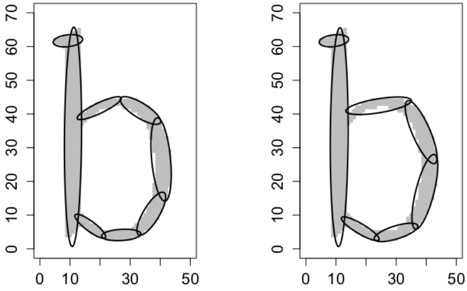
Active Function Cross-Entropy Clustering

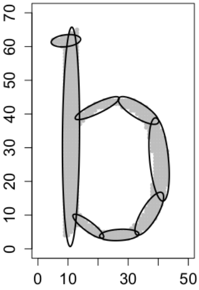
P. Spurek J. Tabor P. Markowicz

Będlewo 2015

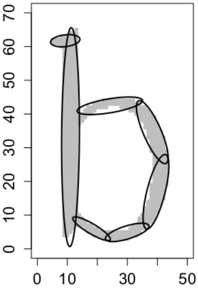


(a) GMM, $k = 8$.

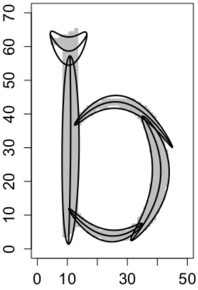
(b) CEC, $k = 7$.



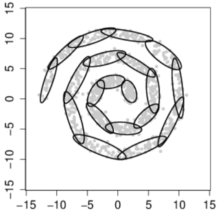
(c) GMM $k = 8$.



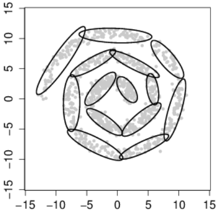
(d) CEC, $k = 7$.



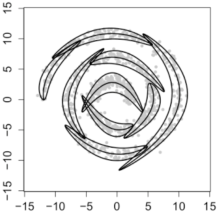
(e) afCEC, $k = 5$.



(f) GMM, $k = 20$.



(g) CEC, $k = 13$.



(h) afCEC, $k = 9$.

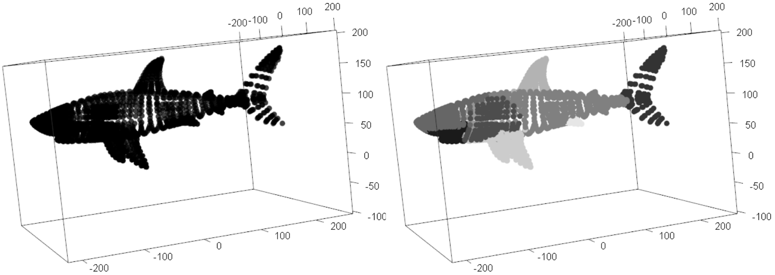


Figure: 3D experiment.

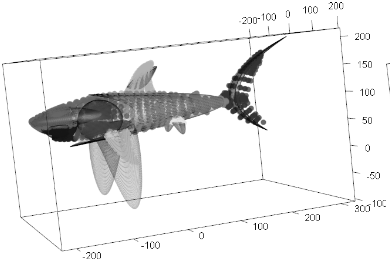


Figure: 3D experiment.

GMM aims at finding $p_1, \ldots, p_k \geq 0$, $\sum_{i=1}^{k} p_i = 1$ and $f_1, \ldots, f_k$ Gaussian densities such that the convex combination

$$f := p_1 f_1 + \ldots + p_k f_k$$

optimally approximates the scatter of our data $X = \{x_1, \ldots, x_n\}$ with respect to MLE cost function

$$\mathrm{MLE}(f, X) := -\sum_{l=1}^{n} \ln(p_1 f_1(x_l) + \ldots + p_n f_n(x_l)).$$

A goal of CEC is to minimize the cost function, which is a minor modification:

$$\mathrm{CEC}(f, X) := -\sum_{l=1}^{n} \ln(\max(p_1 f_1(x_l), \ldots, p_n f_n(x_l))).$$

