ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΚΑΙ

ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΥΠΟΛΟΓΙΣΤΙΚΗ ΝΟΗΜΟΣΥΝΗ

(ΕΡΓΑΣΤΗΡΙΟ)

ΕΑΡΙΝΟ ΕΞΑΜΗΝΟ 2023-2024

2η Σειρά ασκήσεων

Ονομ/νυμα και αρ.μητρώου

___________ΓΕΩΡΓΙΟΣ ΣΙΓΑΛΑΣ___________

________________18387102____________________

Εργαστηριακή ομάδα

________________ΟΜΑΔΑ Γ___________________

Αιγάλεω _16_/_4_/_2024_

1. Σκοπός και περίληψη της άσκησης

Η δεύτερη εργαστηριακή άσκηση πραγματεύεται εύρεση γραμμικών μοντέλων για προβλέψεις μεγεθών με βάση των εισόδων του συστήματος. Στην πρώτη άσκηση δίνονται δεδομένα σε μορφή αρχείου, για τα δεδομένα αυτά ισχύει η μορφή της συνάρτησης ‘y = ax + b’ και με βάση τυπολογικό που δίνεται γίνεται εύρεση των συντελεστών a και b μέσω της μεθόδου των ελαχίστων τετράγωνων αλλά και συντελεστές ποιότητας με τους οποίος χαρακτηρίζουμε άμα έχουμε προσεγγίσει σωστά τους συντελεστές της εξίσωσης. Εφόσον έχουμε την εξίσωση για το γραμμικό μοντέλο μπορούμε να βάλουμε ως εισόδους τα επιθυμητά μεγέθη και να προβλέψουμε την συμπεριφορά του συστήματος. Αντίστοιχα και η δεύτερη άσκηση έχει παρόμοια διαδικασία υλοποίησης, με την διαφορά ότι πλέον έχουμε περισσότερες εισόδους άρα αυξάνεται η περιπλοκότητα εφόσον έχουμε περισσότερους συντελεστές βαρύτητας (a) να υπολογίσουμε άρα πλέον αναλύεται ένα πολυδιάστατο γραμμικό μοντέλο. Πιο συγκεκριμένα στην συνέχεια της εργαστηριακής άσκησης θα αναλυθούν γραφήματα σύγκρισης των πραγματικών τιμών με βάση τα δεδομένα και των προβλέψεων που θα υλοποιηθούν και σχολιασμός των σφαλμάτων της κάθε διαδικασίας.

2. Διαδικασία υλοποίησης της άσκησης

Άσκηση 1: Συχνοτική απόκριση ενισχυτή

Στην εργαστηριακή άσκηση περιέχεται ένα αρχείο-πίνακας 'ampldat.mat' το οποίο περιέχει τα δεδομένα ενίσχυσης-συχνότητας ενός ενισχυτή. Στην πρώτη στήλη του πίνακα βρίσκονται τα δεδομένα συχνότητας (Hz) και στην δεύτερη τα δεδομένα ενίσχυσης (dB). Για περιοχή συχνοτήτων 1kHz-10MHz η σχέση που ενώνει τα 2 μεγέθη από τις 2 στήλες του αρχείου είναι:

- y = ax + b (1)

y: Ενίσχυση , x: Δεκαδικός λογάριθμος της συχνότητας.

A) Στο πρώτο ερώτημα της άσκησης κατασκευάζεται ένα Matlab πρόγραμμα τύπου function το οποίο δέχεται σαν εισόδους τα δεδομένα για την ενίσχυση και την συχνότητα (x,y) και επιστρέφει σαν εξόδους τα παρακάτω (a,b,yp,mare,r2):

i. Τους συντελεστές a και b την κύριας εξίσωσης (1)

ii. Τις προβλέψεις 'yp' για όλα τα διαθέσιμα δεδομένα συχνότητας σύμφωνα με το μοντέλο yp = ax + b.

iii. Το μέσο απόλυτο σφάλμα % (MARE%) και τον συντελεστή R^2.

