18. Statistic data processing

- Observed data
  - Suppose we have observed n sets of data on m items, $x_1, \ldots, x_m$, as

    $$\{x_1(1), \ldots, x_m(1)\}, \{x_1(2), \ldots, x_m(2)\}, \ldots, \{x_1(n), \ldots, x_m(n)\}.$$

    Let us represent these data as

    $$X = \begin{bmatrix} x_1(1) & \ldots & x_m(1) \\ \vdots & & \vdots \\ x_1(n) & \ldots & x_m(n) \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1 & \ldots & \boldsymbol{x}_m \end{bmatrix} = \begin{bmatrix} \boldsymbol{d}^T(1) \\ \vdots \\ \boldsymbol{d}^T(n) \end{bmatrix},$$

    where $\boldsymbol{x}_i = \begin{bmatrix} x_i(1) \\ \vdots \\ x_i(n) \end{bmatrix}$ is a vector of data on item x_i, and

    $\boldsymbol{d}(k) = \begin{bmatrix} x_1(k) \\ \vdots \\ x_m(k) \end{bmatrix}$ is a vector of k-th samples on all the items.

18. Statistic data processing

- Basic statistics（統計量）
  - Mean vector

$$\bar{\boldsymbol{d}} = \frac{1}{n}\sum_{k=1}^{n}\boldsymbol{d}(k) = \begin{bmatrix} \frac{1}{n}\sum_{k=1}^{n} x_1(k) \\ \vdots \\ \frac{1}{n}\sum_{k=1}^{n} x_m(k) \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_m \end{bmatrix}$$

  - Mean deviation matrix（平均偏差行列）

$$\tilde{X} = X - \begin{bmatrix} \bar{\boldsymbol{d}}^T \\ \vdots \\ \bar{\boldsymbol{d}}^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{d}^T(1) - \bar{\boldsymbol{d}}^T \\ \vdots \\ \boldsymbol{d}^T(n) - \bar{\boldsymbol{d}}^T \end{bmatrix} = \begin{bmatrix} x_1(1) - \bar{x}_1 & \dots & x_m(1) - \bar{x}_m \\ \vdots & & \vdots \\ x_1(n) - \bar{x}_1 & \dots & x_m(n) - \bar{x}_m \end{bmatrix}$$

18. Statistic data processing

- Sample covariance matrix（標本共分散行列）

  $$S = \frac{1}{n}\tilde{X}^T\tilde{X}$$

  whose (i,j)-component is

  $$s_{ij} = \frac{1}{n}\sum_{k=1}^{n}(x_i(k) - \bar{x}_i)(x_j(k) - \bar{x}_j)$$

  which is the sample covariance of item x_i and item x_j.

- Correlation coefficient（相関係数）

  $$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}},\ -1 \leq r_{ij} \leq 1.$$

18. Statistic data processing

- Multiple regression analysis（重回帰分析）
  - Expresses a dependent variable（従属変数，目的変数）y in terms of multiple independent variables（独立変数，説明変数）$x_1, \ldots, x_m$ ：

$$y = a_0 + a_1 x_1 + \cdots + a_m x_m.$$

Suppose we have n sets of data on $x_1, \ldots, x_m$ and y,

$$\{x_1(1), \ldots, x_m(1); y(1)\}, \ldots, \{x_1(n), \ldots, x_m(n); y(n)\}.$$

We assume the relation between the data is written as

$$y(k) = a_0 + a_1 x_1(k) + \cdots + a_m x_m(k) + \varepsilon(k),\ k = 1, \ldots, n,$$

where $\varepsilon(k)$ is measurement noise (measurement error).

18. Statistic data processing

The relation

$$y(k) = a_0 + a_1 x_1(k) + \cdots + a_m x_m(k) + \varepsilon(k),\ k = 1, \ldots, n,$$

can be written in a matrix form as,

$$\underbrace{\begin{bmatrix} y(1) \\ \vdots \\ y(n) \end{bmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{bmatrix} 1 & x_1(1) & \ldots & x_m(1) \\ \vdots & \vdots & & \vdots \\ 1 & x_1(n) & \ldots & x_m(n) \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix}}_{\boldsymbol{a}} + \underbrace{\begin{bmatrix} \varepsilon(1) \\ \vdots \\ \varepsilon(n) \end{bmatrix}}_{\boldsymbol{\varepsilon}},$$

$$\boldsymbol{y} = X\boldsymbol{a} + \boldsymbol{\varepsilon}.$$

Here $\varepsilon(k),\ k = 1, \ldots, n$ are statistically independent（統計的に独立）with each other and are from an identical distribution.

