Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Інфокомунікацій
(повна назва)
Кафедра Інфокомунікаційної інженерії імені В.В. Поповського
(повна назва)

ЗВІТ
з практичного заняття №1

з дисципліни
Прогнозування та моделювання в соціальній сфері
Тема : «Формування прогнозів на основі методу середніх точок»
Варіант №10

Виконав:
студент 2 курсу, групи КУІБ-19-2
Нестеренко Є.В.
(прізвище, ініціали)

Перевірив: завідувач кафедри ІКІ ім. В.В. Поповського
Лемешко О.В.
(посада, прізвище, ініціали)

2021 р.

МЕТА РОБОТИ

Здобуття практичних навичок з побудови прогнозів за допомогою методів крайніх та середніх точок. Оцінка точності побудови прогнозів за множиною показників. Проведення порівняльного аналізу ефективності досліджуваних методів прогнозування за якісними та кількісними критеріями.

ХІД ВИКОНАННЯ

Завдання 1. Отримання індивідуального варіанту завдань, представленого часовим рядом

Варіант завдання, представлений у вигляді часового ряду представлений .

Таблиця 1 – Індивідуальні значення для побудови прогнозу

Період	Завдання 10
	Середня заробітна плата в Україні (екв. дол.)
на 31.12.2009	239,5
на 31.12.2010	289,3
на 31.12.2011	340,7
на 31.12.2012	375,3
на 31.12.2013	393,8
на 31.12.2014	213,8
на 31.12.2015	173,4
на 31.12.2016	221,5
на 31.12.2017	275,3
на 31.12.2018	332,3
на 31.12.2019	430,5
на 31.12.2020	437,6

Завдання 2. Аналіз основних характеристик часового ряду

Математичне очікування для випадкового процесу – це невипадкова функція $M(y_t)$,яка при кожному значенні аргументу дорівнює математичному очікуванню відповідного перетину випадкової функції:

$$M(y_t) = \int_{-\infty}^{-} y_t f(y,t) dy.$$

Дисперсія для випадкової функції – це невипадкова функція $D(y_t)$, значення якої для кожного перетину дорівнюють дисперсії відповідного перетину:

$$D(y_t) = M(y_t - M(y_t))^2 = \int_{-\infty}^{\infty} y_t - M(y_t)^2 f(y,t) dt.$$

Автокореляційна функція використовується для характеристики зв'язку між двома перетинами випадкового процесу і є невипадковою функцією двох аргументів $K_y(t_1,t_2)$, що для кожної пари значень аргументів дорівнює кореляційному моменту відповідних перетинів:

$$K_y(t_1,t_2) = cov(t_1,t_2) = M((y_{t_1} - M(y_{t_1}))(y_{t_2} - M(y_{t_2}))).$$

Нормована кореляційна функція – це невипадкова функція, кожне значення якої дорівнює коефіцієнту кореляції:

$$\rho(t_1,t_2) = \frac{M((y_{t_1} - M(y_{t_1}))(y_{t_2} - M(y_{t_2})))}{\sqrt{D(y_{t_1})D(y_{t_2})}}.$$

Коефіцієнт кореляції Пірсона — показник кореляції (лінійної залежності) між двома змінними X та Y, який набуває значень від −1 до +1 включно. Значення +1 означає, що залежність між X та Y є лінійною, і всі точки функції лежать на прямій, яка відображає зростання Y при зростанні X. Значення −1 означає, що всі точки лежать на прямій, яка відображає зменшення Y при зростанні X. Якщо коефіцієнт кореляції Пірсона = 0, то саме лінійної кореляції між змінними немає. Визначається коефіцієнт Пірсона за формулою:

$$r_{xy} = \frac{\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{n}(x_t - \bar{x})^2 \sum_{t=1}^{n}(y_t - \bar{y})^2}}.$$

Завдання 3. Короткий опис досліджуваного методу прогнозування.