Το πρόγραμμα βρίσκεται στην εικόνα 2.2 και υπάρχουν σχόλια σε κάθε εντολή η οποία χρειάζεται διευκρινήσεις.

Το μέσο απόλυτο σφάλμα % (MARE%) μας δίνει το σφάλμα σχετικό με την πραγματική τιμή και το πεδίο τιμών του είναι (0,+∞), εξ' ορισμού ιδανικά το μέσο απόλυτο σφάλμα πρέπει να είναι κοντά στο μηδέν. Αντίστοιχα το πεδίο τιμών του συντελεστή R^2 είναι (-∞,1) και ιδανικά θέλουμε έναν μεγάλο συντελεστή κοντά στην τιμή 1. Παρακάτω στην εικόνα 2.1 είναι όλοι οι τύποι για τον υπολογισμό των συντελεστών του ερωτήματος (iii) οι οποίοι υπάρχουν αναλυτικά και στο πρόγραμμα.

$$a = \frac{SS_{xy}}{SS_{xx}}, \text{ όπου } SS_{xy} = \sum_{j=1}^{P} x_j y_j - \frac{\sum_{j=1}^{P} x_j \sum_{j=1}^{P} y_j}{P}, \quad SS_{xx} = \sum_{j=1}^{P} (x_j)^2 - \frac{\left(\sum_{j=1}^{P} x_j\right)^2}{P}$$

$$b = \bar{y} - a\bar{x}, \quad \text{όπου} \quad \bar{y} = \frac{\sum_{j=1}^{P} y_j}{P}, \quad \bar{x} = \frac{\sum_{j=1}^{P} x_j}{P}$$

$$MARE\% = 100 \frac{\sum_{j=1}^{P} \frac{|y_j - \hat{y}_j|}{y_j}}{P} \quad \text{και}$$

$$R^2 = 1 - \frac{SSE}{SST}, \text{ όπου } SSE = \sum_{j=1}^{P} (y_j - \hat{y}_j)^2 \text{ και } SST = \sum_{j=1}^{P} (y_j - \bar{y})^2$$

