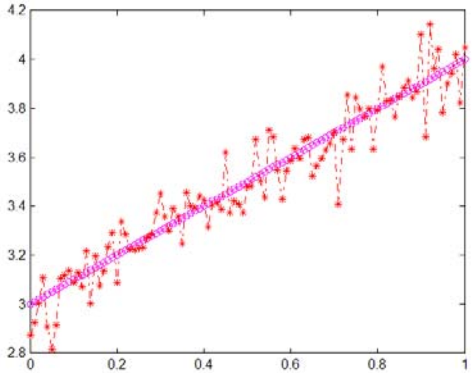
Regression and Linear Regression

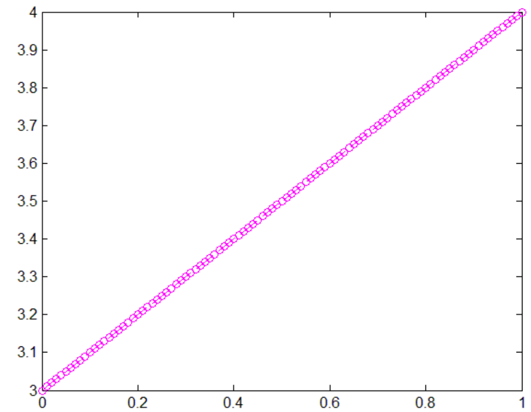
Yan PEI/裴岩 /ペイ　イエン

peiyan@u-aizu.ac.jp

Data=signal + noise

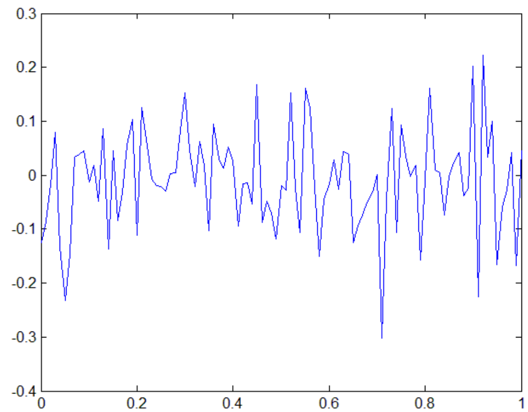


=



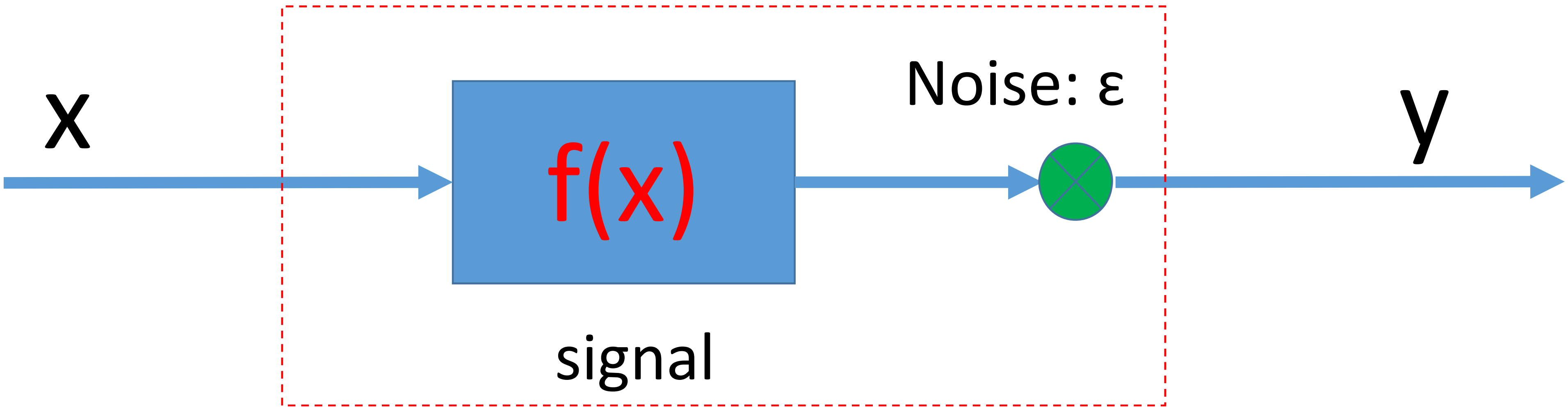
signal

+



noise

Data=signal + noise



$$y=f(x)+\varepsilon$$

Linear regression model

ε:

