# Robot Learning
# 3. Numerical Method-Regression

Jeong-Yean Yang

2020/10/22

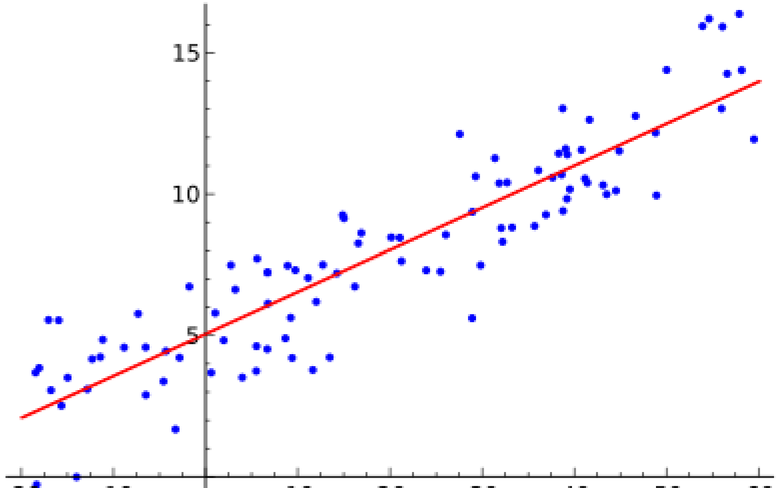
4

# Regression( or Fitting Problem)

# Linear Problem

# Regression

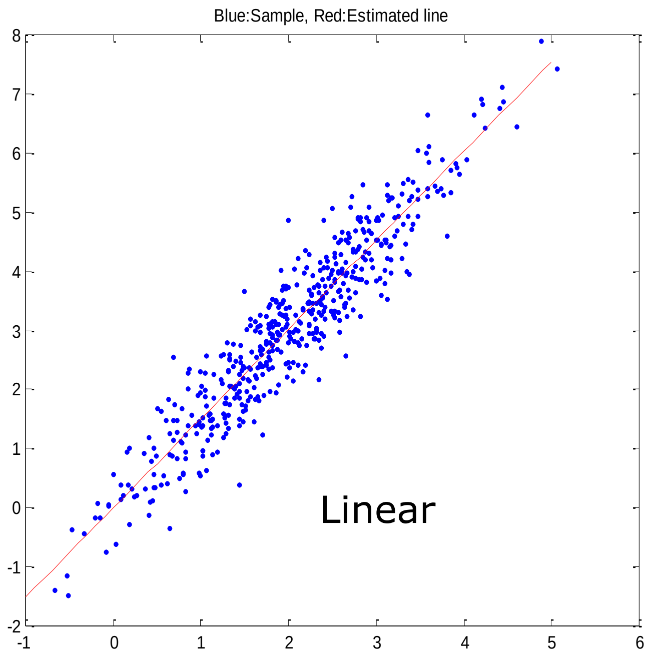
- The Most Important Issue in the field of Learnings.
- What is Regression?
  - Simply, Curve fitting.
  - Goal: find the best Curve or Line.

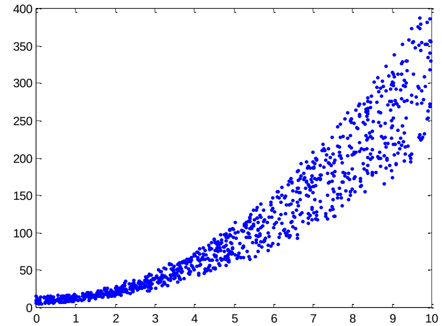



*** Important**
**All samples CANNOT satisfy the one Line!!!**
**→**
**It is an Optimal Problem**

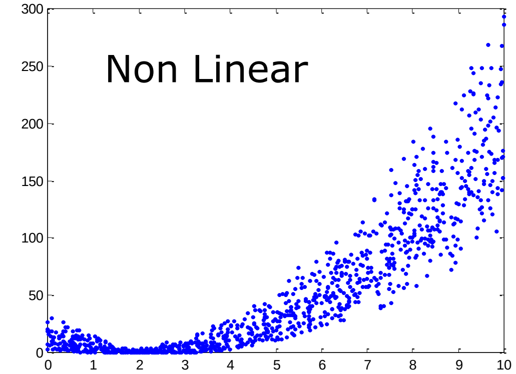
- In other words, Regression is in the filed of Optimization