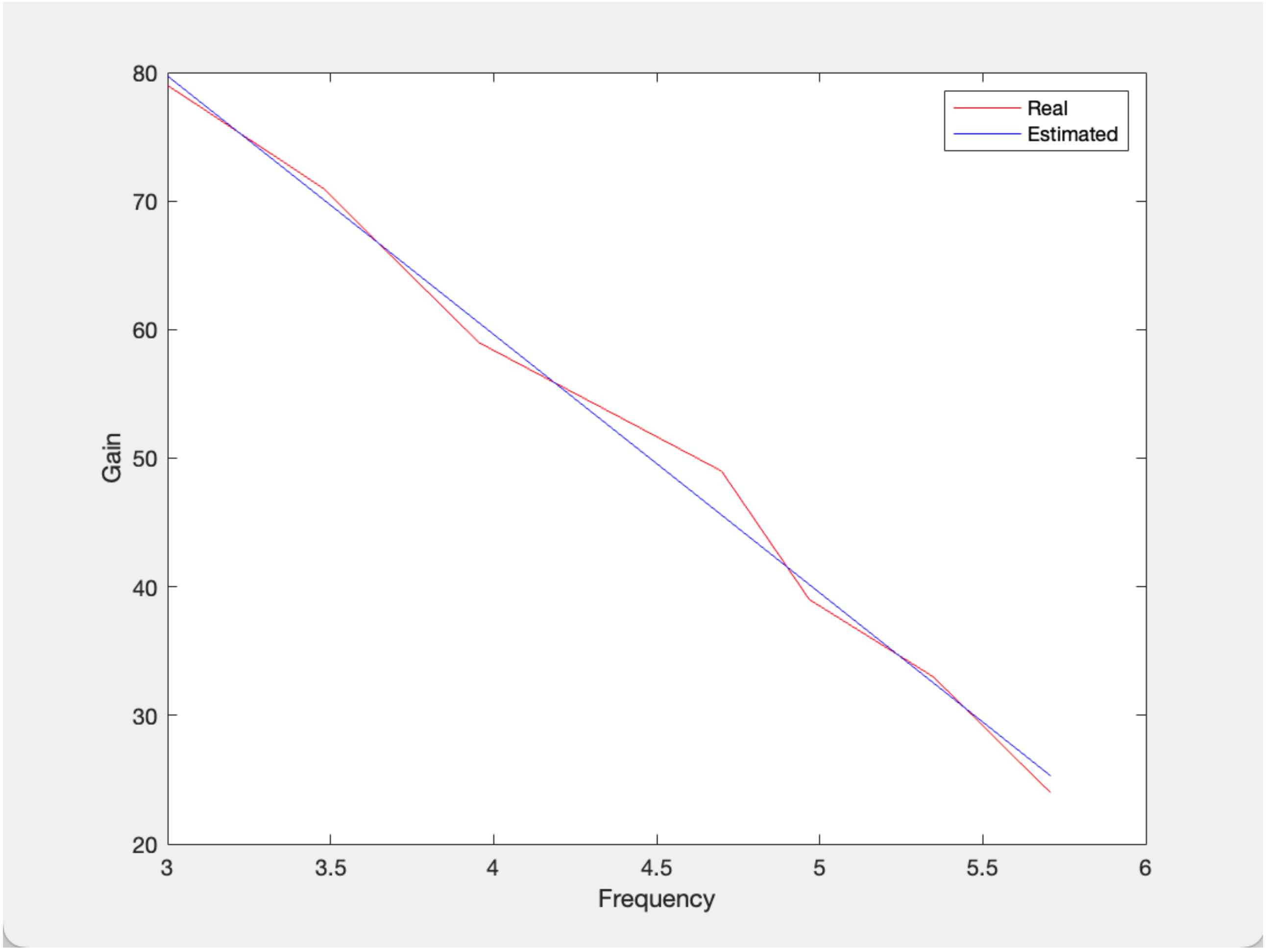
Εικόνα 2.1. Τύποι για τον υπολογισμό των συντελεστών

```
function [a,b,yp,mare,r2]=func(x,y)


p=length(x); %Ο αριθμος των δεδομενων (p) να ειναι ισος οσος ο αριθμος των στοιχειων
             %του διανυσματος x το οποιο περιεχει τα δεδομενα,
             %θα μπορουσαμε να βαλουμε και του y
ssxx = sum(x.^2,[1 p]) - (sum(x,[1 p]).^2/p);
ssxy = sum((x.*y),[1 p]) - ((sum(x,[1 p])*sum(y,[1 p]))/p);
a = ssxy/ssxx;

yavg = sum(y,[1 p])/p; %Υπολογισμος της μεσης τιμης του y
xavg = sum(x,[1 p])/p; %Υπολογισμος της μεσης τιμης του x

b = yavg - (a*xavg);

yp = (a*x)+b; %Υπολογισμος της προβλεψης βαση για ολα τα διαθεσιμα δεδομενα

mare = 100*(sum(abs(y-yp)/y)/p); %Μεσο απολυτο σχετικο σφαλμα

sse = sum((y-yp).^2,[1 p]);
sst = sum((y-yavg).^2,[1 p]);

r2 = 1 - (sse/sst); %Υπολογισμος του συντελεστη R^2
```

Εικόνα 2.2. Το πρόγραμμα Matlab τύπου function για τον υπολογισμό όλων των συντελεστών

B) Στο δεύτερο ερώτημα της πρώτης άσκησης γίνεται υλοποίηση της γραφικής παράστασης με άξονες την ενίσχυση (άξονας y) και τον λογάριθμο της συχνότητας (άξονας x), η οποία απεικονίζει 2 γραφήματα: (Εικόνα 2.3)

- Τα δεδομένα ενίσχυσης-συχνότητας από τα δεδομένα, δηλαδή οι πραγματικές τιμές (κόκκινη γραμμή).
- Η βέλτιστη γραμμή που υπολογίστηκε μέσω της μεθόδου ελαχίστων τετράγωνων δηλαδή οι προβλέψεις (yp) που επιστρέφει η συνάρτηση ‘func’ (μπλε γραμμή).