(1).independent random variable

(2).identical distributed

(3).zero mean and finite variance

So y is also a random variable, the linear regression model is $E(y \mid x) = f(x) + E(\varepsilon) = f(x)$

Linear regression model

The conditional expectation of y is given by

$$
\begin{aligned}
E(y_n \mid x_n) &= w_0 + w_1 x_{n1} + \ldots + w_d x_{nd} \\
&= [1, x_{n1}, \ldots, x_{nd}] \begin{bmatrix} w_0 \\ \vdots \\ w_d \end{bmatrix} \\
&= [1, x^T] w \\
&= \overline{x}^T w
\end{aligned}
$$

Data Generation Model

$$y_n = f(x_n) + \varepsilon_n = E(y_n \mid x_n) + \varepsilon_n = [1, x^T]w + \varepsilon_n$$

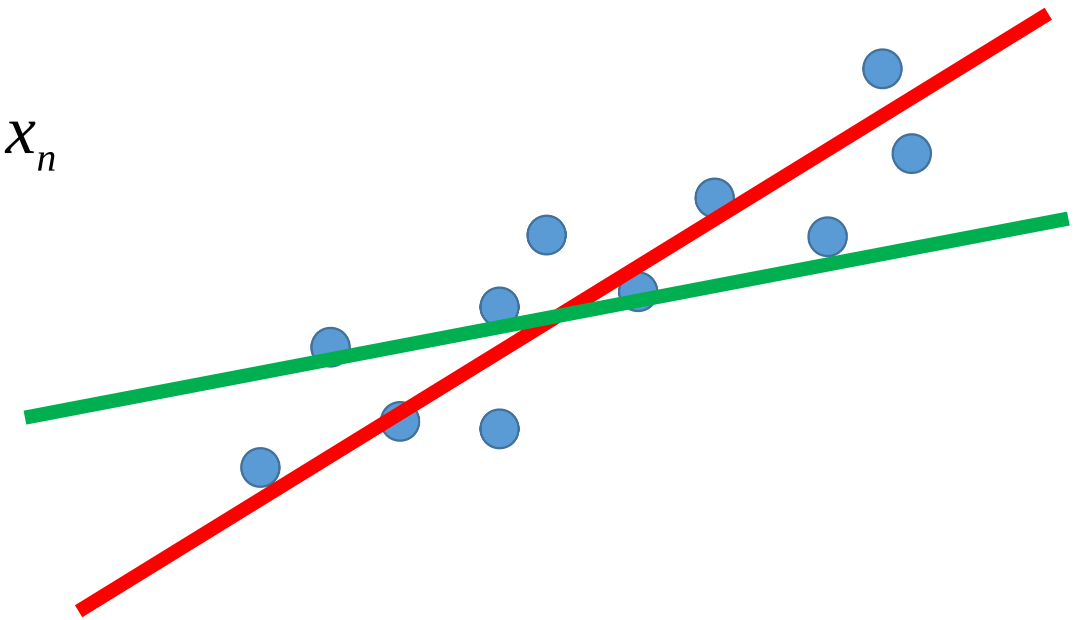
$$if : x_n \in R^1$$

$$then : w = [w_0, w_1] \,\&\, f(x_n) = w_0 + w_1 x_n$$

$$slope \rightarrow w_1$$

$$intercept \rightarrow w_0$$

$$normal\_vector \rightarrow \begin{bmatrix} w_1 \\ -1 \end{bmatrix}$$



Linear Regression

Given N observations x1,…xn and the corresponding y1,…,yn, respectively.

$$X = \begin{bmatrix} x_1^T \\ \vdots \\ x_N^T \end{bmatrix} \qquad \bar{X} = \begin{bmatrix} 1 & x_1^T \\ \vdots & \vdots \\ 1 & x_N^T \end{bmatrix} = \begin{bmatrix} \bar{x}_1^T \\ \vdots \\ \bar{x}_N^T \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

Linear Regression

1. Linear regression analyzes the relationship between two variables, X and y
2. For each subject, we know both X and y, and we want to find the best straight line through the data
3. In some cases, the slop and/or intercept have a scientific meaning
4. In some other cases, we can use the linear regression line as a modal to find new x from y, ot y from x.

