Stochastic Signal Processing

Lesson 4: Sequences of Random Variables & Basic of Estimation

Weize Sun

Last week's review examples

1. Assume that X is a r.v follows exponential distribution with parameter$\lambda > 1$, find the expectation of $Y = e^X$

Note that for exponential distribution:

- Mean: $\mathrm{E(X)} = \int_0^\infty x\lambda e^{-\lambda x} dx = 1/\lambda$
- Variance:$\mathrm{D(X)} = 1/\lambda^2$
- pdf:$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$

Solution:

$$E(Y) = \int_0^\infty e^x \lambda e^{-\lambda x} dx = \int_0^\infty \lambda e^{-(\lambda - 1)x} dx = \frac{\lambda}{\lambda - 1}$$

Last week's review examples

2. Given $Y = 0.5X + 1, Z = -0.5X + 1 \ and \ D(X) = 1$, calculate the covariance matrix of the vector (X, Y, Z)

Note: $\mathrm{cov}(X, Y) = E\left[\left(X - \overline{X}\right)\left(Y - \overline{Y}\right)\right] = E(XY) - E(X)E(Y)$

Solution:

$$\mathrm{k}_{1,1} = D(X) = 1, \mathrm{k}_{2,2} = D(Y) = 0.25, \mathrm{k}_{3,3} = D(Z) = 0.25$$
$$\mathrm{k}_{1,2} = \mathrm{cov}(X, Y) = 0.5, \mathrm{k}_{1,3} = \mathrm{cov}(X, Z) = -0.5$$

$$\begin{aligned}\mathrm{k}_{2,3} = \mathrm{cov}(Y, Z) \quad &= E(-0.25X^2 + 1) - (E(0.5X) + 1)(E(-0.5X) + 1) \\ &= -0.25(E(X^2) - E(X)^2) = -0.25D(X) = -0.25\end{aligned}$$

Therefore:

$$\mathrm{K} = \begin{pmatrix} 1 & 0.5 & -0.5 \\ 0.5 & 0.25 & -0.25 \\ -0.5 & -0.25 & 0.25 \end{pmatrix}$$

General concepts

- A **random vector** is a vector

$$\boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_{\mathrm{n}}] \qquad (4\text{-}1)$$

whose components $\boldsymbol{x}_i$ are random variables. In some books it is called **multi-dimensional r.vs**.

- The probability that $\boldsymbol{X}$ is in a region D of the n-dimensional space equals the probability masses in D :

$$P\{\boldsymbol{X} \in D\} = \int_D f(\boldsymbol{X}) d\boldsymbol{X}\,,\ \boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n] \qquad (4\text{-}2)$$

Where $f(\boldsymbol{X}) = f(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = \frac{\partial^n F(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)}{\partial \boldsymbol{x}_1, \dots, \partial \boldsymbol{x}_n}$ (4-3)

is the joint (or, multivariate) density (pdf) of the random variables x_{i} and

$$F(X) = F(x_1, \dots, x_n) = P\{\boldsymbol{x}_1 \le x_1, \dots, \boldsymbol{x}_n \le x_n\} \qquad (4\text{-}4)$$

is their joint distribution (CDF).

General concepts

- If we substitute in $F(x_1, \dots, x_n)$ certain variables by ∞, we obtain the joint distribution of the remaining variables. If we integrate $f(x_1, \dots, x_n)$ with respect to certain variables, we obtain the joint density of the remaining variables:

$$F(x_1, x_3) = F(x_1, \infty, x_3, \infty)$$

$$f(x_1, x_3) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2, x_3, x_4) dx_2 dx_4 \qquad \text{(4-5)}$$

- $f(x_1, x_3)$ is the joint density of the random variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_3$ and it is in general different from the joint density $f\ (x_2, x_4)$ of the random variables $\boldsymbol{x}_2$ and $\boldsymbol{x}_4$.

- Similarly, the density $f_i(\boldsymbol{x}_1)$ of the random variable $\boldsymbol{x}_1$ will often be denoted by $f(\boldsymbol{x}_i)$.

