Путин Павел Александрович, группа 7-1

Лабораторная работа № 6

**Вариант № 10**

Линейная регрессия

**Цель работы**

Синтезировать заданный алгоритм линейной регрессии. Выполнить проверку значимости полученной модели регрессии.

**Задание**

Получить у преподавателя вариант задания и написать код, реализующий алгоритм линейной регрессии. Получить коэффициенты модели регрессии. Выполнить проверку значимости полученной модели и представить результаты в виде выводов по проделанной работе.

Используя стандартные функции Matlab (regress) построить модель линейной регрессии гармонического ряда с коэффициентами a(1)=3; a(2)=6; a(3)=7; a(4)=-5; a(5)=4; a(6)=-1. Вычислить коэффициент детерминации для следующих объемов обучающей выборки: 50, 100 и 1000. Дисперсия ошибки измерения выходной переменной = 0.1, уровень значимости для проверки гипотез по критерию Фишера = 0.05.

**Код программы (внесённые изменения в шаблон кода выделены)**

%Файл pr35\_1\_lin\_regr. Использование стандартных функций MatLab для построения

%линейной регрессии

clear all; close all;

%% 1. Задание исходных данных

var1=2;%вид восстанавливаемой функции (1-полином; 2-гармонический ряд)%var1

ng=5;% порядок полинома (ряда), ng+1 - количество коэффициентов полинома (ряда)

N=100;%объем обучающей выборки

n=1;%размерность вектора входных переменных (в данном примере фиксирована)

D=0.1; %дисперсия ошибки измерения выходной переменной

gamma=0.05;%уровень значимости для проверки гипотез по критерию Фишера

xmin=0; xmax=1; dx=0.01;%границы и дискретность области определения функций

xd=xmin:dx:xmax; ld=length(xd);

%% 2. Генерация обучающей выборки данных

XN=xmin+(xmax-xmin)\*rand(N,n);

YN=zeros(N,1); a=zeros(ng+1,1);

if var1==1

%исходные данные для полинома

a(1)=1; a(2)=5;

a(3)=10; a(4)=0.5\*10^3; a(5)=0.05\*10^4; a(6)=-0.01\*10^5;

end

if var1==2

%исходные данные для гармонического ряда

% a(1)=0.5; a(2)=1;

% a(3)=10; a(4)=-5; a(5)=0.5; a(6)=-1;

a(1)=3; a(2)=6; a(3)=7; a(4)=-5; a(5)=4; a(6)=-1;

end

%Задание аппроксимируемой функции

if var1==1

p=fliplr(a'); %в обратном порядке

YN=polyval(p,XN)+sqrt(D)\*randn(N,1);

end

if var1==2

YN=a(1)\*ones(N,1)+a(2)\*sin(pi\*XN)+a(3)\*sin(2\*pi\*XN)+a(4)\*sin(3\*pi\*XN)...

+a(5)\*sin(4\*pi\*XN)+a(6)\*sin(5\*pi\*XN)+sqrt(D)\*randn(N,1);

end

%% 3. Обращение к функции вычисления коэффициентов регрессии

X=[ones(N,1),XN];

[a\_,aint,r,rint,stat]=regress(YN,X,gamma);

%a\_-вектор коэффициентов регрессии (n+1)x1;

%bint-матрица интервальных оценок коэффициентов (n+1)x2;%r-вектор остатков Nx1;

%rint-матрица доверительных интервалов остатков Nx2;

%stat- вектор, содержащий значения: статистик R^2 и F, p-значение, RSS/(N-n-1)

RSS=sum(r.^2); mY=mean(X\*a\_); ESS=sum((X\*a\_-mY).^2);

R2=ESS/(RSS+ESS);%коэффициент детерминации

F=(ESS/n)/(RSS/(N-n-1));%статистика Фишера

fgamma=finv(1-gamma,n,N-n-1);%значение F-статистики, превышаемое с вероятностью gamma

p\_value=1-fcdf(F,n,N-n-1);%расчет вероятности превышения полученного значения F

%Проверка вычисляемых статистик в среде

disp('Вычисляемые статистики: R^2, F, p\_value, RSS/(N-n-1)');

disp([R2,F,p\_value,RSS/(N-n-1)]);disp(stat);

%% 4. Построение регрессии и границ восстанавливаемой функции

x=[ones(ld,1),xd']; y=x\*a\_;

ymi=x\*aint(:,1); yma=x\*aint(:,2);

ymin=min(ymi); ymax=max(yma);

%% 5. Визуализация результатов

figure(1);set(gca,'FontSize',12);grid on; hold on;

axis([xmin xmax ymin-0.1 ymax+0.1]);%установка границ поля графика по осям

pg=plot(XN,YN,'ro',xd,y,'-b',xd,ymi,'--k',xd,yma,'--k');

set(pg,'LineWidth',1.25);

title('Полученная регрессионная зависимость','FontName','Courier');

xlabel('X','FontName','Courier');ylabel('Y','FontName','Courier');

str1='N='; str2=num2str(N);

str3=' ng='; str4=num2str(ng);

str5=' D='; str6=num2str(D);

str7=' gamma='; str8=num2str(gamma);

str9=' p-val='; str10=num2str(p\_value);

text(xmin+0.1, 0.5\*ymax,[str1, str2, str3, str4, str5, str6,str7,str8,str9,str10], 'HorizontalAlignment',...