We want to find the value of vector $\boldsymbol{a} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix}$ above.

18. Statistic data processing

– Least squares method（最小2乗法）

- Let us assume $\underline{\mathrm{E}}[\boldsymbol{\varepsilon}] = \mathbf{0},\ \mathrm{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma^2 I.$

  Expectation（期待値）

- We find $\boldsymbol{a}$ that minimizes $e(\boldsymbol{a})$: $e(\boldsymbol{a}) = (\boldsymbol{y} - X\boldsymbol{a})^T(\boldsymbol{y} - X\boldsymbol{a}).$

$$
\begin{aligned}
e(\boldsymbol{a}) &= (\boldsymbol{y}^T - \boldsymbol{a}^T X^T)(\boldsymbol{y} - X\boldsymbol{a}) \\
&= \boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{y}^T X\boldsymbol{a} - \boldsymbol{a}^T X^T\boldsymbol{y} + \boldsymbol{a}^T X^T X\boldsymbol{a}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{a}} e(\boldsymbol{a}) &= \mathbf{0} - (\boldsymbol{y}^T X)^T - X^T\boldsymbol{y} + X^T X\boldsymbol{a} + (\boldsymbol{a}^T X^T X)^T \\
&= -X^T\boldsymbol{y} - X^T\boldsymbol{y} + X^T X\boldsymbol{a} + X^T X\boldsymbol{a} \\
&= 2(X^T X\boldsymbol{a} - X^T\boldsymbol{y})
\end{aligned}
$$

Therefore it is necessary to find $\hat{\boldsymbol{a}}$ that satisfies

$$X^T X\hat{\boldsymbol{a}} = X^T\boldsymbol{y},$$

which is called a normal equation（正規方程式）.

18. Statistic data processing

When the matrix $X^T X$ is non-singular, we obtain

$$\hat{\boldsymbol{a}} = \underset{\text{Pseudo-inverse matrix of } X}{\underline{(X^T X)^{-1} X^T}} \boldsymbol{y}.$$

Since

$$\begin{aligned}\hat{\boldsymbol{a}} &= (X^T X)^{-1} X^T (X\boldsymbol{a} + \boldsymbol{\varepsilon}) \\ &= (X^T X)^{-1} X^T X\boldsymbol{a} + (X^T X)^{-1} X^T \boldsymbol{\varepsilon} \\ &= \boldsymbol{a} + (X^T X)^{-1} X^T \boldsymbol{\varepsilon},\end{aligned}$$

we derive

$$\begin{aligned}\mathrm{E}[\hat{\boldsymbol{a}}] &= \boldsymbol{a} + \mathrm{E}[(X^T X)^{-1} X^T \boldsymbol{\varepsilon}] \\ &= \boldsymbol{a} + (X^T X)^{-1} X^T \underset{\mathbf{0}}{\underline{\mathrm{E}[\boldsymbol{\varepsilon}]}} \\ &= \boldsymbol{a}.\end{aligned}$$

The expectation of derived $\hat{\boldsymbol{a}}$ is equal to the true $\boldsymbol{a}$.
$\hat{\boldsymbol{a}}$ is said to be an unbiased estimator（不偏推定量）.

18. Statistic data processing

- Example
  - Data

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} 5 \\ 7 \\ 9 \\ 11 \end{bmatrix}$$

  - Estimation of $\boldsymbol{a}$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}, \; X^T \boldsymbol{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 5 \\ 7 \\ 9 \\ 11 \end{bmatrix} = \begin{bmatrix} 32 \\ 90 \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{4 \cdot 30 - 10 \cdot 10} \begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{5} \end{bmatrix}$$

$$\hat{\boldsymbol{a}} = (X^T X)^{-1} X^T \boldsymbol{y} = \begin{bmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{5} \end{bmatrix} \begin{bmatrix} 32 \\ 90 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

18. Statistic data processing

- Maximum likelihood estimation（最尤推定）

Estimates the parameter vector $\boldsymbol{a}$ that affects the probabilistic distribution of data vector $\boldsymbol{y}$ by maximizing the conditional probability density function（条件付き確率密度関数）$f(\boldsymbol{y}|\boldsymbol{a})$.

