# Coclustering—a useful tool for chemometrics

Bro et al., *J. Chemometrics* **26**, 256 (2012)

Király Péter

2019. november 27.

klaszterezés célfüggvény: $\sum_j \left| \boldsymbol{x}_j - \boldsymbol{\mu}_{k*(j)} \right|^2$

$$\min \left\| \boldsymbol{X} - \boldsymbol{M}\boldsymbol{B}^{\mathsf{T}} \right\|_F^2, B_{jk} \in \{0,1\}$$

$$\min \left\| \boldsymbol{X} - \left( \boldsymbol{\mu}_1 \boldsymbol{b}_1^{\mathsf{T}} + \ldots + \boldsymbol{\mu}_k \boldsymbol{b}_k^{\mathsf{T}} \right) \right\|_F^2$$

hard: $\boldsymbol{x}_j \in \boldsymbol{C} = \boldsymbol{\mu}\boldsymbol{b}^{\mathsf{T}}$, soft: $\boldsymbol{x}_j \in \{\boldsymbol{C}\}$

koklaszterezés: soft, $\boldsymbol{X} \geq 0$

$$\boldsymbol{G} = \boldsymbol{a}\boldsymbol{b}^{\mathsf{T}}, \boldsymbol{a}, \boldsymbol{b} \in \{0, ?\}$$

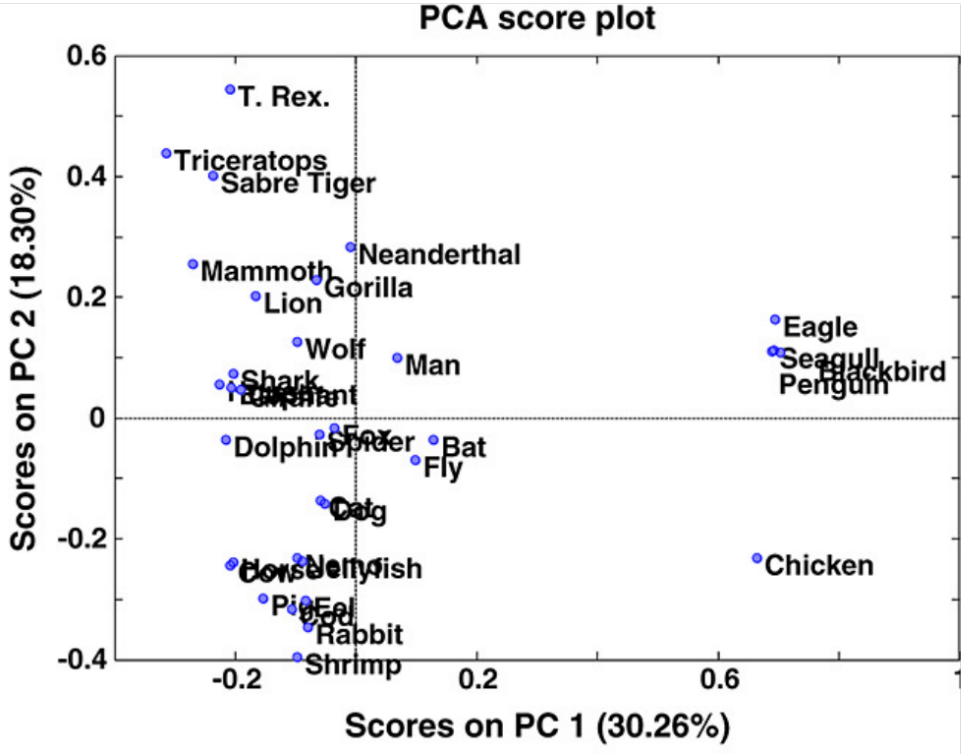
$$\min \left\| \boldsymbol{X} - \boldsymbol{A}\boldsymbol{B}^{\mathsf{T}} \right\|_F^2 + \lambda \textstyle\sum_{i,k} |A_{ik}| + \lambda \sum_{j,k} |B_{jk}|$$

