BLG527E, HW1 Solutions,
ITU, October 31, 2014
StudentName:_____
StudentID:_______

Q1)
Make sure that you read Appendix A of the textbook and the resources on matrices, linear algebra and probability and statistics on ninova.

Given the table of joint probabilities between two discrete random variables X and Y evaluate:
Q1a) P(X=1|Y=-1)
Q1b) Are X and Y independent random variables, why or why not?

	X=-1	X=0	X=1
Y=-1	0.2	0.1	0.05
Y=0	0.05	0.1	0.12
Y=1	0.1	0.15	0.13

Q1c) What is the expected value of 5*X + 4*Y*Y?

Q1ANSWER)

$$P\,(X = x \mid Y = y) = \frac{P\{X = x, Y = y\}}{P\{Y = y\}} = \frac{P(x,y)}{P_Y(y)}$$

where:

$$P(Y = y) = \sum_j P(x_i, y)$$

$$\boldsymbol{P_Y(-1)} = \sum_{x \in S_x} P(x, -1) = P(-1, -1) + P(0, -1) + P(1, -1)$$
$$= 0.2 + 0.1 + 0.05$$
$$= 0.35$$

$$\boldsymbol{P_Y(0)} = \sum_{x \in S_x} P(x, 0) = P(-1, 0) + P(0, 0) + P(1, 0)$$
$$= 0.05 + 0.1 + 0.12$$
$$= 0.27$$

$$\boldsymbol{P_Y(1)} = \sum_{x \in S_x} P(x, 1) = P(-1, 1) + P(0, 1) + P(1, 1)$$
$$= 0.1 + 0.15 + 0.13$$
$$= 0.38$$

$$\boldsymbol{P_Y(y)} = \begin{Bmatrix} 0.35 & if\ y = -1, \\ 0.27 & if\ y = 0, \\ 0.38 & if\ y = 1, \\ 0 & otherwise \end{Bmatrix}$$

Submitting the probabilities in:

$$P\ (X = 1 \mid Y = -1) = \frac{P\{X = 1, Y = -1\}}{P\{Y = -1\}} = \frac{P(1,-1)}{P_Y(-1)}$$

$$= \frac{0.05}{0.35}$$

= 0.1428

Q1b)

Two random variables, x and y, are independent if and only if :

$$\boldsymbol{P(x,y) = P_X(x)P_Y(y)}$$

$$P(X = x) = \sum_i P(x, y_i)$$

$$P(Y = y) = \sum_j P(x_j, y)$$

	X=-1	X=0	X=1	P(Y)
Y=-1	0.2	0.1	0.05	0.35
Y=0	0.05	0.1	0.12	0.27
Y=1	0.1	0.15	0.13	0.38
P(X)	0.35	0.35	0.3	

Multiplying these marginal probabilities $\boldsymbol{P_X(x)}$ $\boldsymbol{P_Y(y)}$, we get:

$\boldsymbol{P_X(x)*P_Y(y)}$	X=-1	X=0	X=1
Y=-1	0.1225	0.1225	0.105
Y=0	0.0945	0.0945	0.081
Y=1	0.133	0.133	0.114

Since (some of the, actually all the) entries in this table differ from the entries in the joint probability table (P(x,y)) these variables are not independent.

Q1c) What is the expected value of 5*X + 4*Y*Y?

	X=-1	X=0	X=1	P(Y)
Y=-1	0.2	0.1	0.05	0.35
Y=0	0.05	0.1	0.12	0.27
Y=1	0.1	0.15	0.13	0.38
P(X)	0.35	0.35	0.3	

E[5*X + 4*Y*Y] = 5E[X]+4E[Y*Y]= 5E[X]+4E[Y^2]=

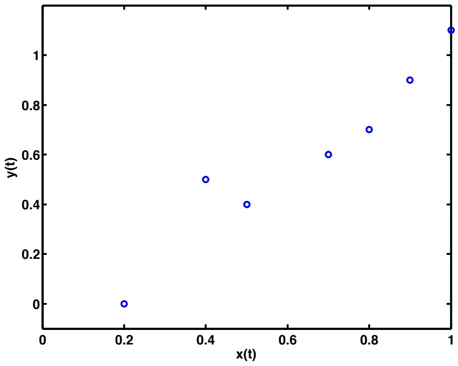
E[X]= $\sum_x xP(x)$=0.35*(-1)+0.35*0+0.3*1=-0.05