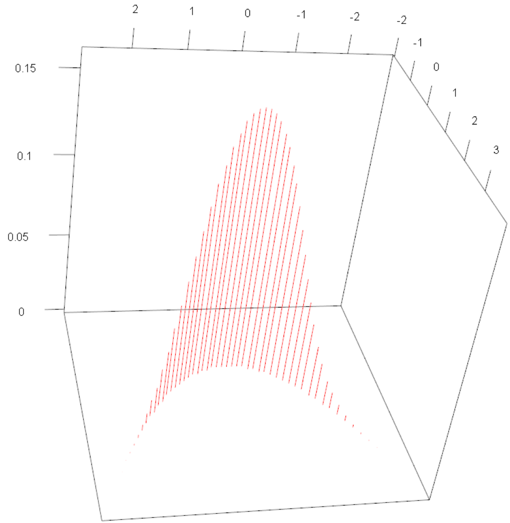


Figure: One dimensional Gaussian density on a curve.

B. Zhang, C. Zhang, X. Yi.
Active curve axis gaussian mixture models.
Pattern recognition, 38, 2351–2362, 2005.

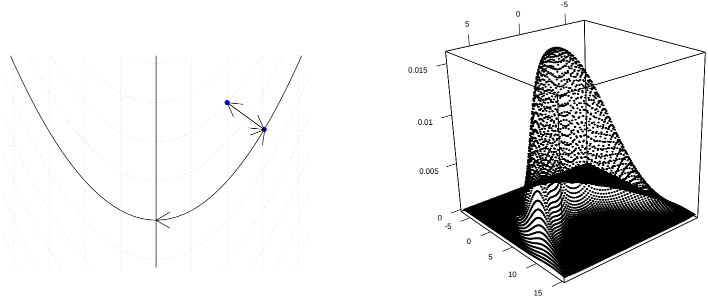


Figure: Two dimensional Gaussian density on a curve.

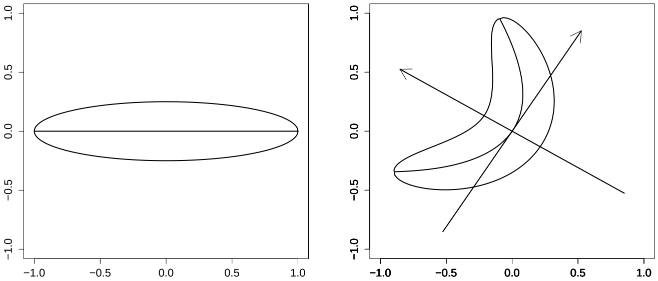


Figure: Comparison of ellipses generated by classical and modified Gaussian densities.

Active curve axis gaussian mixture models

- B. Zhang, C. Zhang, X. Yi.
  Active curve axis gaussian mixture models.
  Pattern recognition, 38, 2351–2362, 2005.
- Z. Ju, H. Liu.
  A unified fuzzy framework for human-hand motion recognition.
  IEEE Transactions on Fuzzy Systems, 19, 901–913, 2011.
- Z. Ju, H. Liu.
  Fuzzy gaussian mixture models.
  Pattern Recognition, 45, 1146–1158, 2012.

Active curve axis gaussian mixture models

Advantages

- Model is very intuitive.
- Works nice in practice (on the plane).

Disadvantages

- It is very hard (or even impossible) to give explicit formula for orthogonal projection and arc length for more complicated curves in higher dimensional spaces.
- The generalized Gaussian density of acaGMM is not a density model (no theoretical background).
- The MLE cost function does not necessarily decrease with iterations (problems with stop condition).

Active curve axis gaussian mixture models

Advantages

- Model is very intuitive.
- Works nice in practice (on the plane).

Disadvantages

- It is very hard (or even impossible) to give explicit formula for orthogonal projection and arc length for more complicated curves in higher dimensional spaces.
- The generalized Gaussian density of acaGMM is not a density model (no theoretical background).
- The MLE cost function does not necessarily decrease with iterations (problems with stop condition).