Метод крайніх точок – це метод, який передбачає обрання двох крайніх точок та проведення між ними прямої. Цей метод використовується під час відсутності достатньої кількості даних. Це один з найпростіших та найшвидших методів прогнозування, пріоритет у якому віддан саме швидкості. При цьому точність прогнозу (особливо на довгій дистанції) зменшується. Пряма, яка проводиться для здобуття прогнозу, має вигляд $\boldsymbol{y=b_0+b_1t}$, а крайні точки мають координати $\boldsymbol{(t_1,y_1)}$ **і** $\boldsymbol{(t_N,y_N)}$.

Оцінки параметрів обчислюються за формулами:

$$\boldsymbol{b_1} = \frac{y_N - y_1}{t_N - t_1},$$

$$\boldsymbol{b_0 = y_1 - b_1 t_1.}$$

Метод середніх точок – це метод, здобутті прогнозів яким, сукупність спостережені розділяється на дві частини у кожній з яких знаходиться середнє арифметичне, а через нього визначаються координати середніх точок, через які проводиться пряма. Як в методі крайніх точок, недолік методу середніх точок є припущення про лінійність прогнозування та наявність деякої експертної складової, завдяки котрій і будуть обрані частина ряду.

Нехай для визначеності перша точка *a* має координати *(t* 1, *y* 1), а друга - *b* - координати *(t* 2, *y* 2). Формули для визначення координат точок *a* та *b*:

$$\boldsymbol{t^1 = \frac{1}{T/2}\sum_{t=1}^{T/2} t,}$$

$$\boldsymbol{y^1 = \frac{1}{T/2}\sum_{t=1}^{T/2} y_t,}$$

$$\boldsymbol{t^2 = \frac{1}{T/2}\sum_{t=\mathrm{T}/2+1}^{T} t,}$$

$$\boldsymbol{y^2 = \frac{1}{T/2}\sum_{t=\mathrm{T}/2+1}^{T} y_{\mathrm{i}}.}$$

Рівняння прямої, яка апроксимує тренд: $\boldsymbol{y=a_0+a_1t}$ і визначається за формулами:

$$\boldsymbol{a_1 = \frac{1}{(T/2)^2}\sum_{t=\mathrm{T}/2+1}^{T}(y_t - y_{t-T/2}),}$$

$$\boldsymbol{a_1 = \frac{y^2-y^1}{t^2-t^2},}$$

$$\boldsymbol{a_0 = y^1 - a_1 t^1}.$$

Завдання 4. Побудова прогнозу з використанням досліджуваного методу. Графічна інтерпретація даних часового ряду та прогнозованих значень.

Використовуючи програму, створену у Matlab, код якої можна побачити на рисунках нижче, побудував прогноз для заданого часового ряду.

```
% модель часового ряду
N=12;% довжина часового ряду, кількість спостережень
t=1:N;
N2=N/2;
%   1      2      3      4      5      6      7      8      9      10     11     12
Y=[239.5 289.3 340.7 375.3 393.8 213.8 173.4 221.5 275.3 332.3 430.5 437.6];
%Y=5*sin(t);%
%Y=10*ones(1,N)+0.01*rand(1,N);
% Y=t;
%Y=rand(1,N).*rand(1,N).*rand(1,N);

Y_avg=mean(Y); % знаходження МО
D=cov(Y); % розрахунок дисперсії
COR=xcorr(Y);% повертає послідовність автокореляції

figure(1)
[acf, lags]=autocorr(Y);
stem(lags,acf)% будує графік функції автокореляції (корелограми)
grid on

% Коефіцієнт кореляції Пірсона
P=((t-mean(t))*(Y-mean(Y))')/sqrt(((t-mean(t))*(t-mean(t))')*((Y-mean(Y))*(Y-mean(Y))'))
% МЕТОД КРАЙНІХ ТОЧОК

% коефіцієнти рівняння прямої y=b1*t+b0
b1=(Y(N2)-Y(1))/(t(N2)-t(1))
b0=Y(1)-b1*t(1)
Y1=b0+b1*t; % реалізація моделі прогнозування
%прогноз на один крок вперед%
shag_1=Y1(N/2+1)
% розрахунок показників помилок прогнозу
```