E[Y^2]= $\sum_y y^2 P(y)$=$(-1)^2$*0.35+$(0)^2$*0.27+$(1)^2$*0.38=0.73

E[5*X + 4*Y*Y]=5*(-0.05)+4*0.73=2.67

Q2) Do this exercise in matlab/python/C/java. Submit your code in ninova with instructions on how to run it.
Q2a) Given the following data points, compute least squares regression line that passes through them.
Q2b) What are your predictions for x=-5, x=4?
Q2c) Compute the least squares model for a polynomial of degree 4 and compare the variances of the linear and degree 4 polynomial model by leaving one data point out at a time.

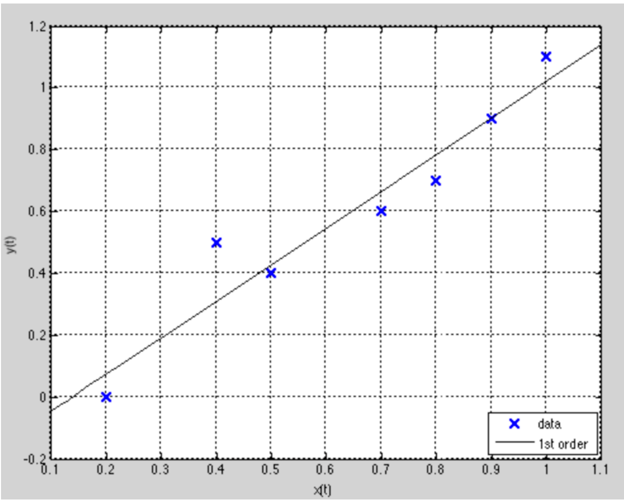
t	**1**	**2**	**3**	**4**	**5**	**6**	**7**
x(t)	**0.2**	**0.5**	**0.4**	**0.7**	**0.8**	**0.9**	**1.0**
y(t)	**0.0**	**0.4**	**0.5**	**0.6**	**0.7**	**0.9**	**1.1**



Q2ANSWER)

The source code for the solution in matlab is given in the appendix. Type "return" to continue when keyboard command is called during execution.

Q2a) y=1.1868*x -0.1629

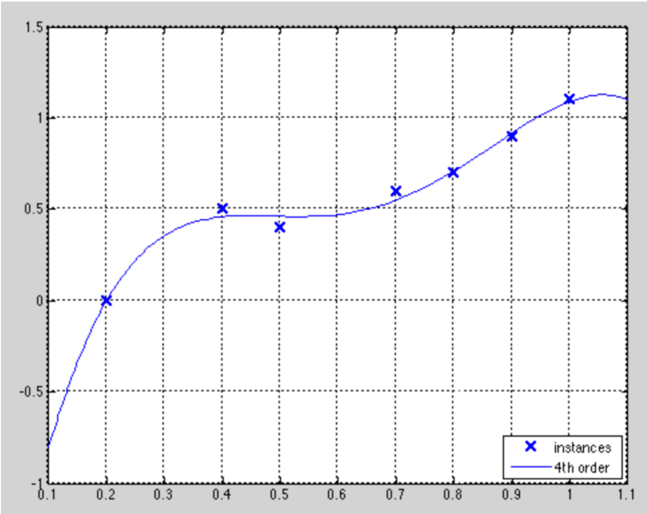


Q2b) Predictions:

x=-5 → y=-6.09683908045977

x=4 → y=4.58419540229885

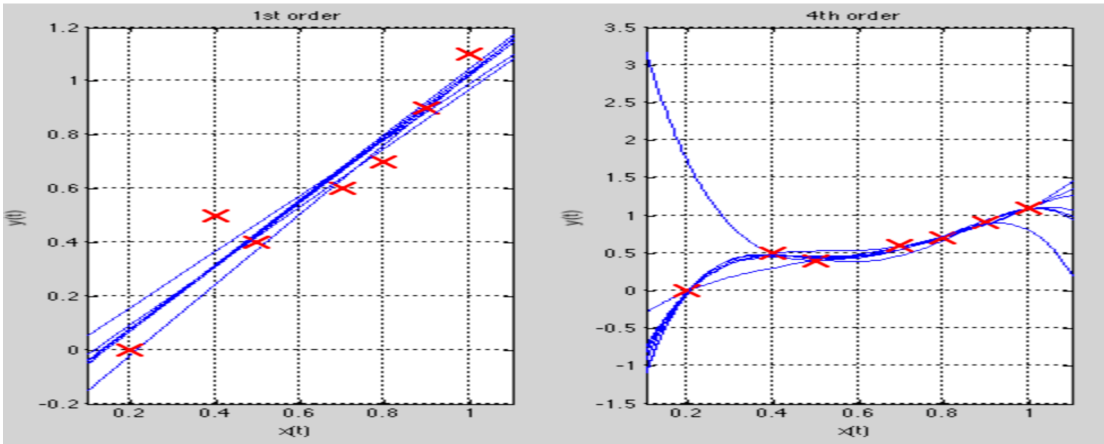
Q2c)

