Application of Semiparametric Density Estimation to Classification

Lasse Holmström
Department of Mathematical Sciences
University of Oulu

Fabian Hoti
Department of Mathematics and Statistics
University of Helsinki
http://www.rni.helsinki.fi/~fjh/

ICPR04, Cambridge, UK

Application of Semiparametric Density Estimation to Classification

1. Introduction
2. Our method sKDE
3. Classification results
4. Exploratory visualization

Introduction

Statistical Pattern Recognition

- a sample $\{(x_1, j_1), \ldots, (x_n, j_n)\}$ is a realization of i.i.d random vectors $[X_i^T, J_i]^T \sim [X^T, J]^T$

- the goal is to use the sample to estimate the class posteriori probabilities that is

$$P(J = j|X = x) = \frac{P_j f_j(x)}{\sum_{i=1}^{c} P_i f_i(x)},$$

for $j = 1, \ldots, c$.

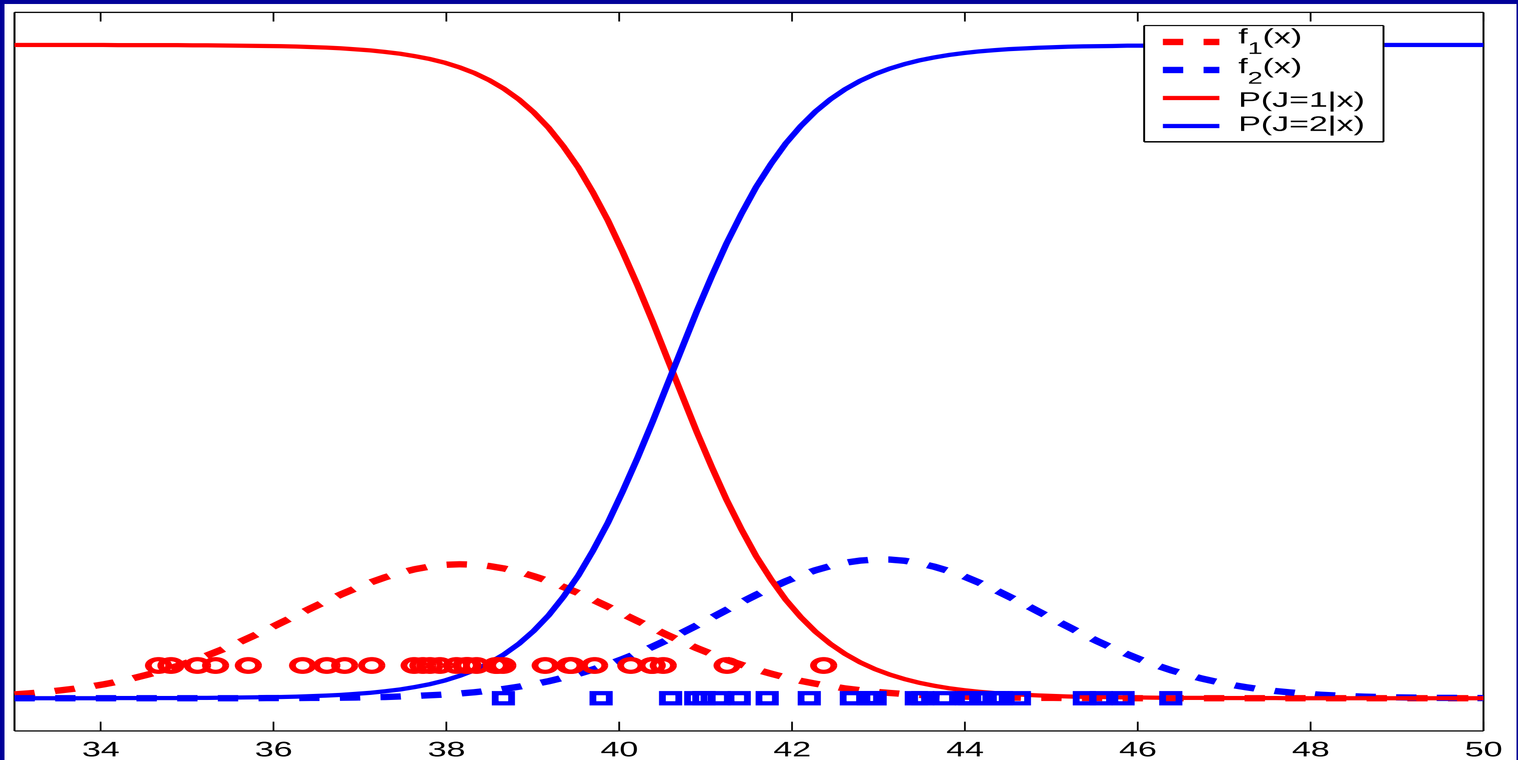
Statistical Pattern Recognition

Bayes classification rule $x \to k$, where k satisfies

$$P(J = k|X = x) > P(J = j|X = x),$$

for all $j \neq k$

A two class example



Nonparametric density estimation

Assume $f(\cdot) \in \mathcal{F}$, where $\mathcal{F}$ is a family of functions with no specific finite parametric form.

Examples:

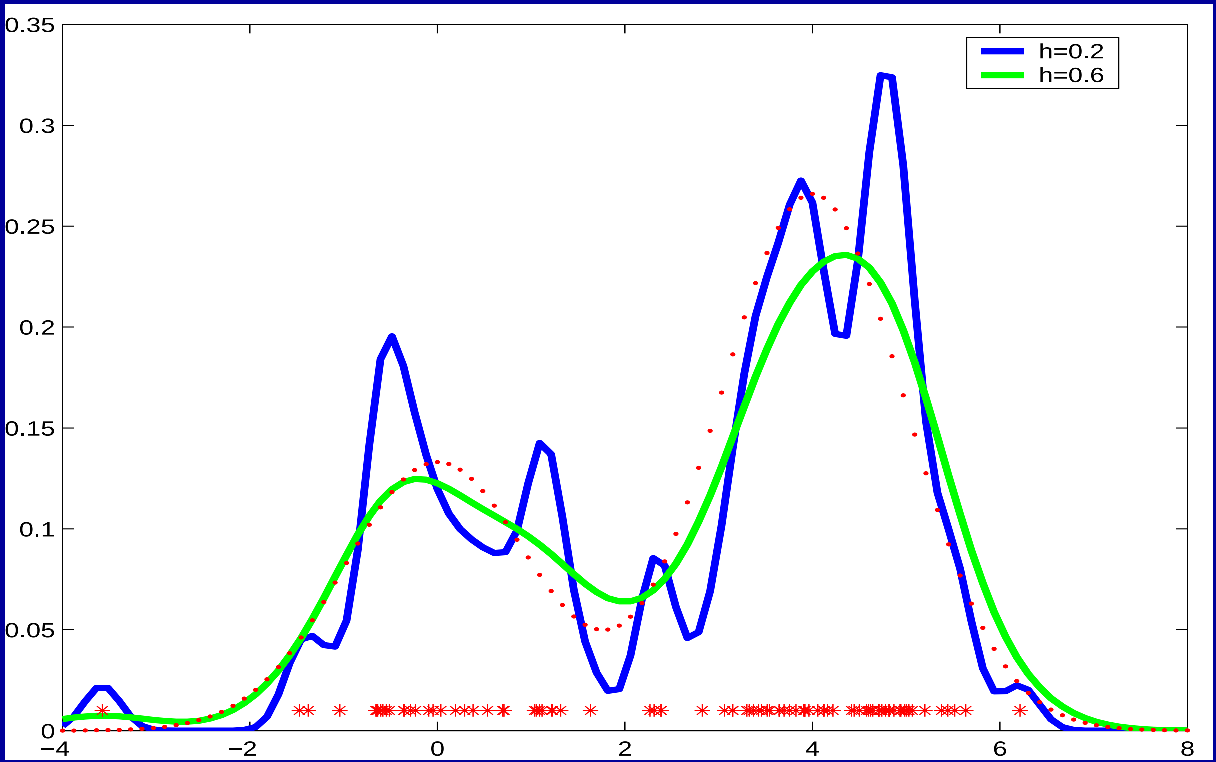
1) the Histogram

2) the Kernel density estimator (KDE)

$$\hat{f}(x) = \frac{1}{n} \sum K_h(x_i - x),$$

where $K_h(x) = \frac{1}{h^d} K(x/h)$ and K is a kernel.