script Ln 16 Col 33

Рисунок 1 – Фаргмент коду програми

```
Y1=b0+b1*t; % реалізація моделі прогнозування
%прогноз на один крок вперед%
shag_1=Y1(N/2+1)
% розрахунок показників помилок прогнозу
e_mkt=Y(N/2+1)-Y1(N/2+1) % помилка прогнозу
delta_mkt=abs(Y(N/2+1)-Y1(N/2+1)) % абсолютна помилка прогнозу
eps_mkt=abs(Y(N/2+1)-Y1(N/2+1))/Y(N/2+1) % відносна помилка прогнозу

% середня абсолютна помилка прогнозу
MAE_mkt=0;
for i=(N/2+1):N
    MAE_mkt=MAE_mkt+abs(Y(i)-Y1(i));
end
MAE_mkt=MAE_mkt/(N/2)

% середня абсолютна відсоткова помилка прогнозу
MAPE_mkt=0;
for i=(N/2+1):N
    MAPE_mkt=MAPE_mkt+abs(Y(i)-Y1(i))/Y(i);
end
MAPE_mkt=100*MAPE_mkt/(N/2)
% середня відсоткова помилка прогнозу
MPE_mkt=0;
for i=(N/2+1):N
    MPE_mkt=MPE_mkt+(Y(i)-Y1(i))/Y(i);
end
MPE_mkt=100*MPE_mkt/(N/2)

%коефіцієнт детермінації
R2_mkt=1-(sum((Y-Y1).^2))/(sum((Y-Y_avg).^2))

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


% МЕТОД СЕРЕДНІХ ТОЧОК

```

script Ln 9 Col 77

Рисунок 2 – Фаргмент коду програми

```
% МЕТОД СЕРЕДНІХ ТОЧОК

t1=sum(t(1:N/2))/(N/2);
t2=sum(t((N/2+1):N))/(N/2);

y1=sum(Y(1:N/2))/(N/2);
y2=sum(Y((N/2+1):N))/(N/2);

% коефіцієнти рівняння прямої y=a1*t+a0
a1=(y2-y1)/(t2-t1)
a0=y1-a1*t1
Y2=a0+a1*t; % реалізація моделі прогнозування
%прогноз на один крок вперед%
shag_1=Y2(N/2+1)
% розрахунок показників помилок прогнозу
e_mst=Y(N/2+1)-Y2(N/2+1) % помилка прогнозу
delta_mst=abs(Y(N/2+1)-Y2(N/2+1)) % абсолютна помилка прогнозу
eps_mst=abs(Y(N/2+1)-Y2(N/2+1))/Y(N/2+1) % відносна помилка прогнозу

% середня абсолютна помилка прогнозу
MAE_mst=0;
for i=(N/2+1):N
    MAE_mst=MAE_mst+abs(Y(i)-Y2(i));
end
MAE_mst=MAE_mst/(N/2)

% середня абсолютна відсоткова помилка прогнозу
MAPE_mst=0;
for i=(N/2+1):N
    MAPE_mst=MAPE_mst+abs(Y(i)-Y2(i))/Y(i);
end
MAPE_mst=100*MAPE_mst/(N/2)

% середня відсоткова помилка прогнозу
MPE_mst=0;
```

Рисунок 3 – Фаргмент коду програми

```
MPE_mst=0;
for i=(N/2+1):N
    MPE_mst=MPE_mst+(Y(i)-Y2(i))/Y(i);
end
MPE_mst=100*MPE_mst/(N/2)

%коефіцієнт детермінації
R2_mst=1-(sum((Y-Y2).^2))/(sum((Y-Y_avg).^2))


figure(2)
plot(t,Y,'-kp',t,Y1,'-.rs',t,Y2,'-b*')
grid on
legend('Y','YmKT','YmST')
xlabel('час');
ylabel('Y');
```