Active curve axis gaussian mixture models

Advantages

- Model is very intuitive.
- Works nice in practice (on the plane).

Disadvantages

- It is very hard (or even impossible) to give explicit formula for orthogonal projection and arc length for more complicated curves in higher dimensional spaces.
- The generalized Gaussian density of acaGMM is not a density model (no theoretical background).
- The MLE cost function does not necessarily decrease with iterations (problems with stop condition).

Active curve axis gaussian mixture models

Advantages

- Model is very intuitive.
- Works nice in practice (on the plane).

Disadvantages

- It is very hard (or even impossible) to give explicit formula for orthogonal projection and arc length for more complicated curves in higher dimensional spaces.
- The generalized Gaussian density of acaGMM is not a density model (no theoretical background).
- The MLE cost function does not necessarily decrease with iterations (problems with stop condition).

Active curve axis gaussian mixture models

Advantages

- Model is very intuitive.
- Works nice in practice (on the plane).

Disadvantages

- It is very hard (or even impossible) to give explicit formula for orthogonal projection and arc length for more complicated curves in higher dimensional spaces.
- The generalized Gaussian density of acaGMM is not a density model (no theoretical background).
- The MLE cost function does not necessarily decrease with iterations (problems with stop condition).

Active curve axis gaussian mixture models

Advantages

- Model is very intuitive.
- Works nice in practice (on the plane).

Disadvantages

- It is very hard (or even impossible) to give explicit formula for orthogonal projection and arc length for more complicated curves in higher dimensional spaces.
- The generalized Gaussian density of acaGMM is not a density model (no theoretical background).
- The MLE cost function does not necessarily decrease with iterations (problems with stop condition).

Active curve axis gaussian mixture models

Advantages

- Model is very intuitive.
- Works nice in practice (on the plane).

Disadvantages

- It is very hard (or even impossible) to give explicit formula for orthogonal projection and arc length for more complicated curves in higher dimensional spaces.
- The generalized Gaussian density of acaGMM is not a density model (no theoretical background).
- The MLE cost function does not necessarily decrease with iterations (problems with stop condition).

Active Function Cross-Entropy Clustering

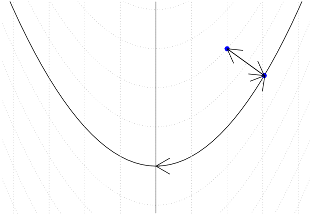


Figure: Difference between acaGMM and afCEC.

The two dimensional Gaussian density for $\mathrm{m}^T = [m_1, m_2]$ and covariance matrix $\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$ is given by following formula

$$N(\mathrm{m}, \Sigma)(\mathrm{x}) = N(m_1, \sigma_1^2)(x_1) N(m_2, \sigma_2^2)(x_2), \tag{1}$$

where in one dimensional case we have

$$N(m, \sigma^2)(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{|x-m|^2}{2\sigma^2}\right) \text{ for } m, \sigma \in \mathbb{R}.$$

1. acaGMM

$$N(\mathrm{m},\Sigma,f)(\mathrm{x}) = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{l_f(p_f(\mathrm{x}),\mathrm{m})^2}{2\sigma_1^2}\right)\cdot\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{\|p_f(\mathrm{x})-\mathrm{x}\|^2}{2\sigma_2^2}\right).$$

2. afCEC

$$N(\mathrm{m},\Sigma,f)([x_1,x_2]) = N(m_1,\sigma_1^2)(x_1)N(m_2,\sigma_2^2)(x_2-f(x_1)).$$