Transformations

- Given k functions $g_1(X), \dots, g_k(X)$, $\quad X = [x_1, \dots, x_n]$, we form the random variables

$$\boldsymbol{y}_1 = g_1(\boldsymbol{X}), \dots, \mathbf{y}_k = g_k(\boldsymbol{X}) \tag{4-6}$$

- The statistics of these random variables can be determined in terms of the statistics of $\boldsymbol{X}$ as in last lesson.
  - If $k < n$, then we could determine first the joint density of the n random variables $\boldsymbol{y_1}, \dots, \boldsymbol{y_k}, \boldsymbol{x_{k+1}}, \dots, \boldsymbol{x_n}$ and then use the generalization of (4-5) to eliminate the $\boldsymbol{x}$'s.
  - If $k > n$, then the random variables $\boldsymbol{y_{n+1}}, \dots, \boldsymbol{y_k}$ can be expressed in terms of $\boldsymbol{y_1}, \dots, \boldsymbol{y_n}$. In this case, the masses in the k space are singular and can be determined in terms of the joint density of $\boldsymbol{y_1}, \dots, \boldsymbol{y_n}$. It suffices, therefore, to assume that $k = n$.

Transformations

- To find the density $f_y(y_1, \ldots, y_n)$ of the random vector $\boldsymbol{Y} = [\boldsymbol{y_1}, \ldots, \boldsymbol{y_n}]$ for a specific set of numbers $y_1, \ldots, y_n$, we solve the system

$$g_1(X) = y_1, \ldots, g_n(X) = y_n \qquad \text{(4-7)}$$

- If this system has no solutions, then $f_y(y_1, \ldots, y_n) = 0$.
- If it has a single solution $X = [x_1, \ldots, x_n]$, then

$$f_y(y_1, \ldots, y_n) = f_x(x_1, \ldots, x_n)|J| \qquad \text{(4-8)}$$

where $J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \cdots & & \cdots \\ \frac{\partial \mathrm{x}_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$ is the jacobian of the transformation (4-7).

注：对多元随机变量的变换(slides 5-7)，这里仅要求了解。

Independence

- The random variables $\boldsymbol{x_1, \dots, x_n}$ are called (mutually) independent if the events $\{\mathbf{x}_1 \leq x_1\}, \dots, \{\mathbf{x}_n \leq x_n\}$ are independent:

$$F(x_1, \dots, x_n) = F(x_1) \cdots F(x_n)$$
$$f(x_1, \dots, x_n) = f(x_1) \cdots f(x_n) \qquad (4\text{-}9)$$

Conditional densities

- Now we define the conditional densities of **multi-dimensional r.vs** or **random vector** as follows:

- For 2 **r.vs**, we have $f(y \mid x) = f(x, y)/f(x)$. Similarly, we conclude that the conditional density of the random variables $\boldsymbol{x_n}, \dots, \boldsymbol{x_{k+1}}$ assuming $\boldsymbol{x_k}, \dots, \boldsymbol{x_1}$ given by

$$f(x_n, \dots, x_{k+1} \mid x_k, \dots, x_1) = \frac{f(x_1, \dots, x_k, \dots, x_n)}{f(x_1, \dots, x_k)} \tag{4-10}$$

- The conditional CDF is:

$$F(x_n, \dots, x_{k+1} \mid x_k, \dots, x_1)$$

$$= \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_{k+1}} f(\alpha_n, \dots, \alpha_{k+1} \mid x_k, \dots, x_1) d\alpha_{k+1} \dots d\alpha_n \tag{4-11}$$

- Chain rule（链式法则）：

$$f(x_1, \dots, x_n) = f(x_n \mid x_{n-1}, \dots, x_1) \cdots f(x_2 \mid x_1) f(x_1) \tag{4-12}$$

Conditional expected values – Discrete type

- For discrete type r.vs, all densities are replaced by probabilities and all integrals by sums.