Рисунок 4 – Фаргмент коду програми

Також були здобуті значення показників точності для методу крайніх та середніх точок за даним часовим рядом. Нижче можна побачити коефіцієнт Пірсона на рис. 5, графік кореляції на рис. 6, значення похибок та дані часового ряду прогнозу для методу крайніх точок на рис.7 та значення похибок та дані часового ряду прогнозу для методу середніх точок на рис.8.

```
P =

    0.3330
```

Рисунок 5 – Коефіцієнт Пірсона

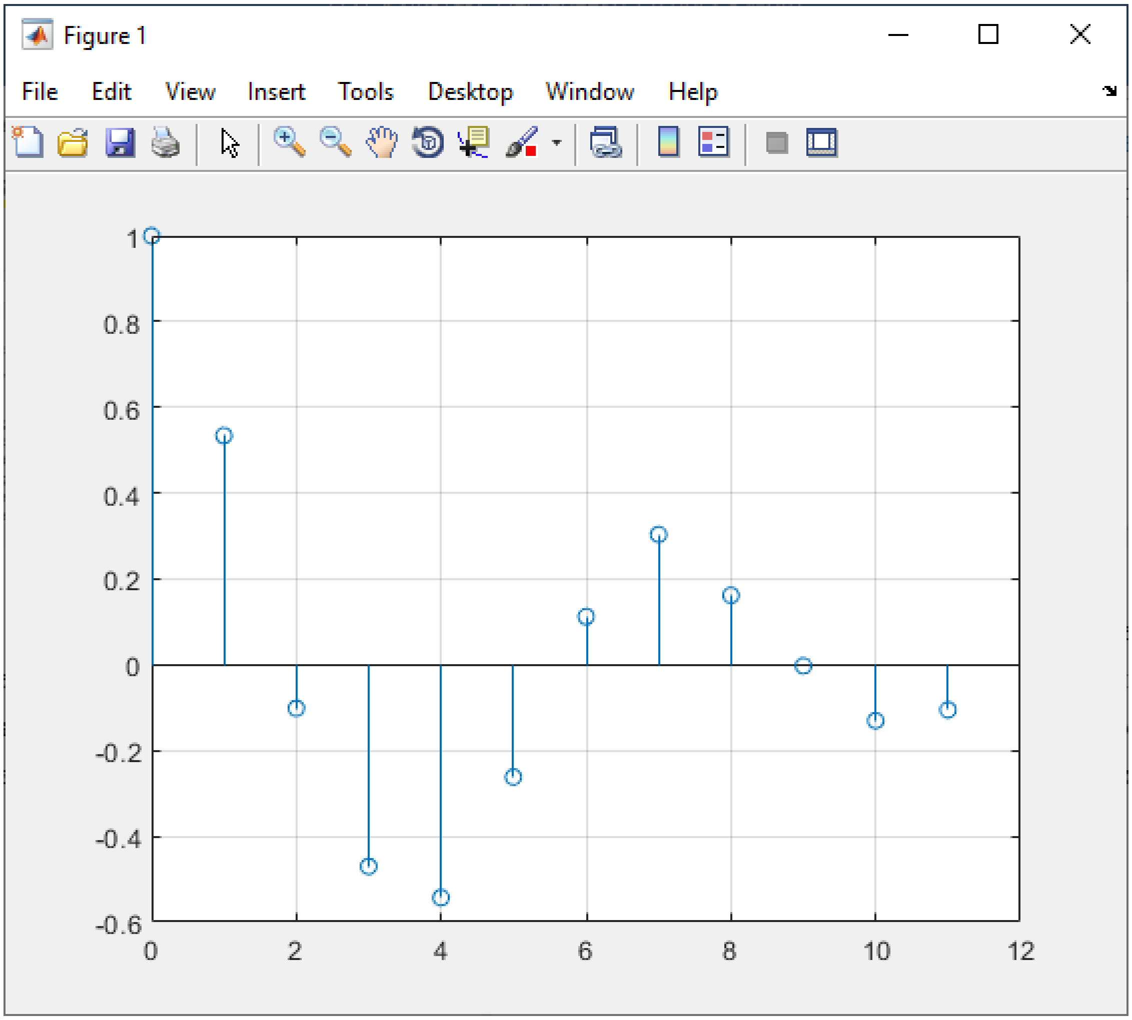

Рисунок 6 – Графік кореляції

Command Window

```
b1 =

   -5.1400


b0 =

  244.6400


shag_1 =

  208.6600


e_mkt =

  -35.2600


delta_mkt =

   35.2600


eps_mkt =

    0.2033


MAE_mkt =

  127.7100


MAPE_mkt =

   35.4561


MPE_mkt =

   28.6780


R2_mkt =

   -1.5538
```