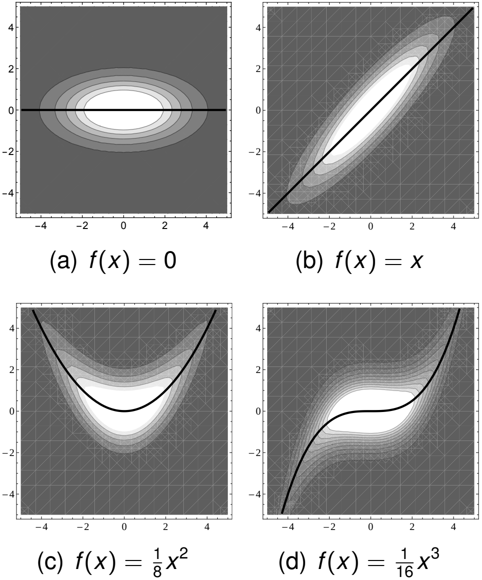
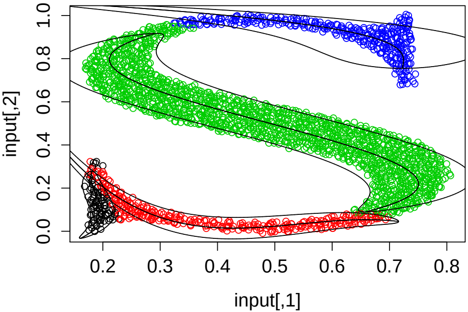


Figure: Level–sets for f–adapted Gaussian Distribution.



(a) $f(x) = a_3x^3 + a_2x^2 + a_1x + a_0$

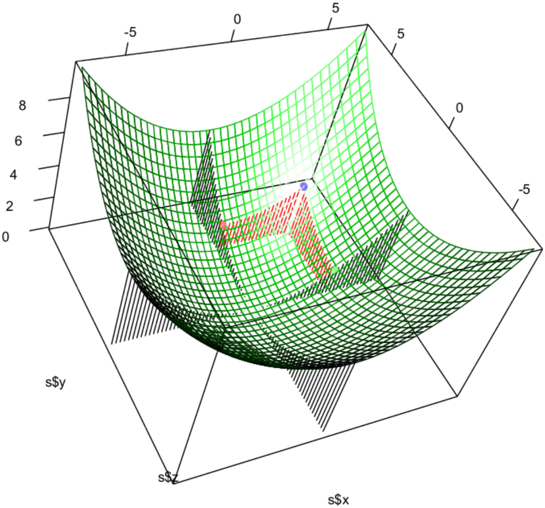
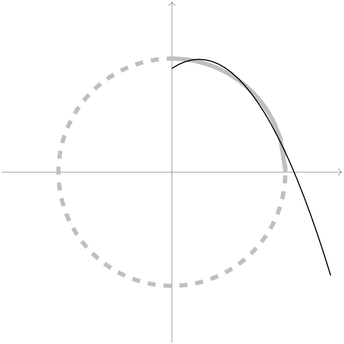
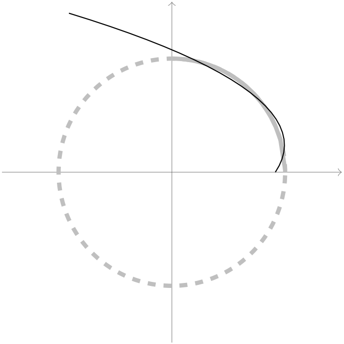


Figure: Example in 3D.



(a) $(x, f(x))$.



(b) $(f(y), y)$.

Figure: Two possible parabola fitting for quadrantal.