- For example, if the random variables $\boldsymbol{x_1},\ \boldsymbol{x_2}, \boldsymbol{x_3}$ take the values a_i, b_k, c_r, respectively , then

$$P\{\boldsymbol{x}_1 = a_i, \boldsymbol{x_2} = b_k, \boldsymbol{x}_3 = c_r\} = \sum_k \sum_r P\{\boldsymbol{x}_1 = a_i \mid b_k, c_r\} P\{\boldsymbol{x}_2 = b_k \mid c_r\} P\{\boldsymbol{x_3} = c_r\} \quad \text{(4-13)}$$

- The conditional mean of the random variables:

$$E\{\boldsymbol{x}_1 \mid x_2, \dots, x_n\} = \int_{-\infty}^{\infty} x_1 f(x_1 \mid x_2, \dots, x_n) dx_1 \qquad \text{(4-14)}$$

This is a function of $x_2, \dots, x_n$; it defines the random variable $E\{\boldsymbol{x_1} \mid \boldsymbol{x_2}, \dots, \boldsymbol{x_n}\}$. Multiplying it by the pdf $f(x_2, \dots, x_n)$ and integrating (**which is, calculate the mean of** $E\{\boldsymbol{x_1} \mid \boldsymbol{x_2}, \dots, \boldsymbol{x_n}\}$) gives:

$$E\{E\{\boldsymbol{x_1} \mid \boldsymbol{x_2}, \dots, \boldsymbol{x_n}\}\} = E\{\boldsymbol{x}_1\} \qquad \text{(4-15)}$$

which is, taking the expetation of '$\boldsymbol{x_1}$ **under all conditions of** $\boldsymbol{x_2}, \dots, \boldsymbol{x_n}$', **we can get the mean of** $\boldsymbol{x_1}$

Conditional expected values

- Another kind of 'chain rule':

$$E\{\boldsymbol{x}_1 \mid x_3\} = \int_{-\infty}^{\infty} E\{\boldsymbol{x}_1 \mid x_2, x_3\} f(x_2 \mid x_3) dx_2 \qquad \text{(4-16)}$$

Proof:

$$\int_{-\infty}^{\infty} E\{\boldsymbol{x}_1 \mid x_2, x_3\} f(x_2 \mid x_3) dx_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f(x_1 \mid x_2, x_3) f(x_2 \mid x_3) dx_1 dx_2$$

$$= (1/f(x_3)) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f(x_1 \mid x_2, x_3) f(x_2 \mid x_3) f(x_3) dx_1 dx_2$$

$$= (1/f(x_3)) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f(x_1, x_2, x_3) dx_1 dx_2 = (1/f(x_3)) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f(x_1, x_2, x_3) dx_2 dx_1$$

$$= (1/f(x_3)) \int_{-\infty}^{\infty} x_1 f(x_1, x_3) dx_1 = E\{\boldsymbol{x}_1 \mid x_3\}$$

- For the discrete case, all densities are replaced by probabilities and all integrals by sums

$$E\{\boldsymbol{x}_1 \mid c_r\} = \sum E\{\boldsymbol{x}_1 | b_k, c_r\} P\{\boldsymbol{x}_2 = b_k \mid c_r\} \qquad \text{(4-17)}$$

Mean square estimation

- The estimation problem is fundamental in the applications of probability. Here we introduce the main ideas using as illustration the estimation of a random variable $\boldsymbol{y}$ in terms of another random variable $\boldsymbol{x}$. Throughout this analysis, the optimality criterion will be the minimization of the mean square value (abbreviation: MS) of the estimation error. (you might see the word 'MSE - mean square error' in some articles or paper sometimes)

- We start with a brief explanation of the underlying concepts in the context of repeated trials, considering first the problem of estimating the random variable $\boldsymbol{y}$ by a constant.

- As we know, the distribution function $F(y)$ of the random variable y determines completely its statistics.
  - This does not mean that if we know $F(y)$ we can predict the value $\boldsymbol{y}(\zeta)$ of $\boldsymbol{y}$ at some future trial.
  - However, we can estimate the unknown $\boldsymbol{y}(\zeta)$ by some number c, and, the knowledge of $F(y)$ can guide us in the selection of c.

Mean square estimation

- If $\boldsymbol{y}$ is estimated by a constant c, then our problem is to select c so as to minimize the error $\boldsymbol{y}(\zeta) - c$ in some sense. A reasonable criterion for selecting c might be 'the average error is close to 0':

$$\frac{\boldsymbol{y}(\zeta_1) - c + \cdots + \boldsymbol{y}(\zeta_n) - c}{n} \simeq 0$$

This would lead to the conclusion that c should equal the mean of $\boldsymbol{y}$.