'left','BackgroundColor',[.8 .8 .8],'FontSize',12);

legend('XN-YN','y=f(x)','y-dy','y+dy');hold off;

**Результаты выполнения задания**

Таблица 1 - Коэффициенты детерминации при разных объёмах выборки

|  |  |  |
| --- | --- | --- |
| **Объём обучающей выборки** | **Коэффициент детерминации** | **Значимость модели** |
| 50 | 0,5113 | 50,2263 |
| 100 | 0,5053 | 100,0915 |
| 1000 | 0,4233 | 732,6471 |

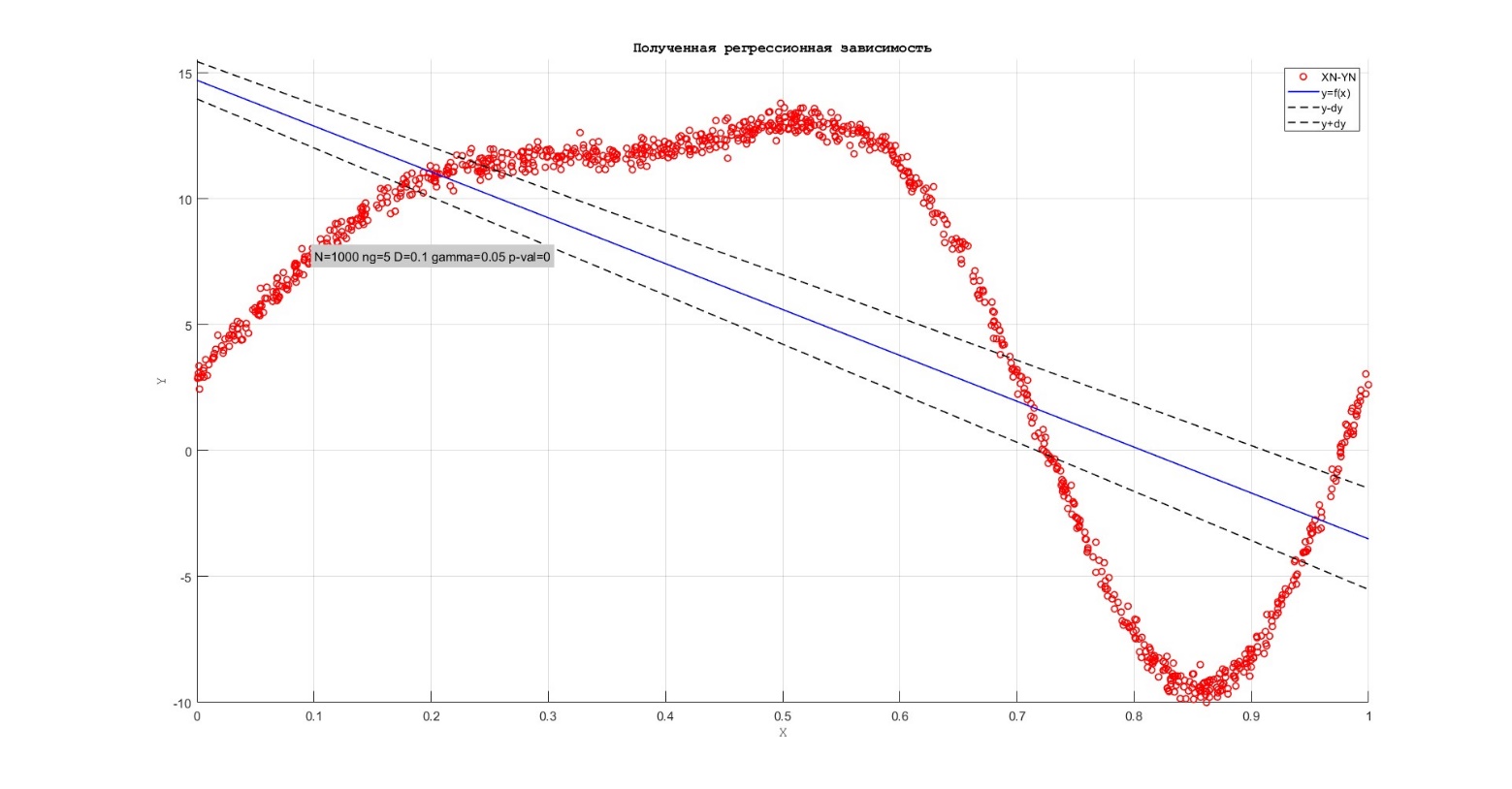
****

Рисунок 1 - Полученная регрессионая зависимость

# Выводы

1. Значимость модели регрессии (статистика Фишера) прямо пропорциональна объёму обучающей выборки
2. Регуляризация требуется в случае сильно зашумлённых данных или при наличии большого количества сильно связанных друг с другом переменных.