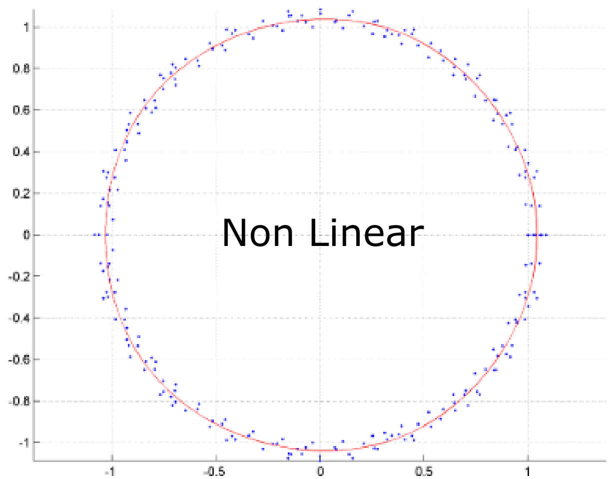
# Regression Types



Linear

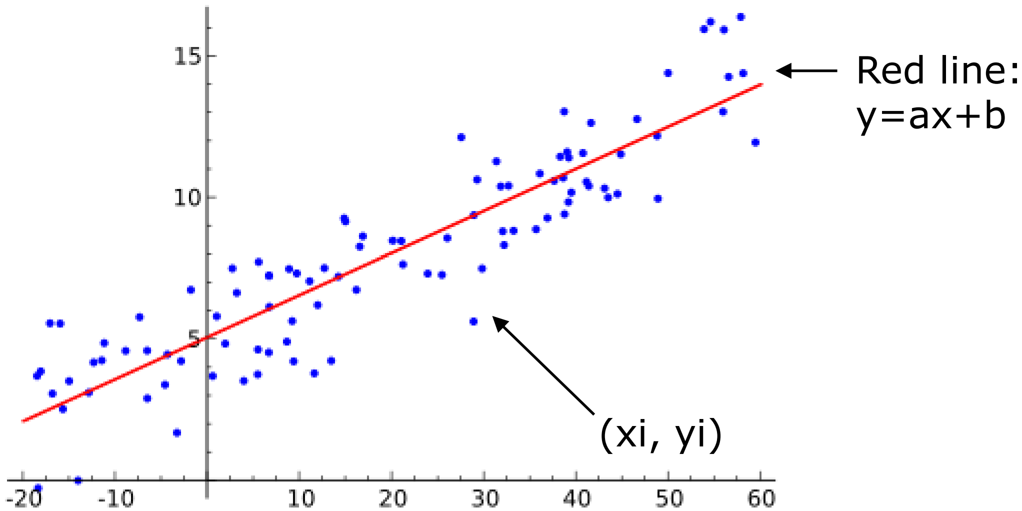



Non Linear
Stochastic

# Linear Regression

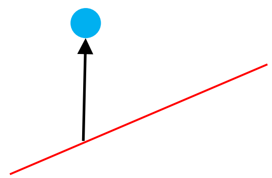
- Samples are given S={(x1,y1), (x2,y2),…, (xN,yN)}




**Knowns: (x1,y1),….(xN,yN)**
**Unknowns: a, b ← Our goal!!**

- Goal: Find the a and b for minimizing Error

# Linear Regression
# Minimize Error Function



Distance :

$$\| y_i - y \|^2$$

$$J = \sum_{i}^{N} e_i^{\,2} = \sum_{i}^{N} \| y_i - y \|^2 = \sum_{i}^{N} \| y_i - (ax_i + b) \|^2$$

- Definition of J
  - Error function or Cost function
- J is often the SUM of Squared Error (Least Square)

N=3
X1 = (x1,y1)=(1,2)
X2 = (x2,y2)=(2,3)
X3 = (x3,y3)=(2,2)

$$J = \sum_{i}^{3} \| y_i - (ax_i + b) \|^2$$

$$= (2-(a+b))^2 + (3-(2a+b))^2 + (2-(2a+b))^2 = J(a,b)$$

# How to Find the Minimum?
# ( in detailed ways later)

- 1. Differentiation

$$J = \sum_{i}^{N} \| y_i - y \|^2 = \sum_{i}^{N} \| y_i - (ax_i + b) \|^2 = J(a,b)$$

When $\dfrac{\partial J}{\partial a} = 0$ and $\dfrac{\partial J}{\partial b} = 0$, J has minima or maxima.