Рисунок 7 – Значення похибок для методу крайніх точок

```
a1 =

    0.5056


a0 =

  306.9639


shag_1 =

  310.5028


e_mst =

 -137.1028


delta_mst =

  137.1028


eps_mst =

    0.7907


MAE_mst =

   87.6083


MAPE_mst =

   32.4343


MPE_mst =

  -11.7763


R2_mst =

    0.0133
```

Рисунок 8 – Значення похибок для методу середніх точок

На рис. 9 відображений побудований за наведеними вище даними графік прогнозу, заснований на методі крайніх (красна лінія) і середніх (синя лінія) точок. Метод середніх точок надав більш точний результат, аніж метод крайніх точок.

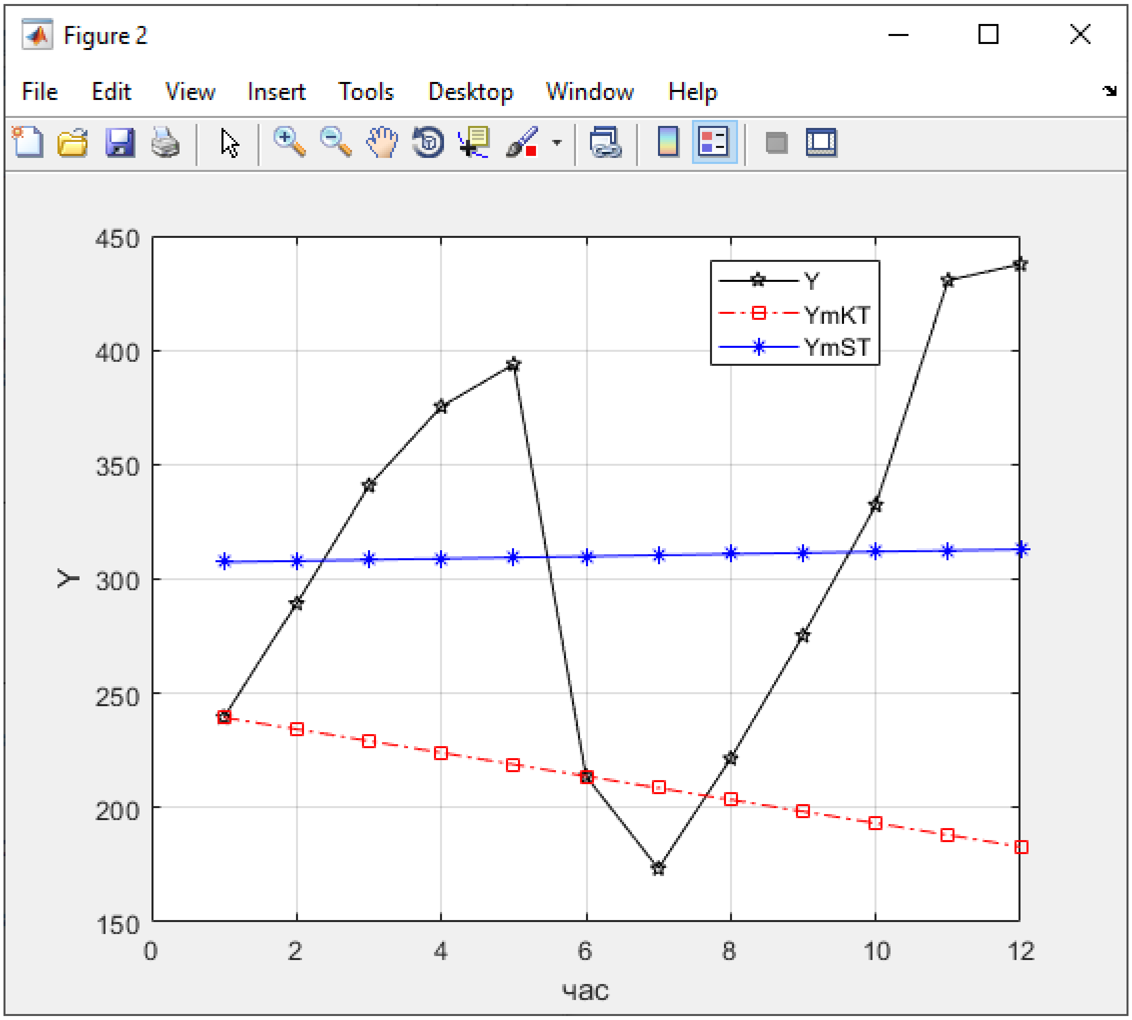

Рисунок 9 - Графік часового ряду та прогнозування за методами крайніх і середніх точок

Завдання 5. Оцінка точності побудованого прогнозу за множиною показників. Занесення отриманих результатів розрахунку в порівняльну таблицю.

Для оцінки точності побудованого прогнозу використовується формула похибки прогнозу, абсолютної похибки прогнозу, середньої абсолютної похибки прогнозу, відносної похибки прогнозу, середньої абсолютної відсоткової похибки прогнозу, середньої відсоткової похибки прогнозу та коефіцієнту детермінації:

$$e_j = y_j - \hat{y}_j,$$

$$\Delta_j = \left| y_j - \hat{y}_j \right|,$$

$$MAE = \frac{\sum_{j=1}^{N} \left| y_j - \hat{y}_j \right|}{N},$$

$$\varepsilon_j = \frac{\left| y_j - \hat{y}_j \right|}{y_j} 100\,\%,$$

$$MAPE = \frac{1}{N} \sum_{j=1}^{N} \frac{\left| y_j - \hat{y}_j \right|}{y_j} 100\,\%,$$