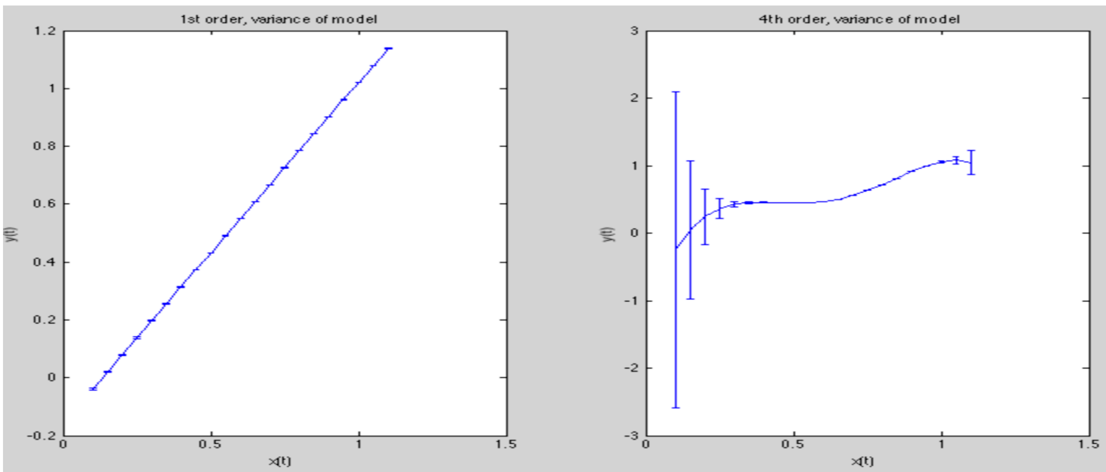


4^{th} order polynomial:

$y= -2.3062+ 19.4848*x-49.4561*x^2+52.8726*x^3-19.5035*x^4$

Models produced (top) and their means and variances (bottom).





Another Answer for Q2c)

2c) least squares model for a polynomial of degree 4:

$$g(x^t|w_4,w_3,w_2,w_1,w_0) = w_4(x^t)^4 + w_3(x^t)^3 + w_2(x^t)^2 + w_1x^t + w_0$$

$$D = \begin{bmatrix} 1 & x^1 & (x^1)^2 & (x^1)^3 & (x^1)^4 \\ 1 & x^2 & (x^2)^2 & (x^2)^3 & (x^2)^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x^N & (x^N)^2 & (x^N)^3 & (x^N)^4 \end{bmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \; r = \begin{bmatrix} r^1 \\ r^2 \\ \vdots \\ r^N \end{bmatrix}$$

$$w = (D^T D)^{-1} D^T r$$

$$g(x) = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + +w_4 x^4$$

For set of $[x, y] = [(0.2\,,0), (0.5\,,0.4), (0.4\,,0.5), (0.7\,,0.6), (0.8\,,0.7), (0.9\,,0.9), (1\,,1.1)]$

$$w^T = [-2.3062 \quad 19.4848 \quad -49.4561 \quad 52.8726 \quad -19.5035]$$

For variance calculation of the linear and degree 4 polynomial model by leaving one data point out at a time:

7 different g(x) functions are found for linear model by leaving one data point out at a time:

$$g_i = \begin{bmatrix} -0.0478 & 1.0435 \\ -0.1507 & 1.1761 \\ -0.2794 & 1.3113 \\ -0.1574 & 1.1959 \\ -0.1694 & 1.2206 \\ -0.1643 & 1.1905 \\ -0.1225 & 1.0957 \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}$$