Example) $y = (x-1)^2$, $y' = 2(x-1) = 0$ $\therefore$ *when* x=1 , y has minimum

- 2. Iterative Method
  - More than thousands of methods exist.
  - Ex) Gradient Descent Method.

$$x_{n+1} = x_n - \eta \nabla J$$

**Why we use Iterative Method in many applications such as NN?**

# Linear Regression Solution

$$J = \sum_i^N \| y_i - y \|^2 = \sum_i^N \| y_i - (ax_i + b) \|^2 = J(a,b)$$

$$\frac{\partial J}{\partial a} = \frac{\partial}{\partial a} \sum_i^N (y_i - y)^2 = \frac{\partial}{\partial a}\left((y_1 - y)^2 + (y_2 - y)^2 + \ldots\right)$$

$$= \sum_i^N \frac{\partial}{\partial a}(y_i - y)^2 = \sum_i^N 2(y_i - y)\frac{\partial}{\partial a}(y_i - y) = \sum_i^N 2(y_i - y)\left(\frac{\partial y_i}{\partial a} - \frac{\partial y}{\partial a}\right)$$

$$= \sum_i^N 2(y_i - y)\left(0 - \frac{\partial y}{\partial a}\right) = \sum_i^N 2(y_i - y)\left(0 - \frac{\partial (ax_i + b)}{\partial a}\right) = \underline{\sum_i^N 2(y_i - y)(-x_i) = 0}$$

$$\frac{\partial J}{\partial b} = \frac{\partial}{\partial b} \sum_i^N (y_i - y)^2 = \sum_i^N 2(y_i - y)\frac{\partial}{\partial b}(y_i - y) =$$

$$\sum_i^N 2(y_i - y)\left(0 - \frac{\partial y}{\partial b}\right) = \sum_i^N 2(y_i - y)\left(0 - \frac{\partial (ax_i + b)}{\partial b}\right) = \underline{\sum_i^N 2(y_i - y)(-1) = 0}$$

# Linear Regression Solution

$$\sum_i^N 2(y_i - y)(-x_i) = \sum_i^N 2(y_i - ax_i - b)(-x_i) = 0$$

$$\therefore a\sum_i^N x_i x_i + b\sum_i^N x_i = \sum_i^N x_i y_i \;---(1)$$

$$\sum_i^N 2(y_i - ax_i - b)(-1) = 0$$

$$\therefore a\sum_i^N x_i + b\sum_i^N 1 = \sum_i^N y_i \;---(2)$$

$$\begin{bmatrix} \sum_i^N x_i x_i & \sum_i^N x_i \\ \sum_i^N x_i & \sum_i^N 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_i^N x_i y_i \\ \sum_i^N y_i \end{bmatrix}$$

**Oops,**
**It is Linear!**

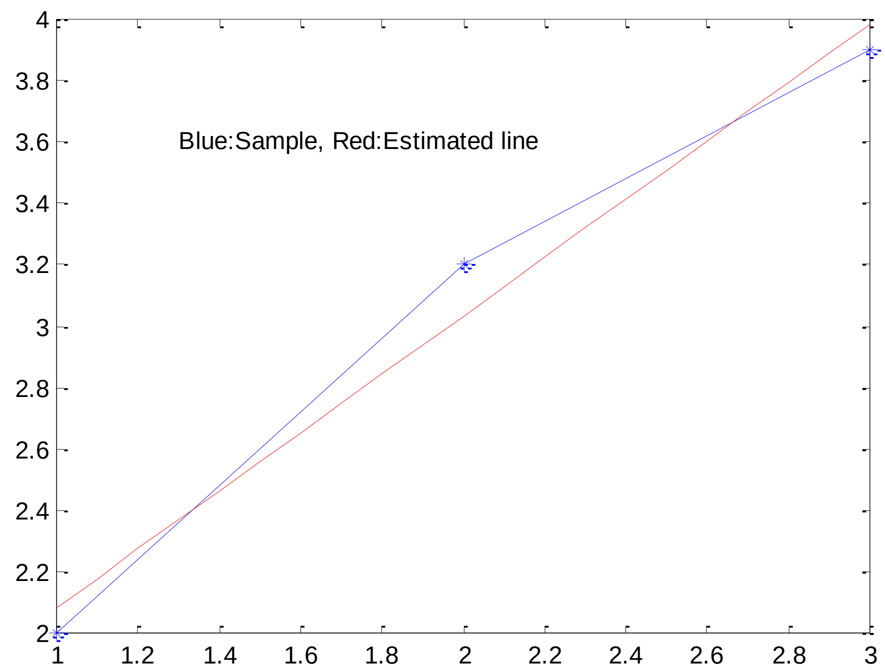
**Matrix calculation is O.K.**

- Test 5.m

# Simple Example Test5.

```
%Linear Regression Simple

X=[ 1 2
    2 3.2
    3 3.9];

N=3;

plot(X(:,1),X(:,2),'-*');

% find a, b
A11 = 1*1 + 2*2 + 3*3;
A12 = 1+2+3;
A21 = A12;
A22 = 1+1+1;
A = [ A11 A12; A21 A22];

B11 = 1*2+2*3.2+3*3.9;
B21 = 2+3.2+3.9;
B = [B11;B21];

X=inv(A)*B

a=X(1);
b=X(2);
```