- Another criterion for selecting c might be the minimization of the average of $|\boldsymbol{y}(\zeta) - c|$. In this case, the optimum c is the median of $\boldsymbol{y}$.

- In our analysis, we consider only MS estimates. This means that c should be such as to minimize the average of $|\boldsymbol{y}(\zeta) - c|^2$. This criterion is in general useful but it is selected mainly because it leads to simple results. In this case, the best c is again the mean of $\boldsymbol{y}$.

Mean square estimation

- The MS estimation of the random variable $\boldsymbol{y}$ by a constant c can be phrased as follows: Find c such that the second moment (MS error)

$$e = E\{(\boldsymbol{y}-c)^2\} = \int_{-\infty}^{\infty} (y-c)^2 f(y)dy \qquad (4\text{-}18)$$

of the difference (error) $(\boldsymbol{y}-c)$ is minimum.

- Clearly, e depends on c and it is minimum if

$$\frac{de}{dc} = \int_{-\infty}^{\infty} 2(y-c)f(y)dy = 0$$

- that is, when $c = \int_{-\infty}^{\infty} yf(y)dy$ (the mean), the minimum can be achieved.

- Thus

$$c = E\{\boldsymbol{y}\} = \int_{-\infty}^{\infty} yf(y)dy \qquad (4\text{-}19)$$

Linear MS Estimation

- The linear estimation problem is the estimation of the random variable $\boldsymbol{y}$ in terms of a linear function $a\boldsymbol{x} + b$ of $\boldsymbol{x}$. The problem now is to find the constants a and b so as to minimize the MS error

$$e = E\{[\boldsymbol{y} - (a\boldsymbol{x} + b)]^2\} \tag{4-19}$$

- Indeed, the MS error e is a function of a and b and it is minimum if $\partial e/\partial a = 0$ and $\partial e/\partial b = 0$. The first equation yields $\frac{\partial e}{\partial a} = E\{2[\boldsymbol{y} - (a\boldsymbol{x} + b)](-\boldsymbol{x})\} = 0$, which is:

- **Orthogonality principle**: when

$$E\{[\boldsymbol{y} - (a\boldsymbol{x} + b)]\boldsymbol{x}\} = 0 \tag{4-20}$$

That means, the optimum linear MS estimate $a\boldsymbol{x} + b$ of $\boldsymbol{y}$ is such that the estimation error $\{\boldsymbol{y} - (a\boldsymbol{x} + b)\}$ is orthogonal to the data $\boldsymbol{x}$. It is fundamental in MS estimation and will be used extensively.

Linear MS Estimation

$$e = E\{[\boldsymbol{y} - (a\boldsymbol{x} + b)]^2\} \tag{4-19}$$

- Note that $\partial e/\partial b = E\{2[\boldsymbol{y} - (a\boldsymbol{x} + b)](-1)\} = 0$, which is $E\{[\boldsymbol{y} - (a\boldsymbol{x} + b)](1)\} = 0$
- **Orthogonality principle**: when

$$E\{[\mathbf{y} - (a\mathbf{x} + b)]\mathbf{x}\} = 0 \tag{4-20}$$

$$E\{[\mathbf{y} - (a\mathbf{x} + b)](\mathbf{1})\} = 0 \tag{4-21}$$

(4-20) & (4-21) can be used to estimate $\boldsymbol{a}$ & $\boldsymbol{b}$, we will introduce the matrix form solution here only.

- In deep learning, the $\boldsymbol{a}$ is referred to as 'weight parameter' and $\boldsymbol{b}$ is as 'bias'.

https://zhuanlan.zhihu.com/p/29815081

1 从感知机到神经网络

感知机的模型，它是一个有若干输入和一个输出的模型，如下图:

输出和输入之间学习到一个线性关系，得到中间输出结果：

$$z = \sum_{i=1}^{m} w_i x_i + b$$

Linear Estimation – the general case

- The linear MS estimate of $\boldsymbol{s}$ in terms of the random variables $\mathbf{x}_i$ is the sum

$$\hat{\mathbf{s}} = a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n + b \qquad (4\text{-}22)$$

where $a_1, \dots, a_n, b$ are $n+1$ constants such that the MS value

$$e = E\{(\mathbf{s} - \hat{\mathbf{s}})^2\} = E\{[\mathbf{s} - (a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n + b)]^2\} \qquad (4\text{-}23)$$

of the estimation error $\mathbf{s} - \hat{\mathbf{s}}$ is minimum.