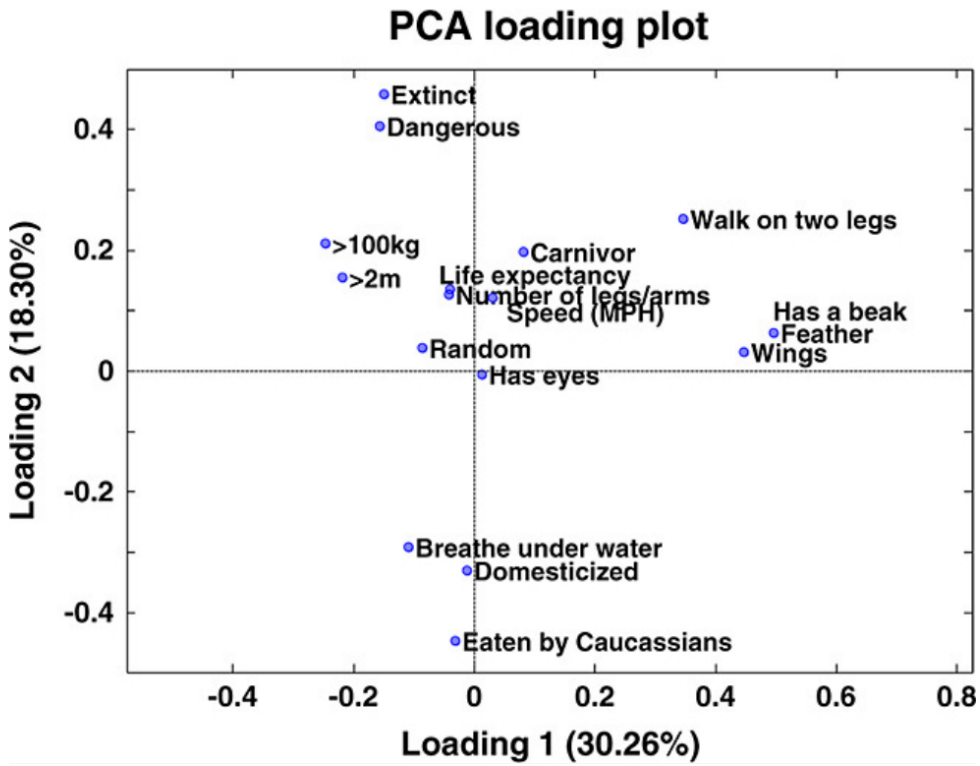
constrained outer product decomposition of $\boldsymbol{X}$,
with sparsity on the latent factors of the bilinear model