Εικόνα 2.3. Δεδομένα ενίσχυσης-συχνότητας

Γ) Στο τρίτο και τελευταίο κομμάτι της πρώτης άσκησης γίνεται υλοποίηση ενός ακόμα function στο Matlab (func2), το οποίο δέχεται σαν εισόδους τους συντελεστές a και b που υπολογίστηκαν με την μέθοδο ελαχίστων τετράγωνων μέσω της συνάρτησης ‘func’ καθώς και μια οποιαδήποτε τιμή ενίσχυσης (dB) και επιστρέφει σαν έξοδο τη συχνότητα (Hz) στην οποία αντιστοιχεί αυτή η τιμή ενίσχυσης. (Εικόνα 2.4)

ask_2_main.m | func.m | func2.m | ask_2_2_main.m | func3.m

```
function x = func2(a,b,y)

k = (y-b)/a;

x = 10^k;
```

Εικόνα 2.4. Function ερωτήματος ‘Γ’

Παρακάτω υπάρχει το συνολικό πρόγραμμα (Εικόνα 2.5) στο οποίο υλοποιούνται και τα 3 ερωτήματα της άσκησης και γίνεται κλήση όλων των συναρτήσεων που αναφέρθηκαν παραπάνω. Πιο συγκεκριμένα για το Γ ερώτημα υπολογίζεται η συχνότητα (g) για τιμή ενίσχυσης ίση με 0dB.

```
clc
clear
load ampldata.mat%Φορτωση του αρχειου που περιεχει τα δεδομενα

x = log10(dat(:,1));%Ο δεκαδικος λογαριθμος της συχνοτητας της πρωτης στηλης του αρχειου ampldata
y = dat(:,2);%Η ενισχυση της δευτερης στηλης του αρχειου ampldata

[a,b,yp,mare,r2] = func(x,y);%Κληση της συναρτησης του ερωτηματος Α

%Σχεδιαση της γραφικης παραστασης του ερωτηματος Β
plot(x,y,'r');%πραγματικες τιμες με κοκκινο χρωμα
hold on
plot(x,yp,'b');%προβλεψεις με μπλε χρωμα
legend('Real','Estimated');
xlabel('Frequency');
ylabel('Gain');

%Κληση της συναρτησης του ερωτηματος Β
g = func2(a,b,0);
```

a	-20.0973
b	140.0178
dat	7x2 double
g	9.2684e+06
mare	[1.7180,0,0,0,0,0,0]
r2	0.9924
x	[3;3.4771;3.9542;4.6990;4...
y	[79;71;59;49;39;33;24]
yp	[79.7260;70.1371;60.5483...

Εικόνα 2.5. Το script της άσκησης 1.

Άσκηση 2: Πρόβλεψη απόδοσης CPU.

Στην $2^{η}$ εργαστηριακή άσκηση κατασκευάζεται ένα γραμμικό μοντέλο το οποίο προβλέπει την σχετική απόδοση ενός επεξεργαστή με βάση τα χαρακτηριστικά του συστήματος. Το αρχείο excel που περιέχει τα δεδομένα με χαρακτηριστικά επεξεργαστών και την σχετική απόδοση τους που θα χρησιμοποιηθεί από το κυρίως πρόγραμμα ονομάζεται Machine_SPU.xlsx.