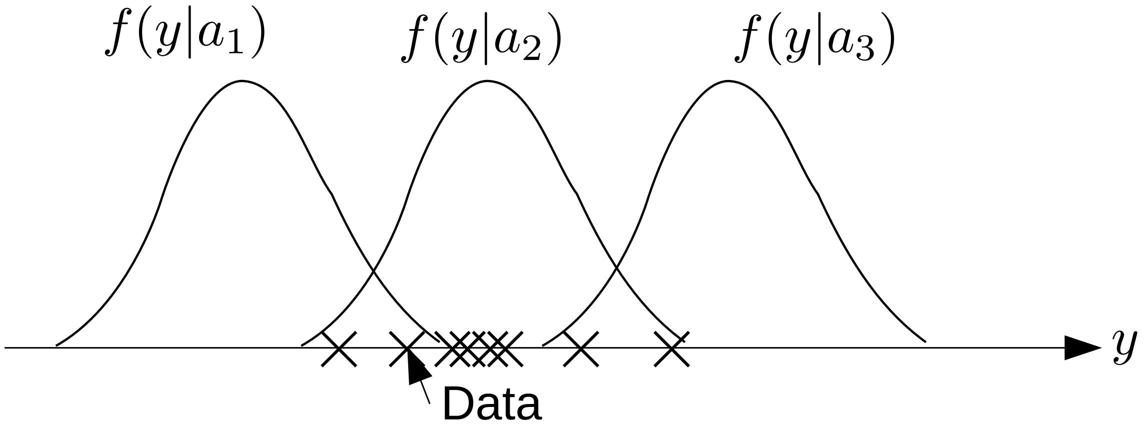
When a certain data vector $\boldsymbol{y}$ is given and substituted into $f(\boldsymbol{y}|\boldsymbol{a})$, $f(\boldsymbol{y}|\boldsymbol{a})$ becomes a function of $\boldsymbol{a}$.
We represent this function as

$$L(\boldsymbol{a}) = f(\boldsymbol{y}|\boldsymbol{a}),$$

which is called a likelihood function（尤度関数）.

The conditional probability density function $f(\boldsymbol{y}|\boldsymbol{a})$ specifies the probabilistic distribution of $\boldsymbol{y}$ when the parameter value $\boldsymbol{a}$ is given.
When a certain value of $\boldsymbol{y}$ is obtained, it is natural to assume that this value of $\boldsymbol{y}$ is obtained because it has a high probability density value (or, roughly speaking it has a high probability).

18. Statistic data processing



Data are obtained because they have high probability density values.
So in the above, $f(y|a_2)$ is the most suitable probability density function and therefore the parameter value a_2 is the most suitable value.

18. Statistic data processing

In the equation

$$\boldsymbol{y} = X\boldsymbol{a} + \boldsymbol{\varepsilon},$$

let us assume that $\boldsymbol{\varepsilon}$ is from an n-dimensional gaussian distribution（ガウス分布，正規分布） with mean vector $\mathbf{0}$ and covariance matrix $\sigma^2 I$.

Then $\boldsymbol{y}$ is from an n-dimensional gaussian distribution with mean vector $X\boldsymbol{a}$ and covariance matrix $\sigma^2 I$.

For a given X, let us assume we obtained data vector $\boldsymbol{y}$. Then the likelihood function is

$$L(\boldsymbol{a}) = f(\boldsymbol{y}|X, \boldsymbol{a}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - X\boldsymbol{a})^T(\boldsymbol{y} - X\boldsymbol{a})\right\}.$$

18. Statistic data processing

$L(\boldsymbol{a})$ is maximized if and only if $\ell(\boldsymbol{a}) = \log L(\boldsymbol{a})$ is maximized.