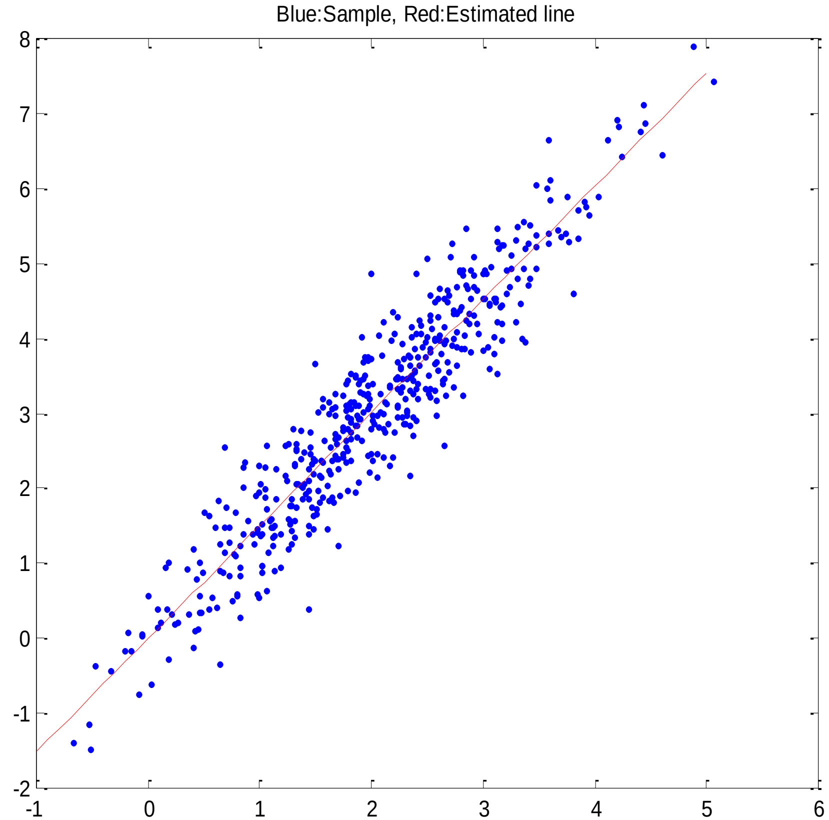
$$\begin{bmatrix} \sum_i^N x_i x_i & \sum_i^N x_i \\ \sum_i^N x_i & \sum_i^N 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_i^N x_i y_i \\ \sum_i^N y_i \end{bmatrix}$$



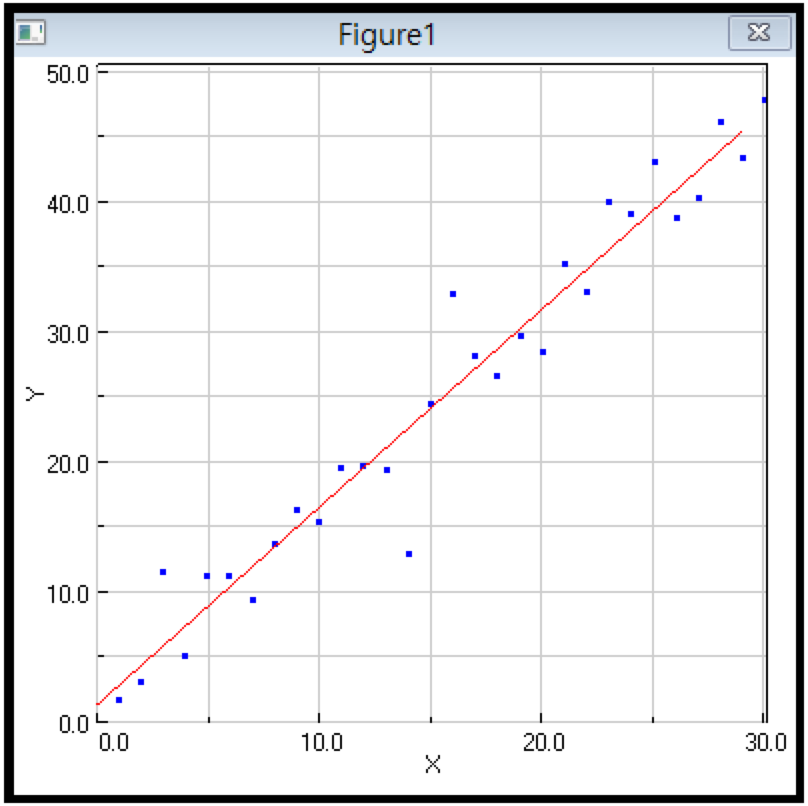
# Many Samples : Test6.m

```
m=[2,3];
s=[1, 1.5;
   1.5 2.5];
N=500;
X=mvnrnd(m,s,N);

plot(X(:,1),X(:,2),'.');

% find a, b
A11 = sum(X(:,1).^2);
A12 = sum(X(:,1));
A21 = A12;
A22 = N;
A = [ A11 A12; A21 A22];

B11 = sum(X(:,1).*X(:,2));
B21 = sum(X(:,2));
B = [B11;B21];

X=inv(A)*B

a=X(1);
b=X(2);

x=-1:0.1:5;
y=a*x + b;
```