Means of linear functions for 7 different samples (x):

$$mean\_g_i = \begin{bmatrix} 0.0793 \\ 0.4322 \\ 0.3146 \\ 0.6674 \\ 0.7851 \\ 0.9027 \\ 1.0203 \end{bmatrix}$$

$$variance(g\_linears) = \frac{1}{NM} \sum_t \sum_i [g_i(x^t) - \bar{g}(x^t)]^2 = 8.2352 \times 10^{-4}$$

7 different g(x) functions are found for degree 4 polynomial model by leaving one data point out at a time:

$$g_i = \begin{bmatrix} 5.5650 & -29.4700 & 60.3417 & -52.8333 & 17.5000 \\ -2.0354 & 16.4173 & -38.2386 & 37.7216 & -12.7630 \\ -0.6983 & 4.9921 & -8.9912 & 7.8301 & -2.0300 \\ -3.2046 & 27.9181 & -74.6048 & 82.1061 & -31.1230 \\ -2.3210 & 19.6588 & -50.1442 & 53.9418 & -20.0458 \\ -2.6594 & 22.8787 & -60.0049 & 65.8824 & -25.0000 \\ -3.1999 & 28.2682 & -77.7729 & 89.6603 & -36.1498 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \\ x^4 \end{bmatrix}$$

Means of degree 4 polynomials for 7 different samples (x):

$$mean\_g_i = \begin{bmatrix} 0.2431 \\ 0.4446 \\ 0.4564 \\ 0.5568 \\ 0.7173 \\ 0.9089 \\ 1.0561 \end{bmatrix}$$

$$variance(g\_poly4) = \frac{1}{NM} \sum_t \sum_i [g_i(x^t) - \bar{g}(x^t)]^2 = 0.0525$$

Variance measures how much g(x) fluctuate around the expected value. As the order of the polynomial increases, small changes in the dataset cause a greater change in the fitted polynomials; thus variance increases. If variance is high, it can be an overfitting situation.

Q3) If **x** and **w** are d dimensional vectors and y is a real number, what is the derivative of E(**w**) with respect to **w**?

g(**x**,**w**) = $x_1 * w_1 + x_2 * w_2 + \ldots + x_d * w_d$ (Model)
X = {($\mathbf{x}^1, y^1$), …, ($\mathbf{x}^N, y^N$)} (Training Data)

$$E(\mathbf{w}) = \frac{1}{2N} \sum_{t=1}^{N} (g(\mathbf{x}^t, \mathbf{w}) - y^t)^2 \qquad \frac{dE}{d\mathbf{w}} = ?$$

Q3) ANSWER

$$\frac{d[g(x^t, w) - y^t]}{dw} = \frac{d[g(x^t, w)]}{dw} = \left(\frac{d[g(x^t, w)]}{dw_1}, \frac{d[g(x^t, w)]}{dw_2}, \ldots, \frac{d[g(x^t, w)]}{dw_d} \right)$$
$$= \left(\frac{d[x_1 * w_1 + \ldots + x_d * w_d]}{dw_1}, \ldots, \frac{d[x_1 * w_1 + \ldots + x_d * w_d]}{dw_d} \right)$$
$$= (x_1, x_2, x_3, \ldots, x_d)$$
$$= x^t$$

$$\frac{dE(w)}{dw} = \frac{d[\frac{1}{2N} \sum_{t=1}^{N} (g(x^t, w) - y^t)^2]}{dw}$$
$$= \frac{1}{2N} \sum_{t=1}^{N} 2 * (g(x^t, w) - y^t) * \left(\frac{d[g(x^t, w) - y^t]}{dw} \right)$$
$$= \frac{1}{N} \sum_{t=1}^{N} (g(x^t, w) - y^t) * x^t$$

Q4) Given a random sample X = {x^1 , ..., x^N} where each x^i is a nonnegative real number and are i.i.d. distributed according to the probability density function (pdf):

$$f_X(x) = k\frac{x}{a^2}e^{-\frac{x^2}{a^2}}$$

Q4a) What is the value of k to make $f_X(x)$ a valid pdf (a is a constant).