Since

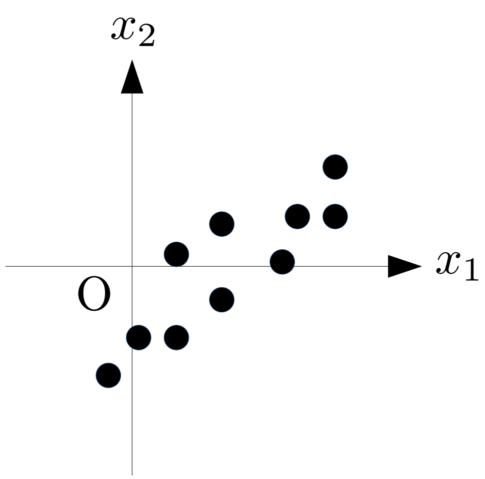
$$\ell(\boldsymbol{a}) = \underbrace{\log \frac{1}{(2\pi\sigma^2)^{n/2}}}_{\text{Constant}} - \frac{1}{2\sigma^2}(\boldsymbol{y} - X\boldsymbol{a})^T(\boldsymbol{y} - X\boldsymbol{a}),$$

$\ell(\boldsymbol{a})$ is maximized when $(\boldsymbol{y} - X\boldsymbol{a})^T(\boldsymbol{y} - X\boldsymbol{a})$ is minimized.

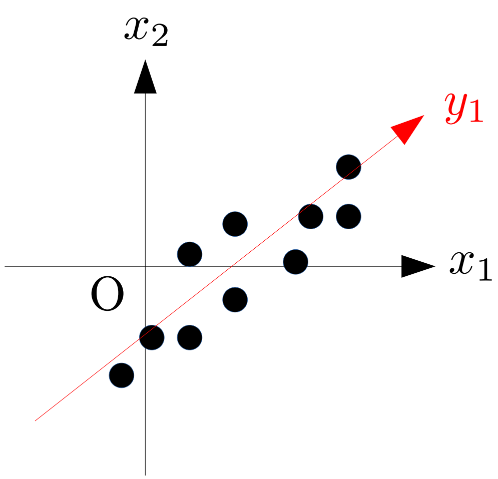
Therefore, in this case where the measurement noise is from a gaussian distribution, the maximum likelihood estimator is equal to the least squares estimator.

18. Statistic data processing

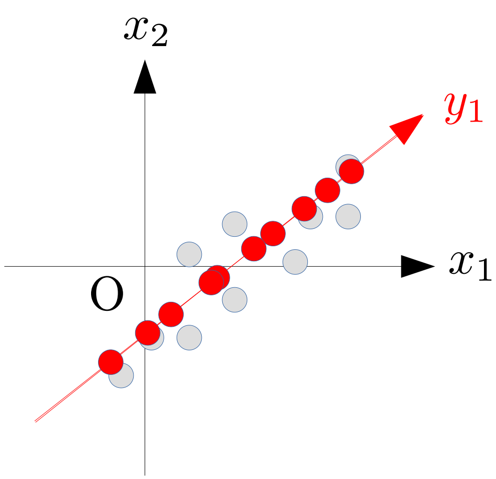
- Principal Component Analysis (PCA)（主成分分析）
  - PCA reduces dimensionality of high-dimensional data without loosing much information



Original **2-dimensional** data



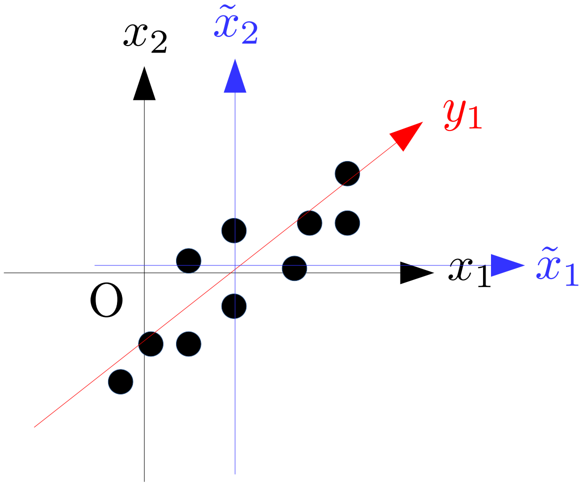
A **1-dimensional** y_1 coordinate



y_1 coordinate can display the data distribution fairly well

Reason: variance on the y_1 axis is large

18. Statistic data processing



- Shift the coordinate axes so that the center of data distribution becomes the origin

$$\tilde{x}_i(k) = x_i(k) - \bar{x}_i,\ \bar{x}_i = \frac{1}{n}\sum_{k=1}^{n} x_i(k)$$

- Generate new coordinates

$$y_1(k) = \sum_{i=1}^{m} w_i \tilde{x}_i(k) = \boldsymbol{w}^T \tilde{\boldsymbol{d}}(k),$$