- Orthogonality principle: P is minimum if the error $\mathbf{s} - \hat{\mathbf{s}}$ is orthogonal to the data $\mathbf{x}_i$:

$$E\{[\mathbf{s} - (a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n + b)]\mathbf{x}_i\} = 0, \text{ for } i = 1, \dots, n \qquad (4\text{-}24)$$

and $E\{[\mathbf{s} - (a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n + b)](1)\} = 0$ for estimating b

Linear Estimation – the matrix form, and the real operation

- Example 4-1: The data of watermelon sales in a certain market at time t are as follows (q_t represents the quantity of watermelon sold at time t, and p_t represents the sales price at time t)

p_t	q_t
2.5	1
2.0	3
1.5	5
1.0	7
0.5	9
0	11

Calculate the $a\&b$ from $q_t = ap_t + b$

Linear Estimation – the matrix form, and the real operation

- Example 4-1: The data of watermelon sales in a certain market at time t are as follows (q_t represents the quantity of watermelon sold at time t, and p_t represents the sales price at time t)

p_t	q_t
2.5	1
2.0	3
1.5	5
1.0	7
0.5	9
0	11

Calculate the $a\&b$ from $q_t = ap_t + b$

Solution:

$$q_t = 11 - 4p_t$$

Linear Estimation – the matrix form, and the real operation

- Example 4-1.2: if there some randomness of the relationship:

p_t	q_t	Probability
2.5	0	0.25
	1	0.50
	2	0.25
2.0	2	0.25
	3	0.50
	4	0.25
...	...	...
0	10	0.25
	11	0.50
	12	0.25

Calculate the $a\&b$ from $q_t \approx ap_t + b$?

Linear Estimation – the matrix form, and the real operation

- Example 4-1.2: if there some randomness of the relationship:

p_t	q_t	Probability
	0	0.25
2.5	1	0.50
	2	0.25
	2	0.25
2.0	3	0.50
	4	0.25
…	…	…
	10	0.25
0	11	0.50
	12	0.25

Calculate the $a\&b$ from $q_t \approx ap_t + b$?

Solution:

$$q_t = 11 - 4p_t + \varepsilon_t$$

The ε_t represents the error of the regression at time t

Linear Estimation – the matrix form, and the real operation

- How to calculate $a\&b$ by minimizing $\boldsymbol{e} = E\{[\mathbf{y} - (a + b\mathbf{x})]^2\}$ in matrix form?

$$y_1 = a + bx_1 + \varepsilon_1$$
$$y_2 = a + bx_2 + \varepsilon_2$$
$$\dots$$
$$y_M = a + bx_M + \varepsilon_M$$

In matrix form

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_M \end{bmatrix} A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_M \end{bmatrix} \boldsymbol{z} = \begin{bmatrix} a \\ b \end{bmatrix} \boldsymbol{e} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix}$$

$$\mathbf{y} = A\boldsymbol{z} + \boldsymbol{e}$$

target：find $\boldsymbol{z} = \begin{bmatrix} a \\ b \end{bmatrix}$ to min $\| \boldsymbol{e} \|_2^2 = \| \mathbf{y} - A\boldsymbol{z} \|_2^2$.

$$\Rightarrow \boldsymbol{z} = A^{+}\mathbf{y}$$

Where $A^{+} = (A^T A)^{-1} A^T$ is the Pseudo inverse

Linear Estimation – the matrix form, and the real operation

- How to calculate $a\&b$ by minimizing $\boldsymbol{e} = E\{[\mathbf{y} - (a + b\mathbf{x})]^2\}$ in matrix form?