Linear Regression Model in Matrix Form

Given N observations x1,…xn and the corresponding y1,…,yn, respectively.

$$\begin{cases} y_1 \approx \bar{x}_1^T w + \varepsilon_1 \\ \vdots \\ y_N \approx \bar{x}_N^T w + \varepsilon_N \end{cases} \rightarrow \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \approx \begin{bmatrix} \bar{x}_1^T \\ \vdots \\ \bar{x}_N^T \end{bmatrix} w + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

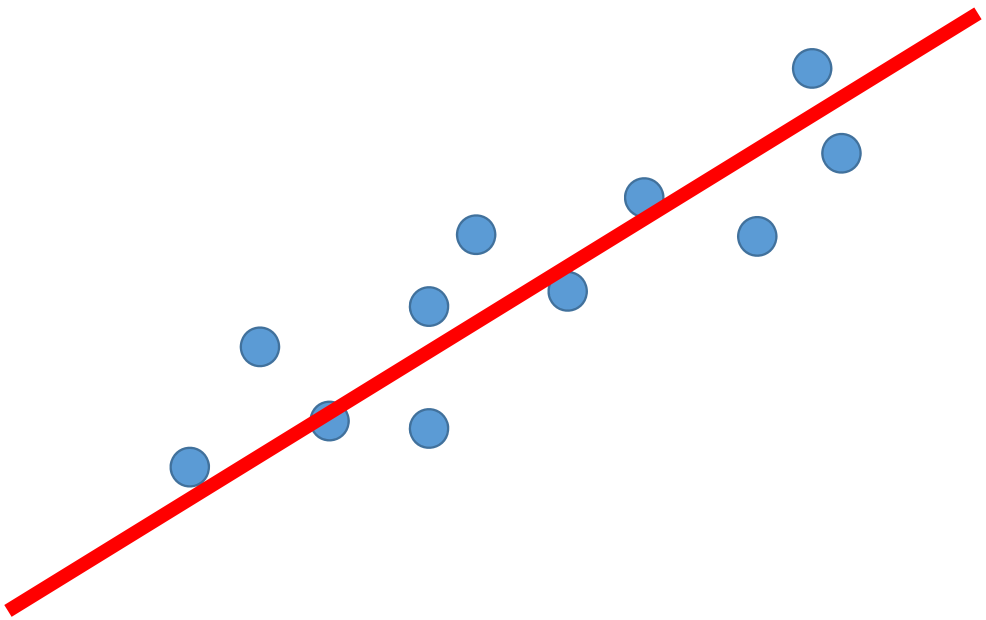
$$\rightarrow y \approx \bar{X} w + \varepsilon$$

Estimating Linear Regression Model

The target objective of this problem is

$$w = \arg\min_{w} \| y - \bar{X}w \|^2$$

RSS: residual sum of squares is minimum



Estimating Linear Regression Model

Solve this problem

$$\| y - \bar{X}w \|^2 = (y - \bar{X}w)^T (y - \bar{X}w)$$
$$= y^T y - 2w^T \bar{X}^T y + w^T \bar{X}^T \bar{X}w$$

$$\frac{\partial}{\partial w} \| y - \bar{X}w \|^2 = \frac{\partial}{\partial w}(y^T y - 2w^T \bar{X}^T y + w^T \bar{X}^T \bar{X}w)$$
$$= -2\bar{X}^T y + 2\bar{X}^T \bar{X}w = 0$$
$$\rightarrow \bar{X}^T \bar{X}w = \bar{X}^T y$$ Normal Equations