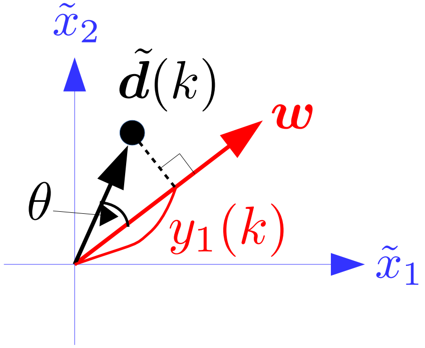
$$\tilde{\boldsymbol{d}}(k) = \begin{bmatrix} \tilde{x}_1(k) \\ \vdots \\ \tilde{x}_m(k) \end{bmatrix} = \begin{bmatrix} x_1(k) - \bar{x}_1 \\ \vdots \\ x_m(k) - \bar{x}_m \end{bmatrix} = \boldsymbol{d}(k) - \bar{\boldsymbol{d}},$$

$$i = 1, \ldots, m,\ k = 1, \ldots, n$$

We only need to determine the new coordinate direction, and thus the norm of vector $\boldsymbol{w}$ does not matter. Therefore, we set

$\|\boldsymbol{w}\|_2 = \sqrt{\boldsymbol{w}^T\boldsymbol{w}} = 1$.

$$\boldsymbol{w}^T \tilde{\boldsymbol{d}}(k) = \|\boldsymbol{w}\|_2 \|\tilde{\boldsymbol{d}}(k)\|_2 \cos\theta = \|\tilde{\boldsymbol{d}}(k)\|_2 \cos\theta$$



18. Statistic data processing

We find the value of $\boldsymbol{w}$ that maximizes the variance calculated from the data (the sample variance) of y_1 under the condition that $\boldsymbol{w}^T\boldsymbol{w} = 1$.

$$
\begin{aligned}
s_{y1} &= \frac{1}{n}\sum_{k=1}^{n}\left(y_1(k) - \bar{y}_1\right)^2 = \frac{1}{n}\sum_{k=1}^{n}\left(\boldsymbol{w}^T\tilde{\boldsymbol{d}}(k) - 0\right)^2 \\
&= \frac{1}{n}\sum_{k=1}^{n}\left(\boldsymbol{w}^T\tilde{\boldsymbol{d}}(k)\right)\left(\boldsymbol{w}^T\tilde{\boldsymbol{d}}(k)\right)^T \\
&= \frac{1}{n}\sum_{k=1}^{n}\boldsymbol{w}^T\left(\boldsymbol{d}(k) - \bar{\boldsymbol{d}}\right)\left(\boldsymbol{d}(k) - \bar{\boldsymbol{d}}\right)^T\boldsymbol{w} \\
&= \boldsymbol{w}^T\left[\frac{1}{n}\sum_{k=1}^{n}\left(\boldsymbol{d}(k) - \bar{\boldsymbol{d}}\right)\left(\boldsymbol{d}(k) - \bar{\boldsymbol{d}}\right)^T\right]\boldsymbol{w} \\
&= \boldsymbol{w}^T S\boldsymbol{w}
\end{aligned}
$$

18. Statistic data processing

- S is a symmetrical matrix, and thus is diagonalized as follows (see page 139):

$$S = T\mathit{\Lambda}T^{-1},\ \mathit{\Lambda} = T^{-1}ST,$$

$T = \begin{bmatrix} \boldsymbol{t}_1 & \dots & \boldsymbol{t}_m \end{bmatrix}$: a matrix consisting of eigen vectors of S,

$$T^{-1} = T^T,$$

$$\therefore T^T T = \begin{bmatrix} {\boldsymbol{t}_1}^T \\ \vdots \\ {\boldsymbol{t}_m}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{t}_1 & \dots & \boldsymbol{t}_m \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix},$$

$$ {\boldsymbol{t}_i}^T \boldsymbol{t}_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

$$\mathit{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \lambda_m \end{bmatrix},\ \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m.$$

18. Statistic data processing

Define $\boldsymbol{u}$ as $\boldsymbol{w} = T\boldsymbol{u}$, then we have,

$$\boldsymbol{w}^T S\boldsymbol{w} = (T\boldsymbol{u})^T S(T\boldsymbol{u}) = \boldsymbol{u}^T T^T ST\boldsymbol{u} = \boldsymbol{u}^T T^{-1} ST\boldsymbol{u} = \boldsymbol{u}^T \mathit{\Lambda}\boldsymbol{u}$$

$$= \begin{bmatrix} u_1 & \dots & u_m \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \lambda_m \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} = \sum_{i=1}^{m} \lambda_i {u_i}^2,$$

$$\boldsymbol{w}^T \boldsymbol{w} = (T\boldsymbol{u})^T (T\boldsymbol{u}) = \boldsymbol{u}^T T^T T\boldsymbol{u} = \boldsymbol{u}T^{-1}T\boldsymbol{u} = \boldsymbol{u}^T \boldsymbol{u} = 1.$$

