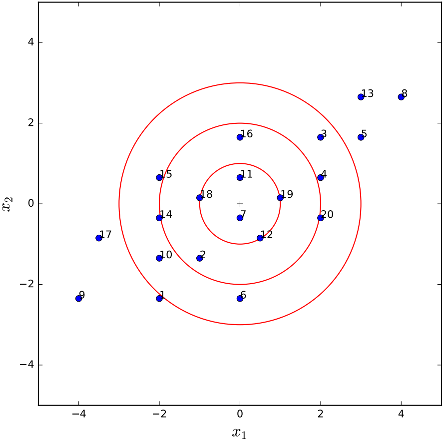
Mahalanobis távolság

De Maesschalck, *Chemom. Intell. Lab. Syst.* **50**, 1 (2000)

Király Péter

2019. május 16.

Grafikus interpretáció



Adatok: $n = 20$ megfigyelés

x_1, x_2 változók

ED (euclidean distance, euklideszi távolság)

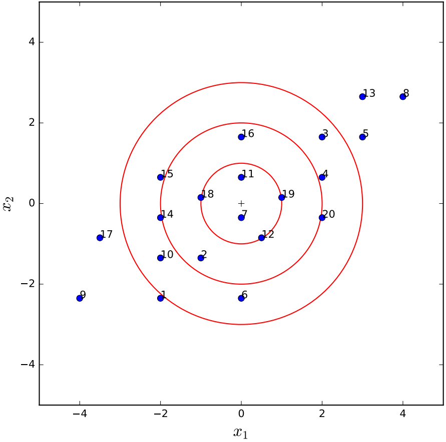
$$\mathrm{ED}_i = \sqrt{(x_{i1} - \bar{x}_1)^2 + (x_{i2} - \bar{x}_2)^2}$$

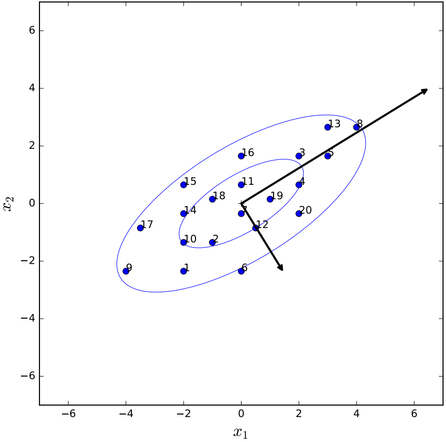
$$\mathrm{ED}_i = \sqrt{(x_{i1} - \bar{x}_1)^2 + (x_{i2} - \bar{x}_2)^2}$$

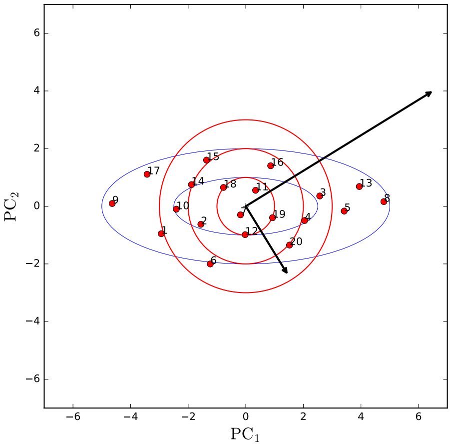
MD (Mahalanobis distance, Mahalanobis távolság)

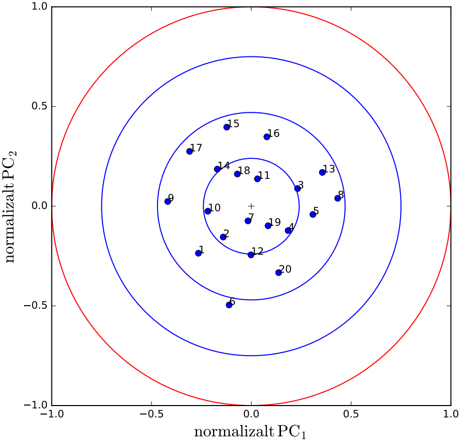
$$\mathrm{MD}_i = \sqrt{\left(\frac{x_{i1}-\bar{x}_1}{\sigma_1}\right)^2 + \left[\left\{\left(\frac{x_{i2}-\bar{x}_2}{\sigma_2}\right) - \rho_{12}\left(\frac{x_{i1}-\bar{x}_1}{\sigma_1}\right)\right\}\frac{1}{\sqrt{1-\rho_{12}^2}}\right]^2}$$

ha $\rho_{12} = 0 : \mathrm{ED}_i = \mathrm{MD}_i$


4
2
0
−2
−4
−4
−2
0
2
4
x_1
x_2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20


6
4
2
0
−2
−4
−6
−6
−4
−2
0
2
4
6
x_1
x_2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20


PC_1
PC_2


normalizalt PC_1
normalizalt PC_2

$$\mathrm{MD}_i^{\mathrm{o}} = \sqrt{(\boldsymbol{x}_i - \bar{\boldsymbol{x}})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^{\mathsf{T}}}$$

$$\mathrm{ED}_i^{\mathrm{o}} = \sqrt{(\boldsymbol{x}_i - \bar{\boldsymbol{x}})\boldsymbol{I}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^{\mathsf{T}}}$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{\frac{-(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)}$$

$$f(\boldsymbol{x}_i) = \frac{1}{(2\pi)^{p/2}\sqrt{|\boldsymbol{\Sigma}|}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})^{\mathsf{T}}}$$

$$\mathrm{ED}_i^{\mathrm{o}} = \mathrm{ED}_i^{\mathrm{t}} = \sqrt{t_i t_i^{\mathsf{T}}} \neq \mathrm{ED}_i^{\mathrm{u}}$$

$$\mathrm{MD}_i^{\mathrm{o}} = \mathrm{MD}_i^{\mathrm{t}} = \mathrm{MD}_i^{\mathrm{u}} = \sqrt{n-1} \cdot \mathrm{ED}_i^{\mathrm{u}}$$

- igaz, ha ∀ PC-t felhasználunk a távolságok számolásához

Kiugró értékek azonosítása

$\text{MD}^2 \Leftrightarrow \chi^2(p-1)$

modellérzékenység:

$h_i = \frac{1}{n} + \frac{(\text{MD}_i^{\text{o}})^2}{n-1} \Leftrightarrow \frac{2g}{n}$