20211210 itpass セミナー 主成分分析入門

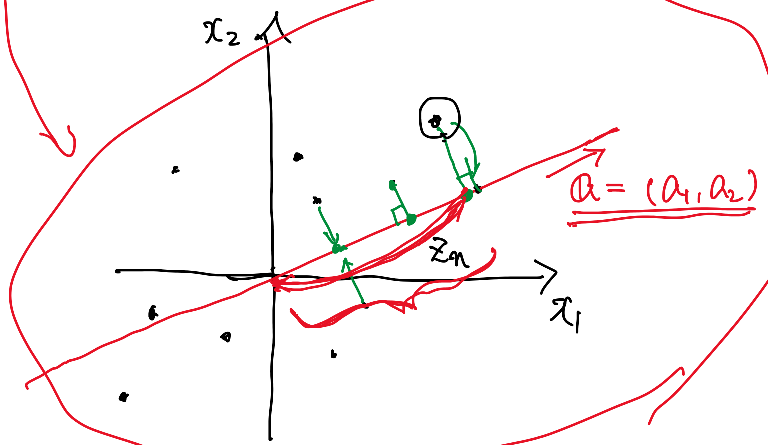
2021年12月10日 金曜日　15:23

2個のデータの組がN個

$(x_{1,1}, x_{1,2}), (x_{2,1}, x_{2,2}), \cdots, (x_{N,1}, x_{N,2})$

平均を引くと

$$\underline{x'_n} = (x'_{n,1}, x'_{n,2}) = (x_{n,1} - \bar{x}_1, x_{n,2} - \bar{x}_2) = \left(x_{n,1} - \frac{1}{N}\sum_{i=1}^{N} x_{i,1},\ x_{n,2} - \frac{1}{N}\sum_{i=1}^{N} x_{i,2}\right)$$



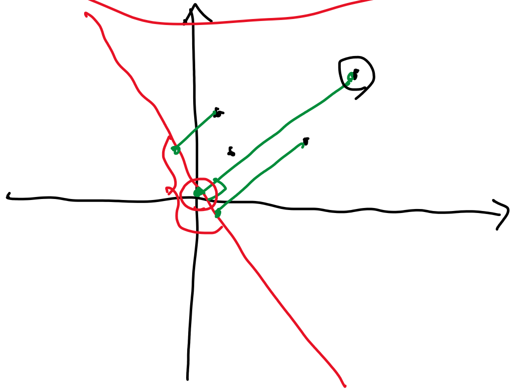
$$z_n = x'_n \cdot a = x'_{n,1} a_1 + x'_{n,2} a_2$$

したがってやることは、

z_n^2 がもっともばらつく a を探す

$|z_n|$

z_n の分散が最も大きい a を探す。



$$\sigma_z^2 = \frac{1}{N-1}\sum_{n=1}^{N}(z_n - \bar{z})^2 = \frac{1}{N-1}\sum_{n=1}^{N}\{(x'_n - \underbrace{\bar{x}'}_{=0}) \cdot a\}^2 = \frac{1}{N-1}\sum_{n=1}^{N}(x'_n \cdot a)^2$$

$$\frac{\partial \sigma_z^2}{\partial a_1} = \frac{\partial \sigma_z^2}{\partial a_2} = 0, \quad a_1^2 + a_2^2 = 1$$

ラグランジュの未定乗数法を用いると解ける。

定数 λ を用いて

$$F(a, \lambda) = \sigma_z^2 - \lambda(a_1^2 + a_2^2 - 1)$$

を定義して

$\partial F \quad \partial F \quad \partial F$

$$\boxed{\frac{\partial F}{\partial a_1} = 0 ,\quad \frac{\partial F}{\partial a_2} = 0 ,\quad \frac{\partial F}{\partial \lambda} = 0}$$

$$\frac{\partial F}{\partial a_1} = 2 \cdot \frac{1}{N-1} \sum_{n=1}^{N} (x'_n \cdot a)\, x'_{n,1} - 2\lambda a_1$$

$$= 2 \cdot \frac{1}{N-1} \sum_{n=1}^{N} (x'_{n,1} \cdot a_1 + x'_{n,2} a_2)\, x'_{n,1} - 2\lambda a_1$$

$$= \underbrace{2 \cdot \frac{1}{N-1} \sum_{n=1}^{N} (x'_{n,1})^2}_{x'_{n,1}\text{ の分散 } \sigma_{x'_1}^2} \cdot a_1 + \underbrace{2 \cdot \frac{1}{N-1} \sum_{n=1}^{N} x'_{n,1} \cdot x'_{n,2}}_{x'_{n1}\text{ と } x'_{n2}\text{ の共分散 } C_{x'_1 x'_2}} \cdot a_2 - 2\lambda a_1$$

$$\boxed{\begin{aligned} &= 2\sigma_{x'_1}^2 \cdot a_1 + 2 C_{x'_1 x'_2} \cdot a_2 - 2\lambda a_1 = 0 \\ \frac{\partial F}{\partial a_1} &= 2 C_{x'_1 x'_2}\, a_1 + 2\sigma_{x'_2}^2\, a_2 - 2\lambda a_2 = 0 \end{aligned}}$$

$$\frac{\partial F}{\partial \lambda} = -(a_1^2 + a_2^2 - 1) = 0$$

$$\begin{pmatrix} \sigma_{x'_1}^2 & C_{x'_1 x'_2} \\ C_{x'_1 x'_2} & \sigma_{x'_2}^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \lambda \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

$$\boxed{\mathbb{V}\, a = \lambda a}$$

分散・共分散行列

$\mathbb{V}$ に対する固有値問題