So, the problem is now to find $\boldsymbol{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}$ that maximizes

$\boldsymbol{u}^T \mathit{\Lambda}\boldsymbol{u} = \sum_{i=1}^{m} \lambda_i {u_i}^2$ satisfying the condition $\boldsymbol{u}^T \boldsymbol{u} = 1$.

18. Statistic data processing

The solution is, considering $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$,

$$\boldsymbol{u}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \boldsymbol{w}_1 = T\boldsymbol{u}_1 = \boldsymbol{t}_1.$$

The eigen vector associated with
the maximum eigen value λ_1.

The maximum value of $\boldsymbol{w}^T S \boldsymbol{w} = \boldsymbol{u}^T \varLambda \boldsymbol{u}$ is λ_1.

18. Statistic data processing

(Another solution)

- The value of $\boldsymbol{w}$ that maximizes $\boldsymbol{w}^T S \boldsymbol{w}$ under the condition $\boldsymbol{w}^T \boldsymbol{w} = 1$ is specified by the condition:

$$\frac{\partial}{\partial \boldsymbol{w}} \left[\boldsymbol{w}^T S \boldsymbol{w} - v \left(\boldsymbol{w}^T \boldsymbol{w} - 1 \right) \right] = \boldsymbol{0},$$

  using the Lagrange multiplier
- Note that S is a symmetrical matrix, and then we have

$$\frac{\partial}{\partial \boldsymbol{w}} \left[\boldsymbol{w}^T S \boldsymbol{w} - v \left(\boldsymbol{w}^T \boldsymbol{w} - 1 \right) \right] = 2S\boldsymbol{w} - 2v\boldsymbol{w} = \boldsymbol{0}.$$

  Therefore $S\boldsymbol{w} = v\boldsymbol{w}$ is obtained, which implies that v is an eigen value of S and that $\boldsymbol{w}$ is its associated eigen vector
- Moreover, we are maximizing $\boldsymbol{w}^T S \boldsymbol{w} = \boldsymbol{w}^T (v\boldsymbol{w}) = v\boldsymbol{w}^T \boldsymbol{w} = v$, and thus v must be the maximum eigen value λ_1 of S and $\boldsymbol{w}$ must be its associated eigen vector $\boldsymbol{t}_1$ (see the part explaining the induced norm of matrices)

18. Statistic data processing

- $y_1 = {\boldsymbol{w}_1}^T \tilde{\boldsymbol{x}}$ obtained so far is called the first principal component（第1主成分）
- The second principal component is defined as the component other than the first component that maximizes the variance. The second principal component is assured to be different from the first component by making their corresponding direction vectors orthogonal to each other: ${\boldsymbol{w}_2}^T \boldsymbol{w}_1 = 0$

- The p-th principal component is obtained by maximizing the variance under the condition that its direction vector $\boldsymbol{w}$ is orthogonal to any of $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{p-1}$:

  Maximize $\boldsymbol{w}^T S \boldsymbol{w}$

  subject to $\boldsymbol{w}^T \boldsymbol{w}_i = 0, \; i = 1, \ldots, p-1, \; \boldsymbol{w}^T \boldsymbol{w} = 1.$

18. Statistic data processing

The problem of finding the second principal component is expressed as follows using the diagonalized matrix:

Under the following conditions:

$$\boldsymbol{u}^T\boldsymbol{u}_1 = \begin{bmatrix}u_1 & \dots & u_m\end{bmatrix}\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix} = u_1 = 0,\ \boldsymbol{u}^T\boldsymbol{u} = 1,$$

maximize $\boldsymbol{u}^T\mathit{\Lambda}\boldsymbol{u} = \sum_{i=1}^{m}\lambda_i {u_i}^2 = \sum_{i=2}^{m}\lambda_i {u_i}^2$.