$$\begin{bmatrix} \sum_i^N x_i x_i & \sum_i^N x_i \\ \sum_i^N x_i & \sum_i^N 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_i^N x_i y_i \\ \sum_i^N y_i \end{bmatrix}$$



# Linear regression l2regline ex/ml/l2regline




import l2regline

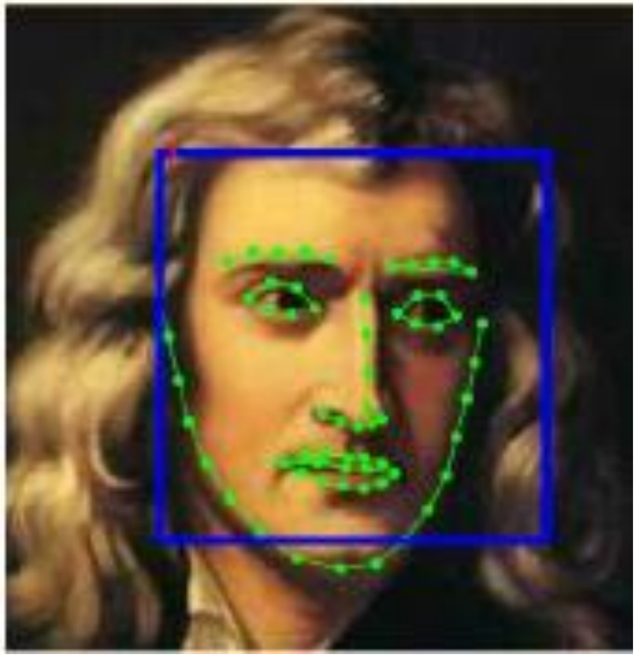
l2regline.test()

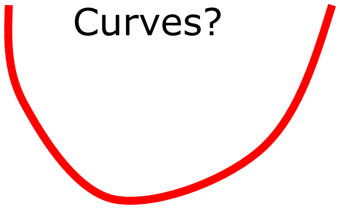
$$\begin{bmatrix} \sum_i^N x_i x_i & \sum_i^N x_i \\ \sum_i^N x_i & \sum_i^N 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_i^N x_i y_i \\ \sum_i^N y_i \end{bmatrix}$$

$$J = \sum_i^N \| y_i - y \|^2 = \sum_i^N \| y_i - (ax_i + b) \|^2 = J(a, b)$$

# Question

- Regression works very well with **Matrix Operation**
  (or Linear algebra)
- Well, How about the Curve?





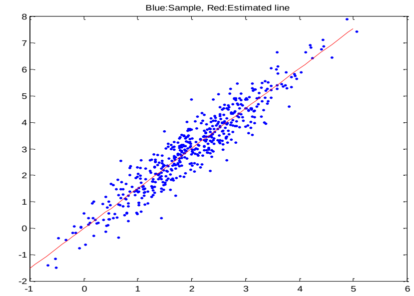
4

# Regression( or Fitting Problem)

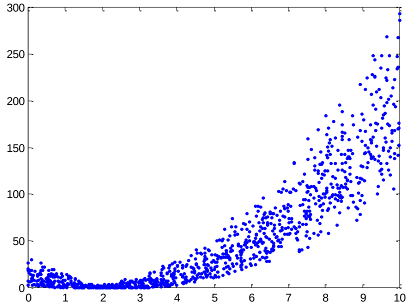
# Non-Linear Problem

# Question:
# Curves are all Non Linear?

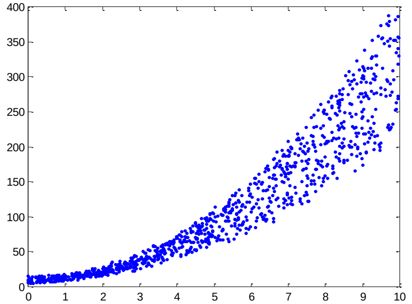
- Straight Line is Linear




- Some curves are Non Linear



- But, other curves are Linear.

