

ΕΠΕΞΕΡΓΑΣΙΑ ΚΑΙ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

Τα δεδομένα που προκύπτουν από τους αισθητήρες σε πολλές περιπτώσεις είναι μετρήσεις κάποιου μεγέθους (π.χ. δυναμικού, θερμοκρασίας, πίεσης, κ.ά) κατά την διάρκεια χρονικού διαστήματος (π.χ. μεταβολή θερμοκρασιών περιβάλλοντος κατά την διάρκεια 1 μήνα ή 1 έτους). Οι μετρήσεις αυτές αποτελούν μία χρονοσειρά ή σήμα, κατά την οποία ένα μέγεθος, που προκύπτει από μία καταμέτρηση ενός αισθητήρα, μεταβάλλεται σε συνάρτηση με τον χρόνο. Για να προκύψουν χρήσιμα συμπεράσματα από τις μετρήσεις (δεδομένα), τα σήματα αυτά υποβάλλονται σε κατάλληλες επεξεργασίες: α) για την βελτίωση της ποιότητας των σημάτων (π.χ. καταστολή θορύβου από ανεπιθύμητες παρεμβολές όπως θόρυβος), και β) σε στατιστικές αναλύσεις για τον υπολογισμό παραμέτρων (π.χ. μέσες τιμές, διακυμάνσεις, ελάχιστες - μέγιστες τιμές θερμοκρασιών) και την εξαγωγή χρήσιμων στοιχείων που χαρακτηρίζουν την μεταβολή των μετρήσεων με τον χρόνο.

ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ: Αφορά την κατάλληλη μορφοποίηση των σημάτων και συνίσταται: α) στην αφαίρεση μετρήσεων από το σήμα που δεν συνιστούν μετρήσεις (NaN), λόγω κάποιας βλάβης, και την αντικατάστασή τους με τιμές μέσω όρων από προηγούμενες και επόμενες τιμές του σήματος, β) την αφαίρεση από τον πίνακα ακραίων μετρήσεων (outliers) και αντικατάστασή τους με τιμές μέσω όρων από προηγούμενα και επόμενα σημεία του πίνακα, γ) την παρεμβολή τιμών ενδιάμεσα των σημείων του σήματος, ώστε να αυξηθούν τα σημεία του σήματος, δ) την κατάλληλη επεξεργασία του σήματος, όπως εξομάλυνση του σήματος από επιπρόσθετο υψηλοχρηθό θόρυβο στο σήμα.

α) Αφαίρεση και αντικατάσταση μη-μετρήσεων (NaN) στο σήμα

Έστω σήμα που προκύπτει από μετρήσεις θερμοκρασιών περιβάλλοντος μίας ημέρας ανά 4 ώρες $X=[5, 7, 6, \text{NaN}, 10, 13, 38]$, με λάθος μέτρηση (NaN) στην 4^η χρονικά θέση του πίνακα. Σε αυτή την περίπτωση πρέπει να αφαιρεθεί και να αντικατασταθεί με τον μέσον όρο της προηγούμενης και επόμενης χρονικά τιμής του σήματος, $(6+10)/2 = 8$ και το σήμα που θα προκύψει θα είναι το: $X1=[5, 7, 6, 8, 10, 13, 38]$