Το γραμμικό μοντέλο έχει πλέον επτά εισόδους στο σύστημα και είναι της μορφής :

$y = (\sum_{i=1}^{6} a(i)x(i)) + b$, όπου :

y: Η σχετική απόδοση του επεξεργαστή
x_1: Χρόνος κύκλου (ns)
x_2: Ελάχιστη κεντρική μνήμη (KB)
x_3: Μέγιστη κεντρική μνήμη (KB)
x_4: Μνήμη Cache (KB)
x_5: Ελάχιστος αριθμός καναλιών
x_6: Μέγιστος αριθμός καναλιών

A) Στο πρώτο ερώτημα της άσκησης κατασκευάζεται ένα Matlab πρόγραμμα τύπου function το οποίο δέχεται σαν εισόδους τα χαρακτηριστικά του επεξεργαστή και την σχετική απόδοση και επιστρέφει σαν έξοδο:

- Τους συντελεστές ‘$α_i$’, i=1,2,…,6 και στον σταθερό όρο b.
- Τις προβλέψεις ‘yp’ για όλα τα διαθέσιμα δεδομένα σύμφωνα με το μοντέλο :

$$yp = (\sum_{i=1}^{6} a(i)x(i)) + b$$

- Το μέσο απόλυτο σχετικό σφάλμα % (MARE%) και τον συντελεστή R^2 .

Χρησιμοποιείται η μέθοδος των ελαχίστων τετράγωνων όπως και στην άσκηση 1 με την διαφορά ότι πλέον υπάρχουν περισσότερες είσοδοι άρα και περισσότερους συντελεστές (‘$α_i$’, i=1,2,…,6). Οι είσοδοι είναι στην μορφή πινάκων (X και Y) και οι συντελεστές (A) αντίστοιχα και υπολογίζονται σύμφωνα με τον τύπο:

$A = (X' \cdot X)^{-1} \cdot X' \cdot Y$ όπου :

X΄: O ανάστροφος πίνακας X
$X' \cdot X$: Πίνακας συνδιασποράς (covariance matrix), δείχνει την συσχέτιση των εισόδων μεταξύ τους.

Άρα ο γενικός τύπος της μεθόδου στην μορφή πινάκων είναι:

$$Y = A \cdot X$$

Για να υπολογίσουμε και τον σταθερό όρο ‘b’ που υπάρχει στην γενική μορφή της μεθόδου ελαχίστων τετράγωνων προσθέτουμε μια στήλη με άσσους στο τέλος του πίνακα Χ ώστε όταν γίνει ο πολλαπλασιασμός των στοιχείων με τον πίνακα των συντελεστών (Α), ο σταθερός όρος b να υπάρχει ως τελευταία μεταβλητή στο διάνυσμα Y. Παρακάτω υπάρχει το πρόγραμμα func3 που αναλύθηκε παραπάνω (Εικόνα 2.6)