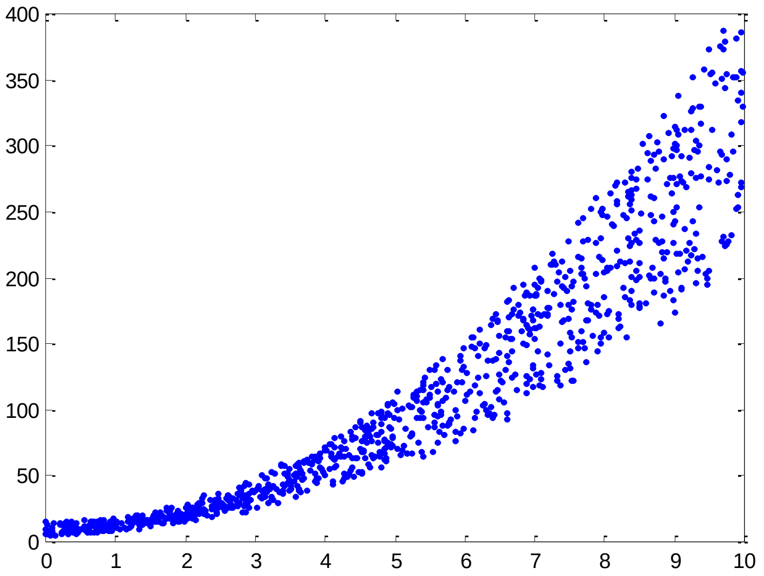


- Let's Answer

  Which Curves are Nonlinear..

# Regression with $y = ax^2 + b$

```
N=1000
x=10*rand(N,1);

for i=1:N
    yo = rand*10-5;
    a  = 3+1*(rand*2-1);
    b  = 2+1*(rand*2-1);

    y(i) = a*(x(i))^2+b;
end

plot(x,y,'.');

X=[ x,y'];

save X1 X;
```




Generate data
with Gen1.m

# Regression Model

$$J = \sum_{i}^{N} \| y_i - y \|^2 = \sum_{i}^{N} (y_i - a x_i^{\ 2} - \mathrm{b})^2$$

$$\frac{\partial J}{\partial a} = 2 \sum_{i}^{N} (y_i - a x_i^{\ 2} - \mathrm{b})(-x_i^{\ 2}) = 0$$

$$\frac{\partial J}{\partial b} = 2 \sum_{i}^{N} (y_i - a x_i^{\ 2} - \mathrm{b})(-1) = 0$$

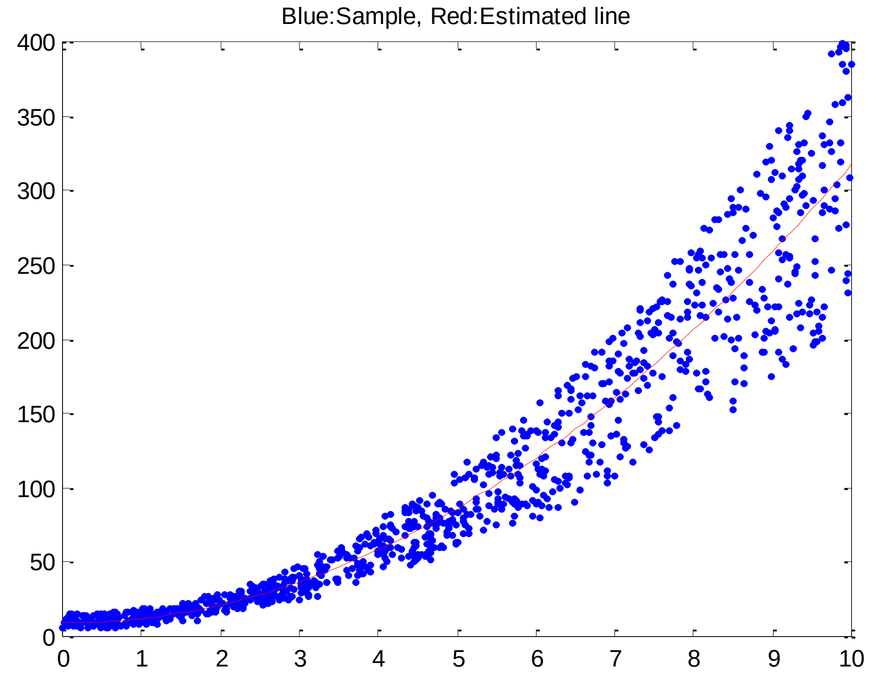
$$\begin{cases} a \sum_{i}^{N} x_i^{\ 4} + b \sum_{i}^{N} x_i^{\ 2} = \sum_{i}^{N} y_i x_i^{\ 2} \\ a \sum_{i}^{N} x_i^{\ 2} + b \sum_{i}^{N} 1 = \sum_{i}^{N} y_i \end{cases}$$