Two parameters to be trained

$$y_1 = a + bx_1 + \varepsilon_1$$
$$y_2 = a + bx_2 + \varepsilon_2$$
$$\dots$$
$$y_M = a + bx_M + \varepsilon_M$$

Vector form

In matrix form

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_M \end{bmatrix} A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_M \end{bmatrix} \boldsymbol{z} = \begin{bmatrix} a \\ b \end{bmatrix} \boldsymbol{e} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix}$$

Training data set, totally M data pieces

Minimize training error

$$\mathbf{y} = A\boldsymbol{z} + \boldsymbol{e}$$

target : find $\boldsymbol{z} = \begin{bmatrix} a \\ b \end{bmatrix}$ to min $\| \boldsymbol{e} \|_2^2 = \| \mathbf{y} - A\boldsymbol{z} \|_2^2$.

$$\Rightarrow \boldsymbol{z} = A^{+}\mathbf{y}$$

Where $A^{+} = (A^T A)^{-1} A^T$ is the Pseudo inverse

Estimation results for training set

Linear Estimation – the matrix form, and the real operation

- How to calculate $a\&b$ by minimizing $\boldsymbol{e} = E\{[\mathbf{y} - (a + b\mathbf{x})]^2\}$ in matrix form?

$k + 1$ parameters to be trained

$$y_1 = a + b_1 x_{11} + \cdots + b_k x_{k1} + \varepsilon_1$$
$$y_2 = a + b_1 x_{12} + \cdots + b_k x_{k2} + \varepsilon_2$$
$$\cdots$$
$$y_M = a + b_1 x_M + \cdots + b_k x_{kM} + \varepsilon_M$$

In matrix form

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_M \end{bmatrix} \; A = \begin{array}{c} \begin{array}{ccccc} & \mathbf{x}_1 & & \mathbf{x}_k \end{array} \\ \begin{bmatrix} 1 & \boxed{\begin{matrix} x_{11} \\ x_{12} \\ \vdots \\ x_{M1} \end{matrix}} & \begin{matrix} \cdots \\ \cdots \\ \ddots \\ \cdots \end{matrix} & \boxed{\begin{matrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kM} \end{matrix}} \end{bmatrix} \end{array} \; \boldsymbol{z} = \begin{bmatrix} a \\ b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix} \; \boldsymbol{e} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix}$$

Training data set, totally M data pieces

Vector form

Minimize training error

$$\mathbf{y} = A\boldsymbol{z} + \boldsymbol{e}$$

$$\boldsymbol{z} = A^{+}\mathbf{y}$$

if $\quad \mathrm{M} \geqslant k + 1$

(M equations, find $k + 1$ values)

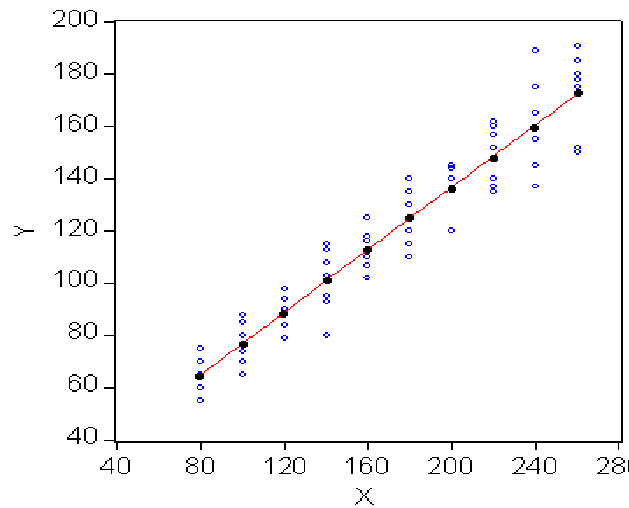
Estimation results for training set

It is well know as 'linear regression'!

Linear Estimation – the matrix form, and the real operation

- The target of the linear regression is: given a series of x_i & y， find a_i & b, so that some law (规律) can be determined
- For example: if there are 60 persons in one village with income x & consumption y as the table below, find the relationship between x & y

Y \ X	80	100	120	140	160	180	200	220	240	260
	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	–	88	–	113	125	140	–	160	189	185
	–	–	–	115	–	–	–	162	–	191
户数	5	6	5	7	6	6	5	7	6	5
总支出	325	462	445	707	678	750	685	1043	966	1211



Solution: $y_i = 17.00 + 0.6x_i$

Then, what will be a person's consumption is his/her income is $x_i = 135$?

Can you use Matlab to do the calculation?

It might be somewhat related to experimental report 4, and will be discussed later