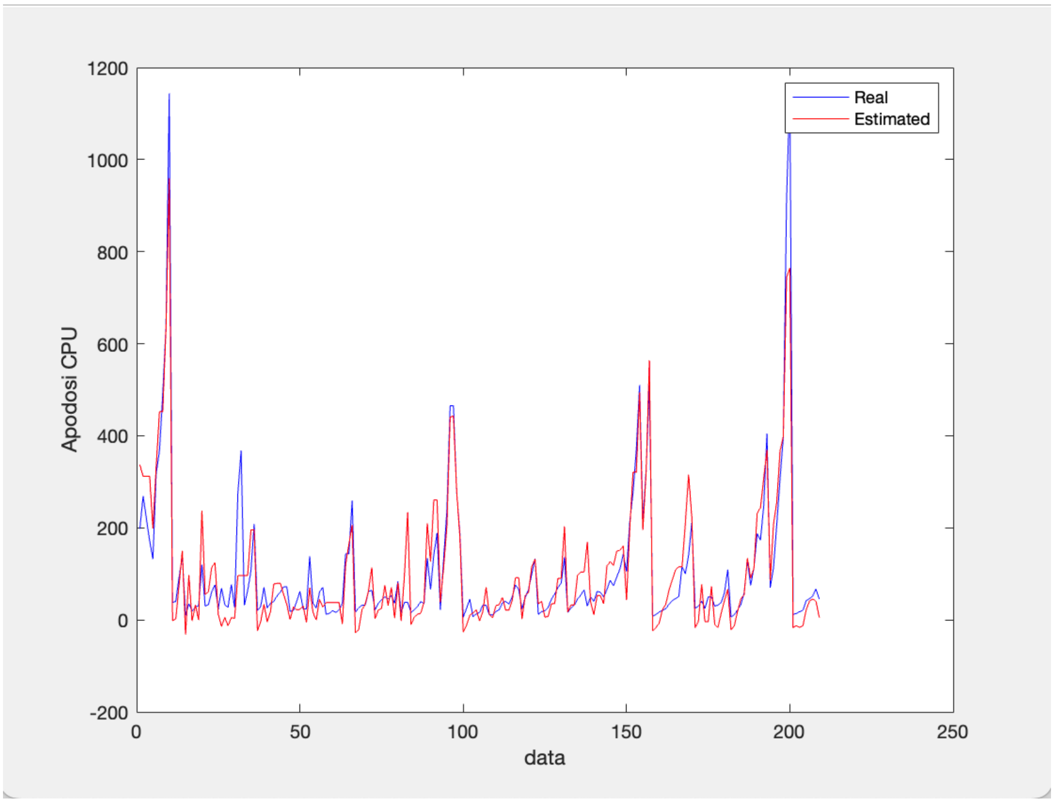
```
function [a,b,yp,mare,r2] = func3(x,y)

a = ((x'*x)^-1)*x'*y; %Υπολογισμος των συντελεστων πινακα Α.

yp = x*a;%Υπολογισμος των προβλεψεων

b = yp(:,end);%Ο σταθερος ορος b

%Υπολογισμος του συντελεστη MARE%
p = length(y);%Ο αριθμος των δεδομενων
mae = sum(y-yp,[1 p])/p;

mare = 100*mae;

%Υπολογισμος του συντελαστη R^2
yavg = sum(y,[1 p])/p;
sse = sum((y-yp).^2,[1 p]);
sst = sum((y-yavg).^2,[1 p]);

r2 = 1 - (sse/sst);
```

Εικόνα 2.6 Το πρόγραμμα Matlab τύπου function για τον υπολογισμό όλων των συντελεστών

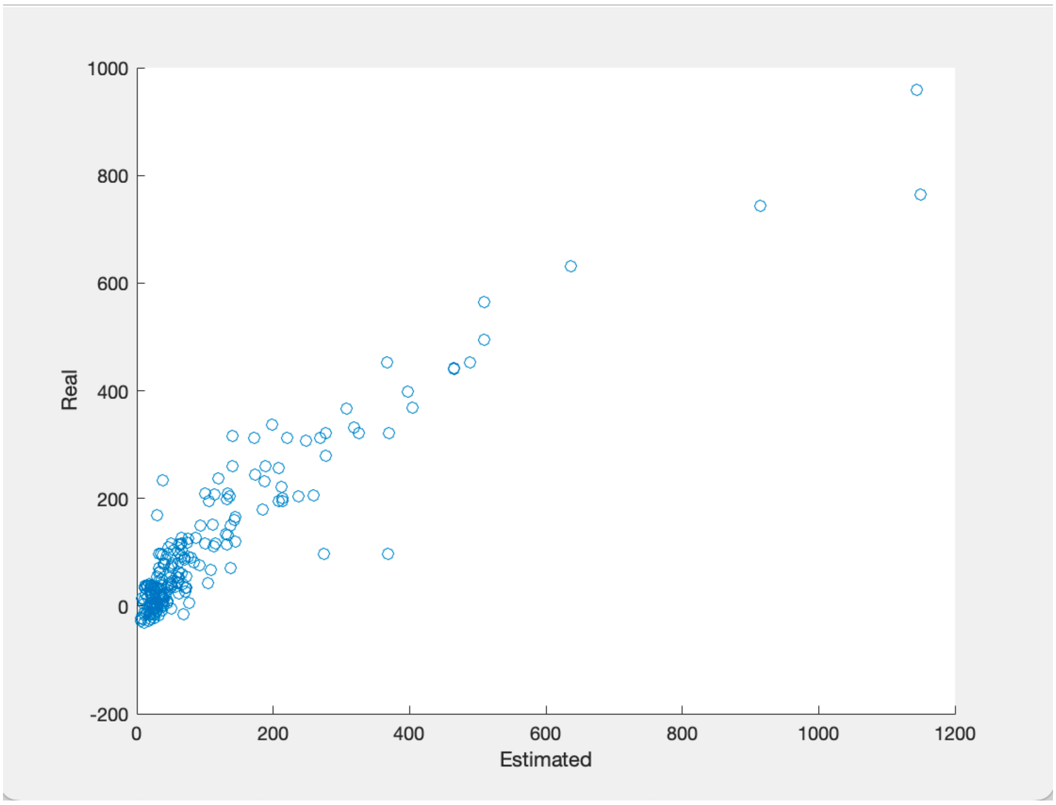
Β) Στο δεύτερο μέρος γίνεται υλοποίηση δυο γραφικών παραστάσεων:

- Γραφική παράσταση των πραγματικών τιμών y για όλα τα διαθέσιμα δεδομένα συναρτήσει του αριθμού του κάθε δεδομένου. Και πάνω στο ίδιο γράφημα η γραφική παράσταση των προβλέψεων (yp) συναρτήσει του αριθμού του κάθε δεδομένου με βάση το μοντέλο. (Εικόνα 2.7)
- Γραφική παράσταση των πραγματικών τιμών συναρτήσει των προβλέψεων (yp) του μοντέλου (Εικόνα 2.8)

Παρατηρούμαι ότι στην δεύτερη γραφική παράσταση υπάρχει κάποια συσσώρευση των αποτελεσμάτων κοντά στην αρχή των αξόνων. Αυτό συμβαίνει διότι οι περισσότεροι επεξεργαστές με βάση τα χαρακτηριστικά τους των δεδομένων είναι χαμηλής απόδοσης άρα χαμηλού y. Ιδανικά θα έπρεπε να έχουμε ευθεία στο γράφημα αλλά είναι η καλύτερη δυνατή πρόβλεψη με βάση τα δεδομένα.



Εικόνα 2.7.



Εικόνα 2.8.

Παρακάτω υπάρχει το συνολικό πρόγραμμα το οποίο δημιουργεί τον πίνακα x διαβάζοντας από το αρχείο excel τα καταλληλά δεδομένα και προσθέτοντας μια στήλη με άσσους στο τέλος του πίνακα. Στην συνέχεια γίνεται κλήση της συνάρτησης func3 και τέλος η απεικόνιση των αποτελεσμάτων με τις γραφικές παραστάσεις. Παρατηρούμαι ότι το σχετικό απόλυτο σφάλμα % είναι αρκετά κοντά στο 0 και αντίστοιχα ο συντελεστής R^2 είναι κοντά στο 1.

```matlab
clc
clear

cpudata = xlsread("Machine CPU.xlsx");%Αναγνωση των δεδομενων απο το αρχειο excel

x = cpudata(:,4:9);
assoi = ones(length(x),1);%δημιουργια πινακα-στηλη με γραμμες ισες με τις γραμμες του x
x = [x assoi];%προσθεση μιας στηλης ασσων στον πινακα x

y = cpudata(:,10);

%Υπολογισμος ολων των συντελεστων
[a,b,yp,mare,r2] = func3(x,y);

%1η Γραφικη παρασταση
plot(1:209,y,'b')
hold on
plot(1:209,yp,'r')
legend('Real','Estimated');
xlabel('data')
ylabel('Apodosi CPU')
hold off

%2η γραφικη πραγματικες-προβλεψεις
scatter(y,yp);
xlabel('Estimated');
ylabel('Real');
```

a	[0.0489;0.0153;0.0056;0.6...
assoi	209x1 double
b	209x1 double
cpudata	209x10 double
mare	1.4435e-11
r2	0.8649
x	209x7 double
y	209x1 double
yp	209x1 double