$$\begin{bmatrix} \sum_{i}^{N} x_i^{\ 4} & \sum_{i}^{N} x_i^{\ 2} \\ \sum_{i}^{N} x_i^{\ 2} & \sum_{i}^{N} 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i}^{N} y_i x_i^{\ 2} \\ \sum_{i}^{N} y_i \end{bmatrix}$$

Matrix Operation
is O.K.

# Test 7

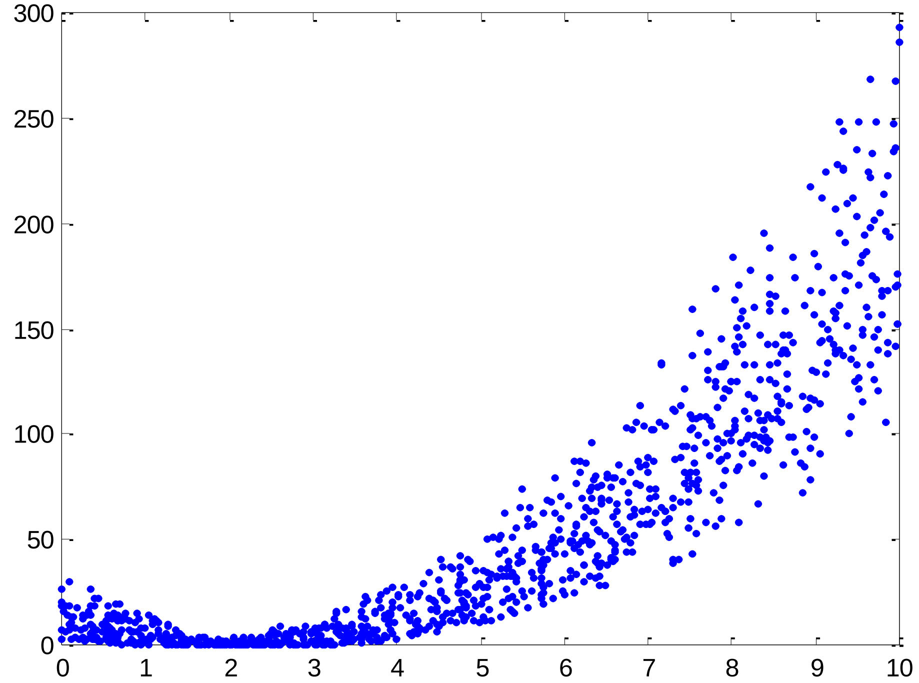
```
% sample generation
gen1
plot(X(:,1),X(:,2),'.');

% find a, b
A11 = sum(X(:,1).^4);
A12 = sum(X(:,1).^2);
A21 = A12;
A22 = N;
A = [ A11 A12; A21 A22];

B11 = sum(X(:,2).*X(:,1).^2);
B21 = sum(X(:,2));
B = [B11;B21];

X=inv(A)*B

a=X(1);
b=X(2);

x=0:0.1:10;
y=a*x.^2 + b;
hold
plot(x,y,'r');

title('Blue:Sample, Red:Estimated line');
```




```
X =

    3.0853
    9.0465
```

# Regression with $y = a(x-b)^2$

```
N=1000
x=10*rand(N,1);

for i=1:N
    yo = rand*10-5;
    a  = 3+1*(rand*2-1);
    b  = 2+1*(rand*2-1);

    y(i)  = a*(x(i)-b)^2;
end

plot(x,y,'.');

X=[ x,y'];
```