Ακολουθεί σχετικό πρόγραμμα:

```
clc; clear all;
t=[0, 4, 8, 12, 16, 20, 24]; %ΩΡΕΣ ΗΜΕΡΑΣ ΠΟΥ ΓΙΝΟΝΤΑΙ ΜΕΤΡΗΣΕΙΣ ΘΕΡΜΟΚΡΑΣΙΑΣ (ΑΝΑ 4 ΩΡΕΣ)
x=[5, 7, 6, NaN, 10, 13, 38]; %ΕΣΤΩ ΟΙ ΜΕΤΡΗΣΕΙΣ ΘΕΡΜΟΚΡΑΣΙΩΝ ΑΙΣΘΗΤΗΡΑ
fprintf('\nΩΡΕΣ: '); fprintf('%4.1f ',t(1:end)); fprintf('\n');
fprintf('\nTEMP: '); fprintf('%4.1f ',x(1:end)); fprintf('\n');

%1.ΕΛΕΓΧΟΣ ΓΙΑ NaN ΣΤΟΙΧΕΙΟ ΣΤΟ x

idNaN=zeros(length(x),1); %ΑΡΧΙΚΟΠΟΙΗΣΗ ΠΙΝΑΚΑ ΠΟΥ ΘΑ ΑΠΟΘΗΚΕΥΕΙ ΤΙΣ ΘΕΣΕΙΣ NaN ΣΤΟ x - (7 ΤΙΜΕΣ)
n=0; %ΜΕΤΡΗΤΗΣ ΑΡΙΘΜΟΥ NaN
for i=1:length(x) %ΕΛΕΓΧΟΣ ΟΛΩΝ ΤΩΝ ΣΤΟΙΧΕΙΩΝ ΤΟΥ ΣΗΜΑΤΟΣ
    if ((isnan(x(i)))) %ΑΝ ΥΠΑΡΧΕΙ NaN (ΕΝΤΟΛΗ MATLAB)
        n=n+1; %ΑΥΞΗΣΗ ΜΕΤΡΗΤΗ
        idNaN(n)=i; %ΑΠΟΘΗΚΕΥΣΗ ΘΕΣΗΣ NaN - ΔΕΙΚΤΗΣ ΘΕΣΗΣ ΟΠΟΥ ΥΠΑΡΧΕΙ ΤΟ NaN
        x(i)=0; %ΜΗΔΕΝΙΣΜΟΣ ΣΤΟΙΧΕΙΟΥ NaN
    end
end
```

```

%ΕΚΤΥΠΩΣΗ ΠΙΝΑΚΑ ΟΠΟΥ ΤΑ NaN ΕΧΟΥΝ ΑΝΤΙΚΑΤΑΣΤΑΘΕΙ ΜΕ ΜΗΔΕΝΙΚΑ
fprintf('\nΠΙΝΑΚΑΣ ΜΕ 0 ΣΤΙΣ ΘΕΣΕΙΣ ΤΩΝ NaN:\n');
fprintf('TEMP: '); fprintf('%4.1f ',x(1:end)); fprintf('\n');

%2.ΓΕΜΙΣΜΑ ΤΩΝ ΘΕΣΕΩΝ ΠΟΥ ΗΤΑΝ ΜΕ NaN, ΜΕ ΤΙΜΕΣ ΜΕΣΩΝ ΟΡΩΝ ΑΠΟ ΑΛΛΕΣ ΤΙΜΕΣ ΤΟΥ ΠΙΝΑΚΑ x (ΠΡΟΗΓΟΥΜΕΝΗ ΚΑΙ ΕΠΟΜΕΝΗ ΤΙΜΗ)

x1=x; %ΔΗΜΙΟΥΡΓΙΑ ΝΕΟΥ ΠΙΝΑΚΑ ΠΟΥ ΘΑ ΑΠΟΘΗΚΕΥΕΙ ΤΟ ΔΙΟΡΘΩΜΕΝΟ ΣΗΜΑ
for i=1:n %ΓΙΑ ΟΛΑ ΤΑ NaN (n=ΜΕΤΡΗΤΗΣ ΤΩΝ NaN)
    if(idNaN(i)<length(x) && idNaN(i)>1) %ΑΝ ΘΕΣΗ NaN ΔΙΑΦΟΡΗ ΤΩΝ ΑΚΡΩΝ ΤΙΜΩΝ ΤΟΥ ΠΙΝΑΚΑ x
        % (ΟΧΙ ΠΡΩΤΗ ΟΥΤΕ ΤΕΛΕΥΤΑΙΑ) ΑΝΤΙΚΑΤΑΣΤΑΣΗ ΤΙΜΗΣ ΟΠΟΥ ΥΠΗΡΧΕ NaN ΣΤΟ x ΜΕ ΜΕΣΗ ΤΙΜΗ
        x1(idNaN(i))=(x(idNaN(i)-1)+x(idNaN(i)+1))/2; %ΠΡΟΗΓΟΥΜΕΝΟΥ ΚΑΙ ΕΠΟΜΕΝΟΥ ΑΡΙΘΜΟΥ
    elseif(idNaN(i)==1) %ΑΝ ΘΕΣΗ NaN ΣΤΗΝ ΑΡΧΗ ΤΟΥ x
        x1(idNaN(i))=x(idNaN(i)+1); %ΑΝΤΙΚΑΤΑΣΤΑΣΗ ΜΕ 2ο ΣΤΟΙΧΕΙΟ ΤΟΥ x
    else %ΑΝ ΘΕΣΗ NaN ΣΤΟ ΤΕΛΟΣ ΤΟΥ x ή elseif(idNaN(i)==length(x))
        x1(idNaN(i))=x(idNaN(i)-1); %ΑΝΤΙΚΑΤΑΣΤΑΣΗ ΜΕ ΠΡΟΤΕΛΕΥΤΑΙΟ ΣΤΟΙΧΕΙΟ ΤΟΥ x
    end
end

%ΕΚΤΥΠΩΣΗ ΠΙΝΑΚΑ ΟΠΟΥ ΤΑ NaN ΕΧΟΥΝ ΑΝΤΙΚΑΤΑΣΤΑΘΕΙ ΜΕ ΜΕΣΟΥΣ ΟΡΟΥΣ
fprintf('\nΠΙΝΑΚΑΣ ΟΠΟΥ ΤΑ NaN ΕΧΟΥΝ ΑΝΤΙΚΑΤΑΣΤΑΘΕΙ ΜΕ ΜΕΣΟΥΣ ΟΡΟΥΣ:\n');
fprintf('TEMP: '); fprintf('%4.1f, ',x1(1:end)); fprintf('\n');