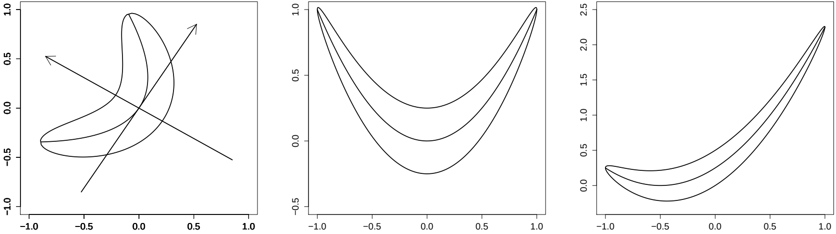
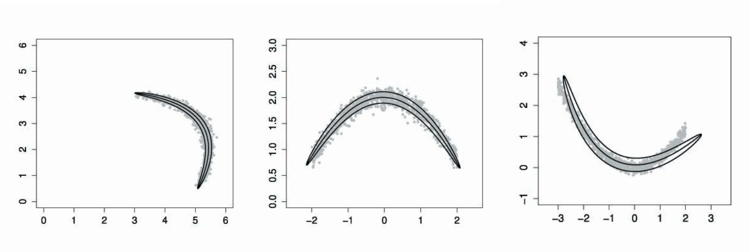
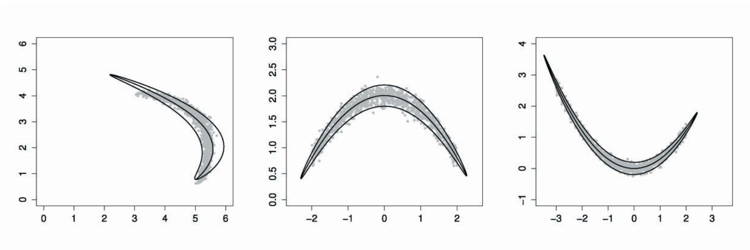


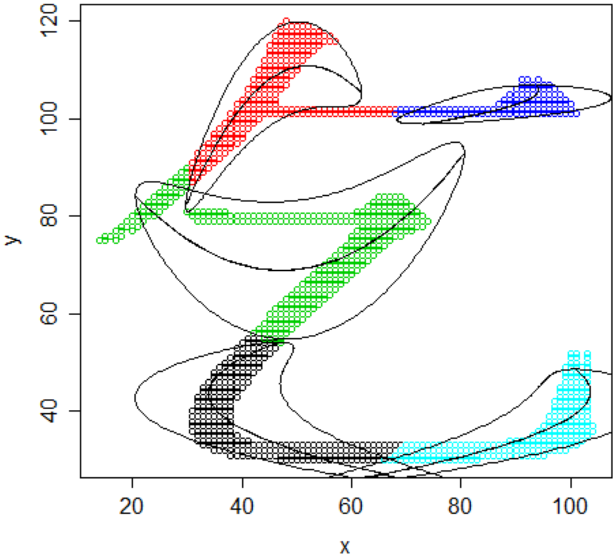
Figure: Examples of ellipses of acaGMM and afCEC.