Εικόνα 2.9. Το πρόγραμμα Matlab για τα ερωτήματα Α,Β με κλήση κατάλληλων συναρτήσεων

Γ) Στο τρίτο μέρος της άσκησης γίνεται προσπάθεια υλοποίησης μιας τρισδιάστατης γραφικής παράστασης που να απεικονίζει τις προβλέψεις του μοντέλου που κατασκευάστηκε στο Α ερώτημα για την απόδοση του επεξεργαστή (άξονας Ζ) συναρτήσει της μέγιστης κεντρικής μνήμης (άξονας Χ) και της μνήμης Cache (άξονας Υ). Θα χρησιμοποιηθούν οι ακόλουθες τιμές ενός επεξεργαστή :

Χρόνος κύκλου : 200 ns
Ελάχιστη κεντρική μνήμη : 3000 KB
Μέγιστη κεντρική μνήμη : 8000 KB – 16000 KB
Μνήμη Cache : 32 KB – 128 KB
Ελάχιστος αριθμός καναλιών : 6
Μέγιστος αριθμός καναλιών : 16

Άρα θα πρέπει να δημιουργηθεί ένας αντίστοιχος πίνακας Χ με επτά στήλες με σταθερές τιμές εκτός από της μέγιστης κεντρικής μνήμης και της μνήμης Cache οι οποίες είναι μεταβαλλόμενες (η έβδομη στήλη είναι γεμάτη άσσους).

Παρακάτω υπάρχει ο κώδικας για την υλοποίηση της τρισδιάστατης γραφικής παράστασης απλά υπήρχε κάποιο πρόβλημα στις διαστάσεις των επίπεδων διότι δεν γνωρίζω τις εντολές για την δημιουργία του σωστού πίνακα x (x2) με βάση τα παραπάνω.

```
k = linspace(8000,16000,100);
c = linspace(32,128,100);
[kk,cc] = meshgrid(k,c);
x2 = [200 3000 k c 6 16 1];
z = x2*a;
mesh(kk,cc,z)
```

Εικόνα 2.10. Προσπάθεια υλοποίησης τρισδιάστατου γραφήματος

Το τελευταίο ερώτημα της άσκησης είναι με βάση το τρισδιάστατο γράφημα του συγκεκριμένου επεξεργαστή τι θα ήταν προτιμότερο, η αύξηση στην μέγιστη κεντρική μνήμη κατά 8000 KB ή την αύξηση της μνήμης Cache κατά 96 KB. Σε αυτό το ερώτημα μπορούμε να απαντήσουμε και χωρίς την γραφική παράσταση εφόσον γνωρίζουμε από το Α ερώτημα τους συντελεστές α για κάθε μορφή εισόδου x του συστήματος μπορούμε να δούμε ποια κλίση (ρυθμός αύξησης) που αντιστοιχεί στο κάθε χαρακτηριστικό είναι μεγαλύτερη άρα θα έχει μεγαλύτερη επίδραση στην τελική απόδοση του επεξεργαστή. Με βάση το πρόγραμμα της εικόνας 2.9 έχουμε τις παρακάτω τιμές για τους συντελεστές α:

A1 = 0.0488549001253196
A2 = 0.0152925719024338
A3 = 0.00557138972510780
A4 = 0.641401426998156
A5 = -0.270357548317551
A6 = 1.48247217046495
b = -55.8939336070244

Εφόσον ισχύει $\alpha_4 > \alpha_3$ τότε η αύξηση της cache έχει περισσότερη επίδραση στην απόδοση του επεξεργαστή.