Solution : $\boldsymbol{u}_2 = \begin{bmatrix}0\\1\\0\\\vdots\\0\end{bmatrix}$, and thus $\boldsymbol{w}_2 = T\boldsymbol{u}_2 = \boldsymbol{t}_2$.

The eigen vector associated with the second largest eigen value λ_2 .

18. Statistic data processing

Similar discussions lead us to:

Solution: $\boldsymbol{u}_p = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$,

p-th entry

$$\boldsymbol{w}_p = T\boldsymbol{u}_p = \boldsymbol{t}_p,$$

The eigen vector associated with the p-th largest eigen value λ_p .

$$y_p = {\boldsymbol{w}_p}^T \tilde{\boldsymbol{d}},$$

variance of data on y_p coordinate: ${\boldsymbol{w}_p}^T S \boldsymbol{w}_p = \lambda_p$.

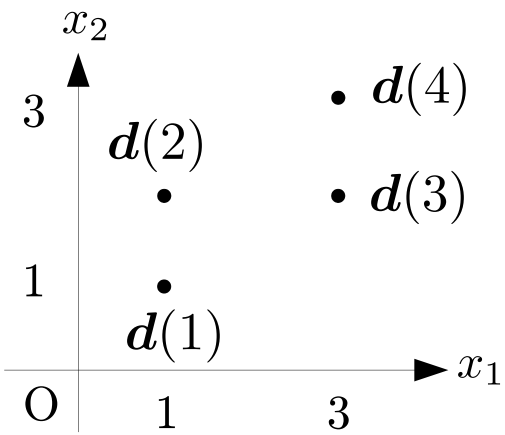
18. Statistic data processing

- Contribution of each principal component is expressed by variance ${\boldsymbol{w}_p}^T S \boldsymbol{w}_p = \lambda_p$
- Using only principal components with significant contributions, the information in the original data can be approximately expressed by lower dimensional data

18. Statistic data processing

- Example
  - Data $\boldsymbol{d}(1) = \begin{bmatrix}1\\1\end{bmatrix}, \boldsymbol{d}(2) = \begin{bmatrix}1\\2\end{bmatrix},$

    $\boldsymbol{d}(3) = \begin{bmatrix}3\\2\end{bmatrix}, \boldsymbol{d}(4) = \begin{bmatrix}3\\3\end{bmatrix}$



  - Mean values

$$\bar{\boldsymbol{d}} = \begin{bmatrix}2\\2\end{bmatrix}$$

  - Sample covariance matrix

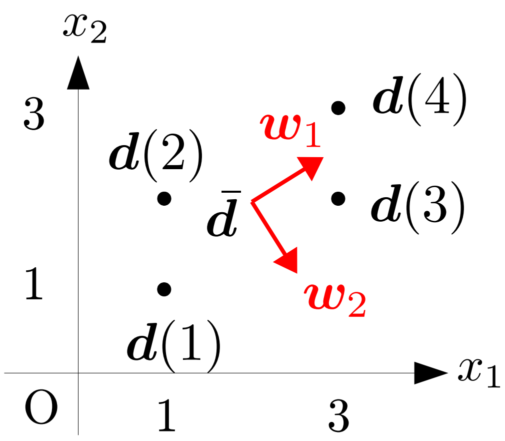
$$S = \frac{1}{4}\begin{bmatrix}1-2 & 1-2 & 3-2 & 3-2\\ 1-2 & 2-2 & 2-2 & 3-2\end{bmatrix}\begin{bmatrix}-1 & -1\\ -1 & 0\\ 1 & 0\\ 1 & 1\end{bmatrix} = \begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}$$

18. Statistic data processing

- Eigen values and eigen vectors of S

$$\lambda_1 = \frac{3+\sqrt{5}}{4} \simeq 1.309, \boldsymbol{t}_1 = \begin{bmatrix} 0.8507 \\ 0.5257 \end{bmatrix} = \boldsymbol{w}_1$$

$$\lambda_2 = \frac{3-\sqrt{5}}{4} \simeq 1.309, \boldsymbol{t}_2 = \begin{bmatrix} 0.5257 \\ -0.8507 \end{bmatrix} = \boldsymbol{w}_2$$


x_2
3
$\boldsymbol{w}_1$
$\boldsymbol{d}(4)$
$\boldsymbol{d}(2)$
$\bar{\boldsymbol{d}}$
$\boldsymbol{d}(3)$
1
$\boldsymbol{w}_2$
$\boldsymbol{d}(1)$
x_1
O
1
3