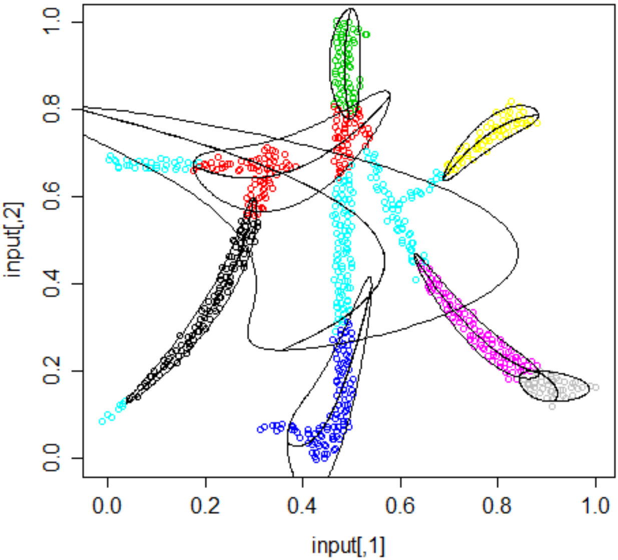


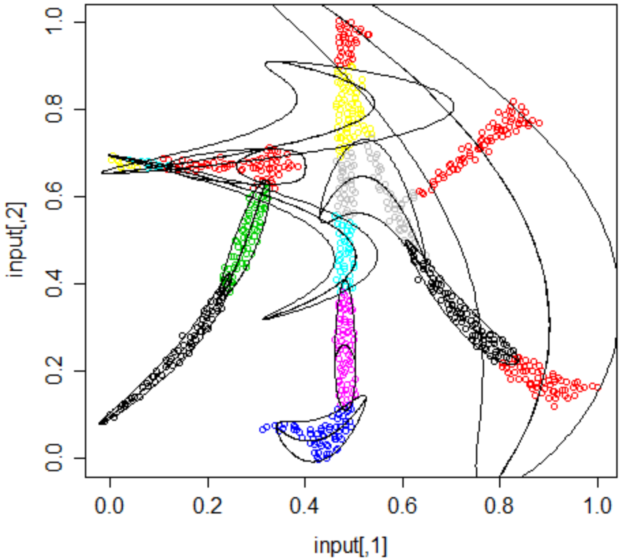
(a) The AcaGMM method.



(b) The afCEC method.


120
100
80
60
40
y
20
40
60
80
100
x

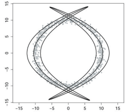

input[,2]
input[,1]
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
input[,2]
0.0
0.2
0.4
0.6
0.8
1.0
input[,1]

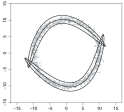
To compare the results we use the standard Bayesian Information Criterion (BIC)

$$BIC = -2LL + k\log(n)$$

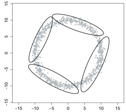
where k is a number of parameters in the model, n is a number of points, and LL is a maximized value of the Log-likelihood function.



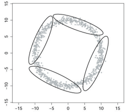
(a) afCEC



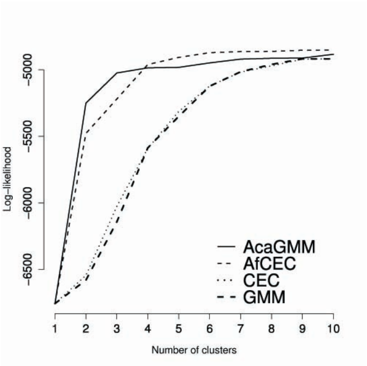
(b) AcaGMM



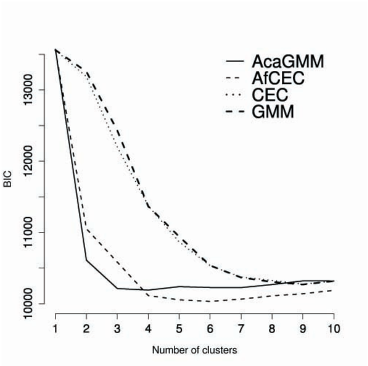
(c) GMM.



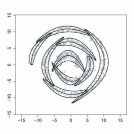
(d) CEC



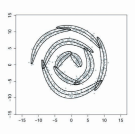
(e) Log-likelihood function.



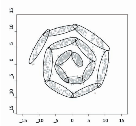
(f) BIC function.



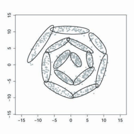
(g) AfCEC



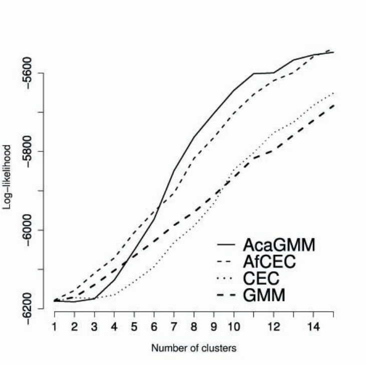
(h) AcaGMM



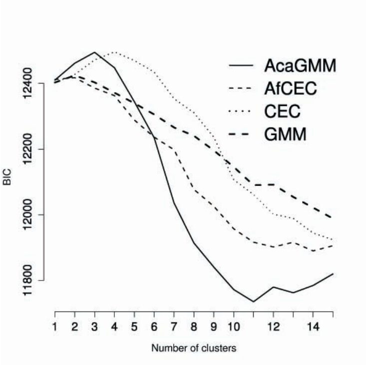
(i) GMM



(j) CEC



(k) Log-likelihood function



(l) BIC function

Thank you for your attention.