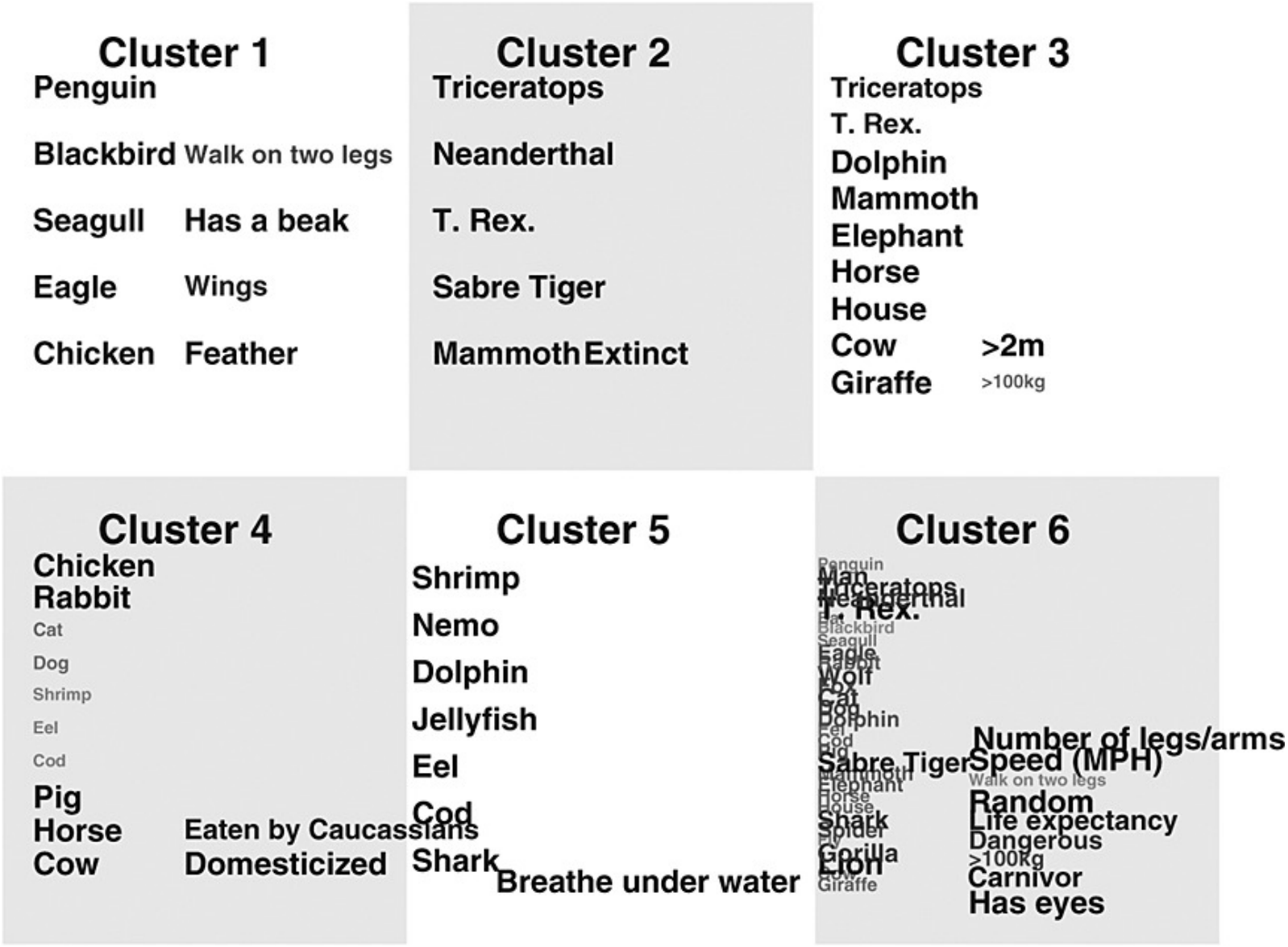
# Játékpélda

| **minták** | | **változók** |
|---|---|---|
| Giraffe | Nemo | Has eyes |
| Cow | Shrimp | Number of legs/arms |
| Lion | Dog | Carnivore |
| Gorilla | Cat | Feather |
| Fly | Fox | Wings |
| Spider | Wolf | Domesticized |
| Shark | Rabbit | Eaten by Caucasians |
| House | Chicken | >100 kg |
| Horse | Eagle | >2 m |
| Elephant | Seagull | Breathe under water |
| Mammoth | Blackbird | Extinct |
| Sabre Tiger | Bat | Dangerous |
| Pig | T. Rex. | Life expectancy |
| Cod | Neanderthal | Random |
| Eel | Triceratops | Has a beak |
| Jellyfish | Man | Walk on two legs |
| Dolphin | Penguin | Speed (MPH) |