Q4b) Write the likelihood and log-likelihood functions for X in terms of the parameter a

Q4c) Find the maximum likelihood estimate for the parameter a

Q4ANSWER)

Q4a) If $f_X(x)$ is a valid pdf, then

$$\int_0^\infty f_X(x)\,dx = 1$$

$$\int_0^\infty k\frac{x}{a^2}e^{-\frac{x^2}{a^2}}\,dx = 1$$

$$-\frac{x^2}{a^2} = u$$

$$\int_0^\infty \left(-\frac{k}{2}\right)\left(-\frac{2x}{a^2}\right)e^{-\frac{x^2}{a^2}}\,dx = 1$$

$$-\frac{k}{2}\int_0^\infty e^u\,du = {}_0^\infty\Big|\left(-\frac{k}{2}\right)e^{-\frac{x^2}{a^2}} = \frac{k}{2} = 1 \Rightarrow k = 2$$

Q4b) Likelihood function

$$l(a|X) = p(X|a) = \prod_{i=1}^{N} p(x^i|a) = \prod_{i=1}^{N}\frac{2x^i}{a^2}e^{-\frac{(x^i)^2}{a^2}}$$

$$= \left(\frac{2^N}{a^{2N}}\right)\left(\prod_{i=1}^{N}x^i\right)\left(e^{-\sum_{i=1}^{N}\frac{(x^i)^2}{a^2}}\right)$$

Log-likelihood function

$$\mathcal{L}(a|X) = \log l(a|X) = \log\left(\left(\frac{2^N}{a^{2N}}\right)\left(\prod_{i=1}^{N}x^i\right)\left(e^{-\sum_{i=1}^{N}\frac{(x^i)^2}{a^2}}\right)\right)$$

$$= \log(2^N) - \log(a^{2N}) + \sum_{i=1}^{N}\log(x^i) - \sum_{i=1}^{N}\frac{(x^i)^2}{a^2}$$

Q4c) Maximum likelihood estimation

$$\frac{d\mathcal{L}(a|X)}{da} = 0$$

$$-\frac{2Na^{2N-1}}{a^{2N}} + \sum_{i=1}^{N}\frac{2(x^i)^2}{a^3} = 0$$

$$-2Na^{-1} + \frac{2}{a^3}\sum_{i=1}^{N}(x^i)^2 = 0$$

$$\sqrt{\frac{\sum_{i=1}^{N}(x^i)^2}{N}} = \hat{a}$$

Q5) Given that çinekop and sarıkanat lengths (two different types of bluefish) are distributed according to Gaussians with means 11 and 18 respectively and both Gaussian have a variance of 3 and assuming that both fish have the same prior probability 0.5 of being caught, what is the best (maximizing the posterior probability of class given length) value of the threshold length that discriminates these two classes?

Q5ANSWER)

Prior probability of çinekop and sarıkanat lengths are Gaussian,

$$p(x|C_i) = \frac{1}{\sqrt{2\pi}\,\sigma_i} exp\left[-\frac{(x-\mu)^2}{2\sigma_i^2}\right]$$

And the discriminant function

$$g_i(x) = \log p(x|C_i) + \log p(C_i)$$

$$g_i(x) = -\frac{1}{2}\log 2\pi - \log s_i - \frac{(x-m_i)^2}{2s_i^2} + \log \hat{P}(C_i)$$

$-\frac{1}{2}\log 2\pi$ is a constant, the $\log \hat{P}(C_i)$ are equal and can be dropped they are common in all $g_i(x)$. Also the variance of çinekop and sarıkanat lengths are equal,

$$g_i(x) = -(x-m_i)^2$$

So, with two classes, the midpoint between two means is the threshold of decision.