Generate data
with Gen2.m

# Regression Model

$$J = \sum_{i}^{N} \| y_i - y \|^2 = \sum_{i}^{N} (y_i - a(x_i - b)^2)^2$$

$$\frac{\partial J}{\partial a} = 2\sum_{i}^{N} (y_i - a(x_i - b)^2)(-(x_i - b)^2) = 0$$

$$\frac{\partial J}{\partial b} = 2\sum_{i}^{N} (y_i - a(x_i - b)^2)(2a(x_i - b)) = 0$$

$$\therefore \sum_{i}^{N} (y_i - a(x_i - b)^2)(-(x_i - b)^2) = f(a,b) = 0$$

$$\therefore \sum_{i}^{N} (y_i - a(x_i - b)^2)(2a(x_i - b)) = g(a,b) = 0$$

**Non Linear Equation → N-R Eq**

# Questions: What is the Key Point?

# That is Linear or Not.

- Don't take it the wrong way owing to Curve types

- Linear Regression,

$y = ax + b$

$$J = \sum_i^N \| y_i - y \|^2 = \sum_i^N \| y_i - (ax_i + b) \|^2$$

$$\begin{bmatrix} \sum_i^N x_i x_i & \sum_i^N x_i \\ \sum_i^N x_i & \sum_i^N 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_i^N x_i y_i \\ \sum_i^N y_i \end{bmatrix}$$

$y = ax^2 + b$

$$J = \sum_i^N \| y_i - y \|^2 = \sum_i^N (y_i - ax_i{}^2 - \mathrm{b})^2$$

$$\begin{bmatrix} \sum_i^N x_i{}^4 & \sum_i^N x_i{}^2 \\ \sum_i^N x_i{}^2 & \sum_i^N 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_i^N y_i x_i{}^2 \\ \sum_i^N y_i \end{bmatrix}$$

$$y = ax^2 + b = a() + b$$

$$\therefore Linear$$

# Definition of Linearity

- Scalar Multiplication

$$if\ \ v \in L,\ \ it\ also\ satisfies\ \ \alpha v \in L$$

- Additivity

$$if\ \ v_1 \in L \text{ and } v_2 \in L,$$
$$then\ it\ also\ satisfies\ \ \alpha v_1 + \beta v_2 \in L$$

* Remind that

$$y = a(x-b)^2 = a([\,]-b) = a[\,] - ab$$

$$\therefore \text{It is NOT Linear!}$$

# Back To pp.19 $y = a(x-b)^2$
# This Regression has Nonlinear Equations

$$\sum_i^N (y_i - a(x_i - b)^2)(-(x_i - b)^2) = f(a,b) = 0$$

$$\sum_i^N (y_i - a(x_i - b)^2)(2a(x_i - b)) = g(a,b) = 0$$

- How to solve it?

1. $$J = \sum_i^N \| y_i - y \|^2$$

$$\frac{\partial J}{\partial a} = f(a,b) = 0, \frac{\partial J}{\partial b} = g(a,b) = 0$$

Nonlinear Newton-Raphson in pp. 19

$$F(\hat{x} + h) = 0 = F(\hat{x}) + J\hat{h}$$

$$\therefore \hat{h} = -J^{-1}F$$

$$\rightarrow \hat{x}_{k+1} = \hat{x}_k + \hat{h} = \hat{x}_k - J^{-1}F$$

- How to solve it?

2. Optimization

For example, Gradient Descent Method (GDM)

$$J = \sum_{i}^{N} \| y_i - y \|^2 = \sum_{i}^{N} (y_i - a(x_i - b)^2)^2$$

$$w = \begin{bmatrix} a \\ b \end{bmatrix}$$

$$w_{n+1} = w_n - \eta \nabla J$$

$$J = \sum_{i}^{N} \| y_i - y \|^2 = \sum_{i}^{N} (y_i - a(x_i - b)^2)^2$$

$$\nabla J = \frac{\partial J}{\partial a} \mathrm{i} + \frac{\partial J}{\partial b} \mathrm{j}$$

$$= \left[ 2\sum_{i}^{N} (y_i - a(x_i - b)^2)(-(x_i - b)^2) \right] \mathrm{i} + \left[ 2\sum_{i}^{N} (y_i - a(x_i - b)^2)(2a(x_i - b)) \right] \mathrm{j}$$

# How to Find the Minimum?

- 1. Differentiation (Linear Equation)

$$J = \sum_{i}^{N} \| y_i - y \|^2 = \sum_{i}^{N} \| y_i - (ax_i + b) \|^2 = J(a,b)$$

$$J = J(w_1, w_2, w_3, \ldots. w_N)$$

$$\frac{\partial J}{\partial w_1} = 0, \quad \frac{\partial J}{\partial w_2} = 0, \quad \frac{\partial J}{\partial w_3} = 0, \ldots \quad \frac{\partial J}{\partial w_N} = 0$$

$$if \;\; it \;\; is \;\; linear \;\; eqs, \;\; then \;\; Aw = b$$

- 2. Iterative Method (Non-Linear Equation)
  - More than thousands of methods exist.
  - Ex) Gradient Descent Method.

$$x_{n+1} = x_n - \eta \nabla J$$