Estimating Linear Regression Model

Solve this problem

$$\bar{X}^T \bar{X} w = \bar{X}^T y$$

$$\rightarrow \hat{w} = (\bar{X}^T \bar{X})^{-1} \bar{X}^T y$$

$$\rightarrow \hat{y} = \bar{X}\hat{w} = \bar{X}(\bar{X}^T \bar{X})^{-1} \bar{X}^T y$$

Kernel based Linear Regression

Kernel-based Linear Regression

Given N observations x1,…xn and the corresponding y1,…,yn, respectively. And a feature mapping and its kernel function

$$X = \begin{bmatrix} \varphi(x_1)^T \\ \vdots \\ \varphi(x_N)^T \end{bmatrix} \qquad \bar{X} = \begin{bmatrix} 1 & \varphi(x_1)^T \\ \vdots & \vdots \\ 1 & \varphi(x_N)^T \end{bmatrix} = [1_{N\times 1}, X] \qquad y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

Kernel based Linear Regression

Normal Equation

$$\frac{\partial}{\partial w}\| y - \bar{X}w \|^2 = \frac{\partial}{\partial w}(y^T y - 2w^T \bar{X}^T y + w^T \bar{X}^T \bar{X}w)$$

$$= -2\bar{X}^T y + 2\bar{X}^T \bar{X}w = 0$$

$$\rightarrow \boxed{\bar{X}^T \bar{X}w = \bar{X}^T y}$$

1 Dual form

$$\bar{X}^T \bar{X}w = \bar{X}^T y$$

$$\rightarrow w = (\bar{X}^T \bar{X})^{-1} \bar{X}^T y$$

$$\rightarrow w = (\bar{X}^T \boxed{\bar{X})(\bar{X}^T \bar{X})^{-2} \bar{X}^T y}$$

$$\rightarrow w = \bar{X}^T \alpha$$

2 Again from normal form

$$\bar{X}^T \bar{X}w = \bar{X}^T y$$

$$\rightarrow \bar{X}^T \bar{X}\bar{X}^T \alpha = \bar{X}^T y$$

$$\rightarrow \bar{X}\bar{X}^T \bar{X}\bar{X}^T \alpha = \bar{X}\bar{X}^T y$$

$$\rightarrow \bar{K}^2 \alpha = \bar{K}y$$

$$\rightarrow \alpha = \bar{K}^{-1} y$$

Kernel based Linear Regression

$$\bar{X}^T \bar{X} w = \bar{X}^T y$$
$$\rightarrow \bar{X}^T \bar{X}\bar{X}^T \alpha = \bar{X}^T y$$
$$\rightarrow \bar{X}\bar{X}^T \bar{X}\bar{X}^T \alpha = \bar{X}\bar{X}^T y$$
$$\rightarrow \bar{K}^2 \alpha = \bar{K} y$$
$$\rightarrow \alpha = \boxed{\bar{K}}^{-1} y$$

note $\bar{K} = \bar{X}\bar{X}^T$

$$= [1_{N\times 1}, X] \begin{bmatrix} 1_{1\times N} \\ X^T \end{bmatrix}$$
$$= 1_{N\times N} + XX^T$$
$$= 1_{N\times N} + K$$

Hence

$$\alpha = \bar{K}^{-1} y = (1_{N\times N} + K)^{-1} y$$
$$y = \bar{X} w = (1_{N\times N} + K)\alpha$$

Kernel based Linear Regression

For a unknown x

$$
\begin{aligned}
f(x) &= [1, \varphi(x)^T]w \\
&= [1, \varphi(x)^T]\bar{X}^T\alpha \\
&= [1, \varphi(x)^T]\begin{bmatrix} 1_{1\times N} \\ X^T \end{bmatrix}\alpha \\
&= (1_{1\times N} + (X\varphi(x))^T)\alpha \\
&= [1 + K(x_1, x), \ldots, 1 + K(x_N, x)]\alpha \\
&= (\alpha_1 + \ldots + \alpha_N) + \alpha_1 K(x_1, x) + \ldots + \alpha_N K(x_N, x)
\end{aligned}
$$