| | Has eyes | Number of legs/arms | Carnivore | Feather | Wings | Domesticized 0 | Eaten by Caucasians | >100 kg | >2 m | Breathe under water | Extinct | Dangerous | Life expectancy | Random | Has a beak | Walk on two legs | Speed |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Giraffe | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 30 | 1 | 0 | 0 | 32 |
| Cow | 1 | 4 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 15 | 3 | 0 | 0 | 30 |
| Lion | 1 | 4 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 15 | 6 | 0 | 0 | 50 |
| Gorilla | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 30 | 2 | 0 | 1 | 25 |
| Fly | 1 | 6 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1 | 7 | 0 | 0 | 5 |
| Spider | 1 | 8 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 8 | 0 | 0 | 1 |
| Shark | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 50 | 4 | 0 | 0 | 30 |
| House | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 100 | 9 | 0 | 0 | 0 |
| Horse | 1 | 4 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 15 | 2 | 0 | 0 | 40 |
| Elephant | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 35 | 6 | 0 | 0 | 25 |
| Mammoth | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 35 | 5 | 0 | 0 | 25 |
| Sabre Tiger | 1 | 4 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 15 | 7 | 0 | 0 | 40 |
| Pig | 1 | 4 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 25 | 8 | 0 | 0 | 11 |
| Cod | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 40 | 9 | 0 | 0 | 2 |
| Eel | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 55 | 1 | 0 | 0 | 20 |
| Jellyfish | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0.7 | 3 | 0 | 0 | 1 |
| Dolphin | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 30 | 5 | 0 | 0 | 35 |
| Nemo | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 6 | 0 | 0 | 4 |
| Shrimp | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 2 | 0 | 0 | 0.5 |
| Dog | 1 | 4 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 13 | 8 | 0 | 0 | 35 |
| Cat | 1 | 4 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 25 | 9 | 0 | 0 | 30 |
| Fox | 1 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 4 | 0 | 0 | 42 |
| Wolf | 1 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 18 | 3 | 0 | 0 | 25 |
| Rabbit | 1 | 4 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 9 | 8 | 0 | 0 | 35 |
| Chicken | 1 | 2 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 15 | 1 | 1 | 1 | 9 |
| Eagle | 1 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 55 | 3 | 1 | 1 | 60 |
| Seagull | 1 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 6 | 1 | 1 | 25 |
| Blackbird | 1 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 18 | 0 | 1 | 1 | 25 |
| Bat | 1 | 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 24 | 4 | 0 | 0 | 8 |
| T. Rex. | 1 | 4 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 40 | 9 | 0 | 1 | 25 |
| Neanderthal | 1 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 50 | 8 | 0 | 1 | 18 |
| Triceratops | 1 | 4 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 30 | 5 | 0 | 0 | 10 |
| Man | 1 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 80 | 2 | 0 | 1 | 28 |
| Penguin | 1 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 4 | 1 | 1 | 25 |


PCA score plot
T. Rex.
Triceratops
Sabre Tiger
Neanderthal
Mammoth
Gorilla
Lion
Eagle
Wolf
Man
Seagull
Blackbird
Penguin
Shark
Dolphin
Bat
Fly
Chicken
Rabbit
Shrimp
Scores on PC 2 (18.30%)
Scores on PC 1 (30.26%)
0.6
0.4
0.2
0
-0.2
-0.4
-0.2
0.2
0.6


PCA loading plot
Extinct
Dangerous
Walk on two legs
>100kg
Carnivor
>2m
Life expectancy
Number of legs/arms
Speed (MPH)
Has a beak
Feather
Random
Wings
Has eyes
Breathe under water
Domesticized
Eaten by Caucassians
Loading 2 (18.30%)
Loading 1 (30.26%)
0.4
0.2
0
-0.2
-0.4
-0.4
-0.2
0
0.2
0.4
0.6
0.8




6 koklaszter modell

## Cluster 1

Penguin
Blackbird
Seagull
Eagle
Chicken

Random
Walk on two legs
Has a beak
Wings
Feather

## Cluster 2

Shrimp
Nemo
Dolphin
Jellyfish
Eel
Cod
Shark

Breathe under water
Eaten by Caucasians

## Cluster 3

Triceratops
Neanderthal
T. Rex.
Wolf
Sabre Tiger
Mammoth
Shark
Gorilla
Lion

Dangerous
Extinct
>100kg

## Cluster 4

Chicken
Rabbit
Cat
Dog
Shrimp
Cod
Pig
Horse
Cow

Eaten by Caucasians
Domesticized

## Cluster 5

Bat
Fly

Random
Wings

## Cluster 6

Triceratops
T. Rex.
Dolphin
Mammoth
Elephant
Horse
House
Cow
Giraffe

>2m
>100kg

## Cluster 7


7 koklaszter modell

- spektrális adatok kezelése problémás ($\lambda$)
- túl sok irreleváns infó problémás
- mellékfürtelemzés
- kód: `http://www.models.life.ku.dk/` $\rightarrow$ SMR
- "chemometric applications in biology" $\stackrel{?}{=}$ biometrics
- általánosítás: Seriation, the method out of a chemist's mind (TG)