$$MPE = \frac{1}{N} \sum_{j=1}^{N} \frac{y_j - \hat{y}_j}{y_j} 100\,\%,$$

$$R^2 = 1 - \frac{\sum_{t=1}^{n} e_t^2}{\sum_{t=1}^{n} (y_t - \overline{y})^2}.$$

Отримані результати точності занесені у таблицю 2.

Таблиця 2 – Отримані у результаті розрахунків дані

Метод прогнозу /показник точності прогнозу	Прогноз (на один часовий інтервал вперед)	Помилка прогнозу	Абсол. помилка прогнозу	Відн. помилка прогнозу	Сер. абс. помилка прогнозу	Сер. абс. відсоткова помилка прогнозу	Сер. відсотк. помилка прогнозу	Коеф. детерм.
Метод крайніх точок	208,6600	-35,26	35,26	0,2033	127,71	35,4561	28,678	0,22
Метод середніх точок	310,5028	-137,1028	137,1028	0,7907	87,6083	32,4343	-11,7763	0,0133

ВИСНОВКИ

Середня абсолютна відсоткова помилка прогнозу методом крайніх точок дорівнює 35.45%, що знаходиться у проміжку між 20 та 50 відсотками і є задовільним результатом. Помилка прогнозу на 1 крок вперед склала 35,26.

Для метода середніх точок середня абсолютна відсоткова помилка прогнозу дорівнює 32.43. Це показує, що прогноз знаходиться у проміжку між 20 та 50 відсотками і є задовільним результатом. Помилка прогнозу на 1 крок вперед склала 137,1028.

Метод крайніх точок дав значно точніший результат прогнозу на 1 крок, якщо порівнювати з методом середніх точок, але середня абсолютна точність прогнозу краща у метода середніх точок.

Отже, можна зробити висновок, що метод крайніх точок підходить більше для експрес-прогнозу, а метод середніх точок – для подальшого прогнозу.