$$g_{çinekop}(x) = g_{sarıkanat}(x)$$

$$\left(x - m_{çinekop}\right)^2 = (x - m_{sarıkanat})^2$$

$$x = \frac{m_{çinekop} + m_{sarıkanat}}{2} = \frac{11+18}{2} = 14.5$$

APPENDIX A

```
%HW1, Q2, by one of the students in the class
%Modified by Zehra Cataltepe
% The code is written in MATLAB R2009b
close all; clear all;
x = [0.2; 0.5; 0.4; 0.7; 0.8; 0.9; 1];
y = [0 ; 0.4; 0.5; 0.6; 0.7; 0.9; 1.1];
plot(x,y,'x','LineWidth',2,'MarkerSize',10);
% Q2.a)computation of least squares regression
A = [size(x,1) sum(x);sum(x) sum(x.^2)];
Y = [sum(y); sum(y.*x)];
w1 = pinv(A)*Y;
f = @(x) w1(1)+w1(2).*x;
hold on; grid on; fplot(f,[0.1 1.1],'k');
xlabel('x(t)');ylabel('y(t)');
legend('data','1st order','Location','SouthEast');
% Q2.b)predictions for x=-5 and x=4
x_new = [-5; 4];
y_new = [ones(size(x_new,1),1) x_new]*w1; %y_new = [-6.0968; 4.5842]

keyboard

%Q2c)
% The code is written in MATLAB R2009b
close all; clear all;
x = [0.2; 0.5; 0.4; 0.7; 0.8; 0.9; 1];
y = [0 ; 0.4; 0.5; 0.6; 0.7; 0.9; 1.1];
plot(x,y,'x','LineWidth',2,'MarkerSize',10);
% Q2.c)computation of least squares regression for a polynomial of 4th order
X = [ones(size(x,1),1) x x.^2 x.^3 x.^4];
w4 = pinv(X'*X)*X'*y;
f = @(a) w4(1)+w4(2).*a+w4(3).*a.^2+w4(4).*a.^3+w4(5).*a.^4;
hold on; grid on; fplot(f,[0.1 1.1]);
legend('instances','4th order','Location','SouthEast');

keyboard

hold on; grid on; fplot(f,[0.1 1.1]);
    xlabel('x(t)');ylabel('y(t)');
    title('1st order');
    subplot(122), plot(xdata,ydata,'rx','LineWidth',2,'MarkerSize',15);
    yy1(i,:) = f(xx) ;
    X = [ones(size(x,1),1) x x.^2 x.^3 x.^4];
    w4 = pinv(X'*X)*X'*y;
    f = @(a) w4(1)+w4(2).*a+w4(3).*a.^2+w4(4).*a.^3+w4(5).*a.^4;
    hold on; grid on; fplot(f,[0.1 1.1]);
    xlabel('x(t)');ylabel('y(t)');
    title('4th order');
    yy4(i,:) = f(xx) ;
end
```

```matlab
%HW1, Q2, by one of the students in the class
%Modified by Zehra Cataltepe
% The code is written in MATLAB R2009b

%Q2c) Variances
close all; clear all;
xdata = [0.2; 0.5; 0.4; 0.7; 0.8; 0.9; 1];
ydata = [0 ; 0.4; 0.5; 0.6; 0.7; 0.9; 1.1];

xx = 0.1:0.05:1.1 ;
yy1 = zeros(size(xdata,1),size(xx,2)) ;
yy4 = zeros(size(xdata,1),size(xx,2)) ;
for i = 1:size(xdata,1)
    if(i==1)
        x = xdata(i+1:size(xdata,1),:);
        y = ydata(i+1:size(ydata,1),:);
    elseif(i==7)
        x = xdata(1:size(xdata,1)-1,:);
        y = ydata(1:size(ydata,1)-1,:);
    else
        x = [xdata(1:i-1,:);xdata(i+1:size(xdata,1),:)];
        y = [ydata(1:i-1,:);ydata(i+1:size(ydata,1),:)];
    end
    subplot(121), plot(xdata,ydata,'rx','LineWidth',2,'MarkerSize',15);
    A = [size(x,1) sum(x);sum(x) sum(x.^2)];
    Y = [sum(y); sum(y.*x)];
    w1 = pinv(A)*Y;
    f = @(x) w1(1)+w1(2).*x;
    hold on; grid on; fplot(f,[0.1 1.1]);
    xlabel('x(t)');ylabel('y(t)');
    title('1st order');
    subplot(122), plot(xdata,ydata,'rx','LineWidth',2,'MarkerSize',15);
    yy1(i,:) = f(xx) ;
    X = [ones(size(x,1),1) x x.^2 x.^3 x.^4];
    w4 = pinv(X'*X)*X'*y;
    f = @(a) w4(1)+w4(2).*a+w4(3).*a.^2+w4(4).*a.^3+w4(5).*a.^4;
    hold on; grid on; fplot(f,[0.1 1.1]);
    xlabel('x(t)');ylabel('y(t)');
    title('4th order');
    yy4(i,:) = f(xx) ;
end

figure
subplot(1,2,1)
errorbar(xx,mean(yy1),std(yy1).*std(yy1));
xlabel('x(t)');ylabel('y(t)');
title('1st order, variance of model') ;
subplot(1,2,2)
errorbar(xx,mean(yy4),std(yy4).*std(yy4))
xlabel('x(t)');ylabel('y(t)');
title('4th order, variance of model') ;
```