```

```

Command Window

ΩΡΕΣ:  0.0  4.0  8.0 12.0 16.0 20.0 24.0

TEMP:  5.0  7.0  6.0 NaN 10.0 13.0 38.0

ΠΙΝΑΚΑΣ ΜΕ 0 ΣΤΙΣ ΘΕΣΕΙΣ ΤΩΝ NaN:
TEMP:  5.0  7.0  6.0  0.0 10.0 13.0 38.0

ΠΙΝΑΚΑΣ ΟΠΟΥ ΤΑ NaN ΕΧΟΥΝ ΑΝΤΙΚΑΤΑΣΤΑΘΕΙ ΜΕ ΜΕΣΟΥΣ ΟΡΟΥΣ:
TEMP:  5.0  7.0  6.0  8.0 10.0 13.0 38.0
fx >> |

```

β) Αφαίρεση από τον πίνακα ακραίων μετρήσεων (outliers)

Ως ακραία τιμή θεωρείται η τιμή στοιχείου του πίνακα που είναι μεγαλύτερη από 3 τυπικές αποκλίσεις των τιμών του πίνακα.

$X > 3 * \text{ΤΥΠΙΚΕΣ ΑΠΟΚΛΙΣΕΙΣ}$

$$\text{όπου τυπική απόκλιση} = SD = \sqrt{\frac{\sum_{i=1}^N (X_i - \text{mean}X)^2}{N-1}} \quad \text{ή} \quad \sqrt{\frac{\sum_{i=1}^N (T_i - \text{mean}T)^2}{N-1}}$$

Σε αυτή την περίπτωση, η τιμή του σήματος αφαιρείται και αντικαθίσταται από τον μέσο όρο των τιμών της χρονικά προηγούμενης και επόμενης τιμής του σήματος.

Έτσι για σήμα:

$T=[5, 7, 10, 15, 13, 25, 5]$ με τυπική απόκλιση έστω **7.11**, η τιμή 25 στην 6^η θέση είναι μεγαλύτερη από το τριπλάσιο της τυπικής απόκλισης και θα αντικατασταθεί από την τιμή $(13+5)/2 = 9$ και το σήμα που θα προκύψει θα είναι το ακόλουθο:

$T=[5, 7, 10, 15, 13, 9, 5]$

Ακολουθεί σχετικό πρόγραμμα:

```
clc; clear all; close all;
t=[0, 4, 8, 12, 16, 20, 24]; %ΩΡΕΣ ΗΜΕΡΑΣ ΠΟΥ ΓΙΝΟΝΤΑΙ ΜΕΤΡΗΣΕΙΣ ΘΕΡΜΟΚΡΑΣΙΑΣ (ΑΝΑ 4 ΩΡΕΣ)
T=[5, 7, 10, 15, 13, 25, 5]; %ΕΣΤΩ ΟΙ ΜΕΤΡΗΣΕΙΣ ΘΕΡΜΟΚΡΑΣΙΩΝ ΑΙΣΘΗΤΗΡΑ
fprintf('t: '); fprintf('%4.1f ',t(1:end));
fprintf('\nT: '); fprintf('%4.1f ',T(1:end));
fprintf('\n');

N=length(T); %ΜΕΓΕΘΟΣ ΠΙΝΑΚΑ ΘΕΡΜΟΚΡΑΣΙΩΝ

s=0; %ΜΗΔΕΝΙΣΜΟΣ ΑΘΡΟΙΣΜΑΤΟΣ
for i=1:N %ΓΙΑ ΟΛΕΣ ΤΙΣ ΘΕΡΜΟΚΡΑΣΙΕΣ
    s=s+T(i); %ΑΘΡΟΙΣΜΑ ΓΙΑ ΕΥΡΕΣΗ ΜΕΣΗΣ ΤΙΜΗΣ
end
meanT=s/N; %ΕΥΡΕΣΗ ΜΕΣΗΣ ΤΙΜΗΣ ΘΕΡΜΟΚΡΑΣΙΩΝ

s=0; %ΜΗΔΕΝΙΣΜΟΣ ΑΘΡΟΙΣΜΑΤΟΣ
for i=1:N %ΓΙΑ ΟΛΕΣ ΤΙΣ ΘΕΡΜΟΚΡΑΣΙΕΣ
    s=s+(T(i)-meanT)^2; %ΑΘΡΟΙΣΜΑ ΓΙΑ ΕΥΡΕΣΗ ΤΥΠΙΚΗΣ ΑΠΟΚΛΙΣΗΣ ΘΕΡΜΟΚΡΑΣΙΩΝ
end
stdT=sqrt(s/(N-1)); %ΕΥΡΕΣΗ ΤΥΠΙΚΗΣ ΑΠΟΚΛΙΣΗΣ ΘΕΡΜΟΚΡΑΣΙΩΝ (Standard Deviation)

fprintf('\nΜΕΣΗ ΤΙΜΗ: %4.2f\nΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ: %4.2f\n',meanT,stdT);
```

```

%ΑΝΤΙΚΑΤΑΣΤΑΣΗ ΤΙΜΩΝ ΣΤΙΣ ΘΕΣΕΙΣ ΤΟΥ ΠΙΝΑΚΑ ΠΟΥ ΥΠΗΡΧΑΝ ΑΚΡΑΙΕΣ ΤΙΜΕΣ
T1=T; %ΔΗΜΙΟΥΡΓΙΑ ΝΕΟΥ ΠΙΝΑΚΑ ΠΟΥ ΘΑ ΑΠΟΘΗΚΕΥΣΕΙ ΤΟ ΔΙΟΡΘΩΜΕΝΟ ΣΗΜΑ
for i=1:N %ΓΙΑ ΟΛΕΣ ΤΙΣ ΤΙΜΕΣ ΤΟΥ ΠΙΝΑΚΑ
    if (abs(T(i))>3*stdX) %ΕΥΡΕΣΗ ΑΚΡΑΙΑΣ ΤΙΜΗΣ - ΑΝ ΜΕΓΑΛΥΤΕΡΗ ΤΟΥ ΖΗΛΑΣΙΟΥ ΤΗΣ ΤΥΠΙΚΗΣ ΑΠΟΚΛΙΣΗΣ
        if(i==1)
            T1(i)=T(i+1); %ΑΝΤΙΚΑΤΑΣΤΑΣΗ ΜΕ 2ο ΣΤΟΙΧΕΙΟ ΤΟΥ Τ
        elseif(i==N)
            T1(i)=T(i-1); %ΑΝΤΙΚΑΤΑΣΤΑΣΗ ΜΕ ΠΡΟΤΕΛΕΥΤΑΙΟ ΣΤΟΙΧΕΙΟ ΤΟΥ Τ
        else
            %ΑΝ ΘΕΣΗ ΑΚΡΑΙΑΣ ΤΙΜΗΣ ΔΙΑΦΟΡΗ ΤΩΝ ΑΚΡΩΝ ΤΟΥ ΠΙΝΑΚΑ
            %ΑΝΤΙΚΑΤΑΣΤΑΣΗ ΤΙΜΗΣ ΟΠΟΥ ΥΠΗΡΧΕ ΑΚΡΑΙΑ ΤΙΜΗ ΣΤΟΝ Τ
            %ΜΕ ΜΕΣΗ ΤΙΜΗ ΠΡΟΗΓΟΥΜΕΝΟΥ ΚΑΙ ΕΠΟΜΕΝΟΥ ΑΡΙΘΜΟΥ
            T1(i)=(T(i-1)+T(i+1))/2;
        end
    end
end
end

```

```

%ΕΚΤΥΠΩΣΗ 2 ΠΙΝΑΚΩΝ (ΑΡΧΙΚΟΥ ΚΑΙ ΔΙΟΡΘΩΜΕΝΟΥ)
fprintf('\nΑΡΧΙΚΟ ΣΗΜΑ: '); fprintf('%4.2f ',T(1:end));
fprintf('\nΔΙΟΡΘΩΜΕΝΟ ΣΗΜΑ: '); fprintf('%4.2f ',T1(1:end));
fprintf('\n');

```

```

%ΣΧΕΔΙΑΣΗ ΣΗΜΑΤΩΝ ΑΡΧΙΚΟΥ ΚΑΙ ΔΙΟΡΘΩΜΕΝΟΥ
figure(1);
plot(t,T,'rs--','markersize',15);
hold on;
plot(t,T1,'b*--','linewidth',5);
grid on;
legend('ΑΡΧΙΚΟ ΣΗΜΑ','ΔΙΟΡΘΩΜΕΝΟ ΣΗΜΑ',2);
xlabel('ΩΡΕΣ');
ylabel('ΘΕΡΜΟΚΡΑΣΙΑ (°C)');
title('ΑΦΑΙΡΕΣΗ ΑΚΡΑΙΩΝ ΤΙΜΩΝ ΣΗΜΑΤΟΣ');

```

