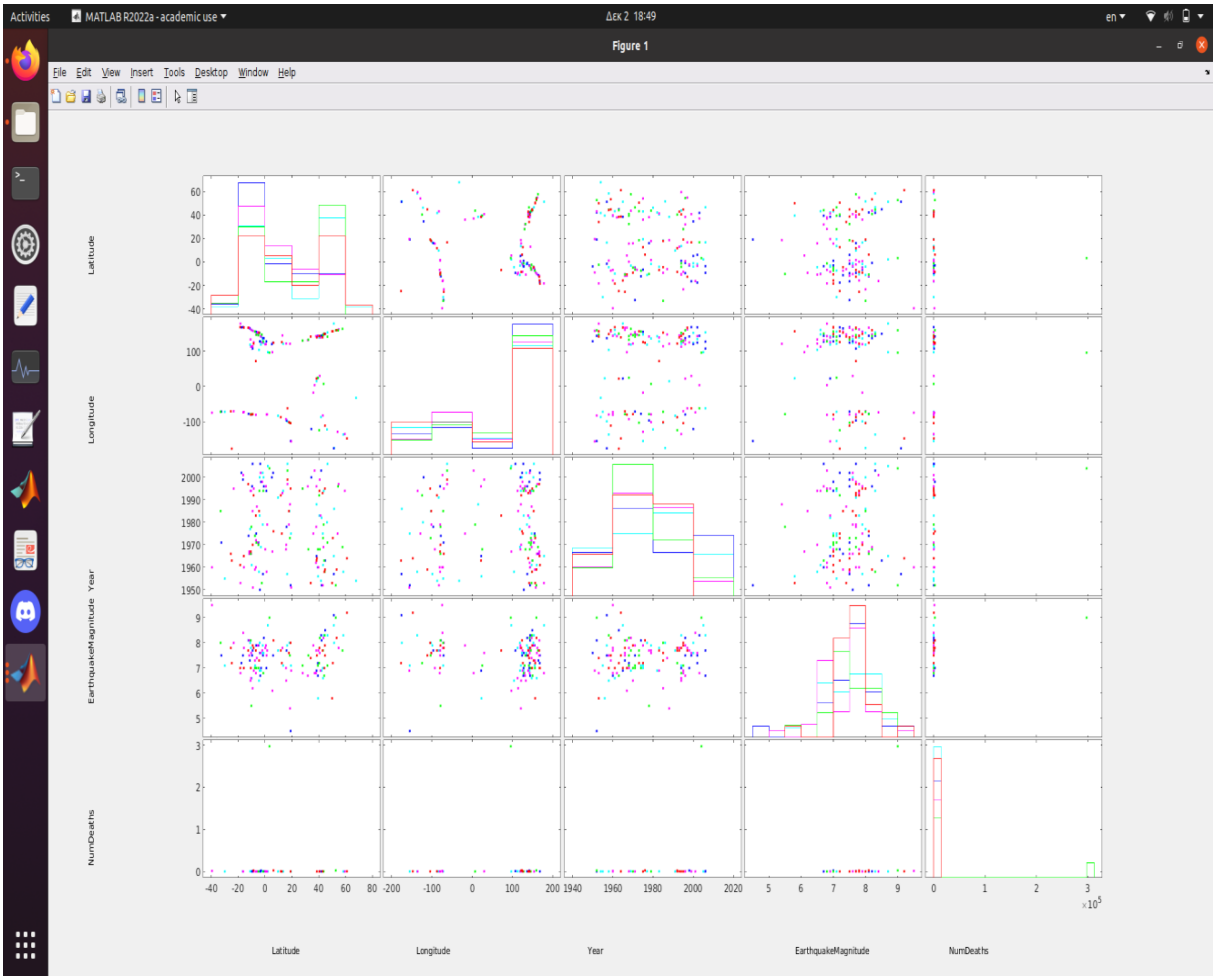
ask3_3_1.m

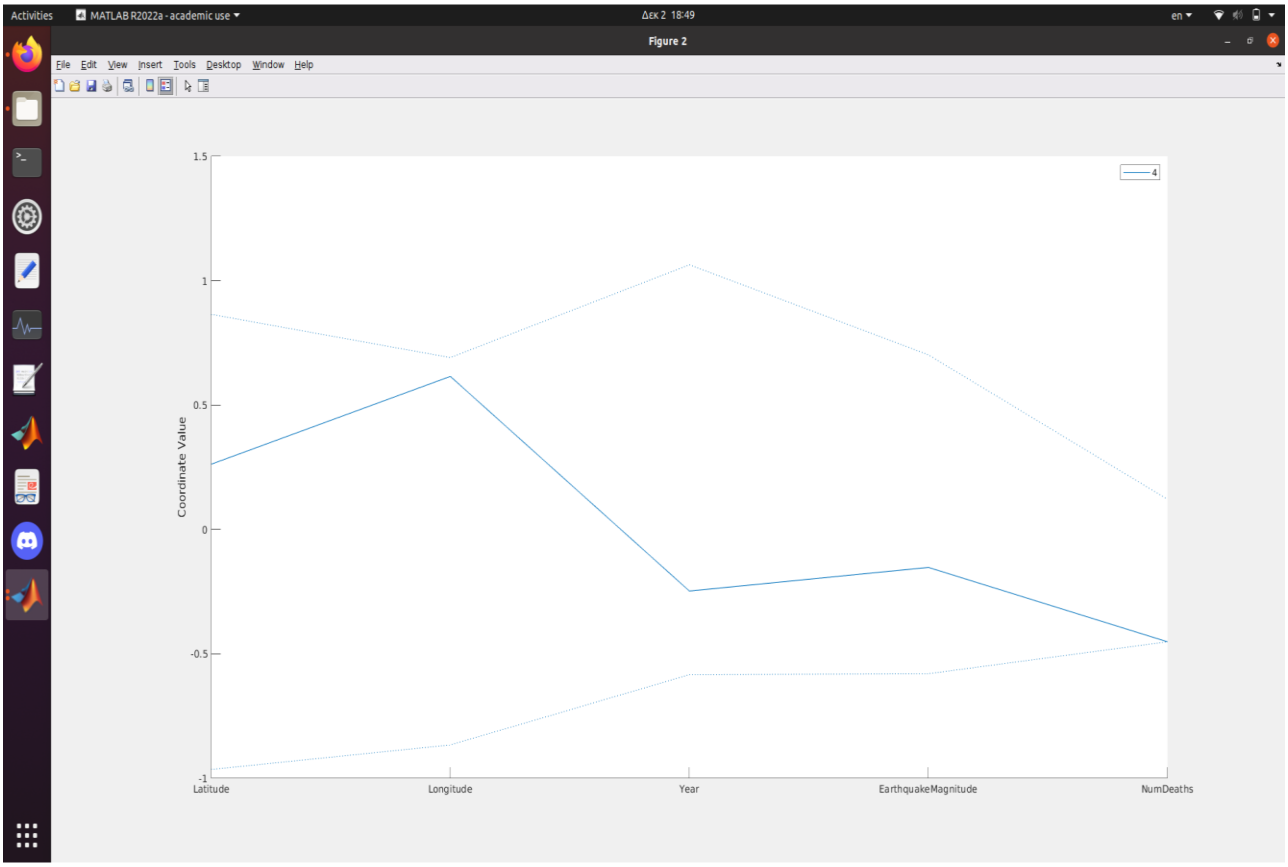
```
X = [T2.Latitude,T2.Longitude,T2.Year,T2.EarthquakeMagnitude,T2.NumDeaths];
varNames = {'Latitude'; 'Longitude'; 'Year'; 'EarthquakeMagnitude'; 'NumDeaths'};

n = size(X, 1);
Cylinders = randi([1, 5], n, 1);
```

```
figure(1)
gplotmatrix(X, [], Cylinders, ['c' 'b' 'm' 'g' 'r'], [], [], false);
text([.08 .24 .43 .66 .83], repmat(-.1,1,5), varNames, 'FontSize',8);
text(repmat(-.12,1,5), [.86 .62 .41 .25 .02], varNames, 'FontSize',8,'Rotation',90);

figure(2)
Cyl468 = ismember(Cylinders,[4 6 8]);
parallelcoords(X(Cyl468,:), 'group',Cylinders(Cyl468), ...
'standardize','on', 'labels',varNames)

parallelcoords(X(Cyl468,:), 'group',Cylinders(Cyl468), ...
'standardize','on', 'labels',varNames, 'quantile',.25)
```




Activities
MATLAB R2022a - academic use
Δεκ 2 18:49
Figure 2
File Edit View Insert Tools Desktop Window Help
Coordinate Value
MPG
Acceleration
Displacement
Weight
Horsepower
4
6
8


Activities
MATLAB R2022a - academic use
Δεκ 2 18:49
Figure 1
File Edit View Insert Tools Desktop Window Help
Latitude
Longitude
Year
EarthquakeMagnitude
NumDeaths


Activities
MATLAB R2022a - academic use
Δεκ 2 18:49
en
Figure 2
File Edit View Insert Tools Desktop Window Help
1.5
1
0.5
0
-0.5
-1
Coordinate Value
4
Latitude
Longitude
Year
EarthquakeMagnitude
NumDeaths