Kernel density estimation



The proposed method

Semiparametric density estimation

The d-dimensional vectors are split as

$$X = (Y, Z), Y \in \mathbb{R}^s, Z \in \mathbb{R}^{d-s}$$

The unknown density function is then factorized

$$f(x) = f_Y(y) f_{Z|Y=y}(z)$$

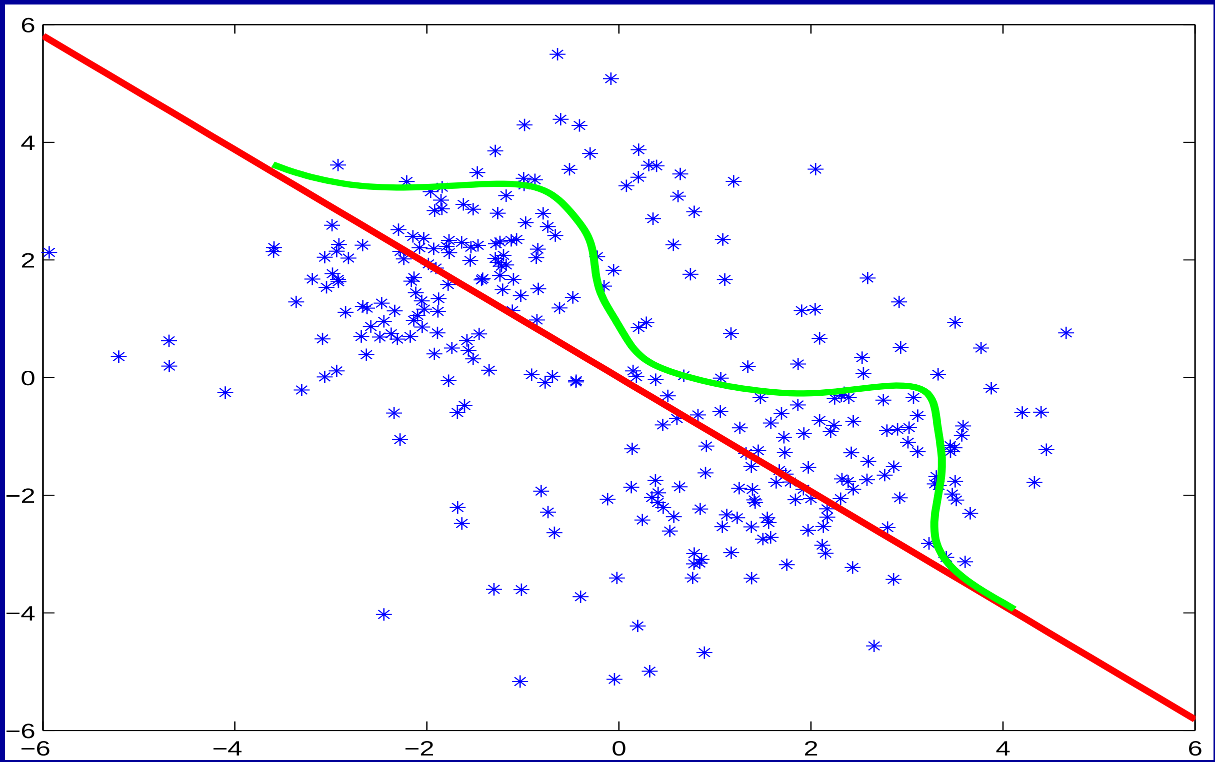
where $Y \sim f_Y$ and $Z|Y = y \sim f_{Z|Y=y}$

Semiparametric density estimation

We estimate f_Y with KDE and we assume $f_{Z|Y=y}(z) \sim N(m(y), C(y))$ where

$$
\begin{aligned}
m(y) &= \mathbb{E}(Z \mid Y = y) \\
C(y) &= \mathbb{E}[(Z - m(y))^T (Z - m(y)) \\
&\quad \mid Y = y]
\end{aligned}
$$

Semiparametric density estimation



Semiparametric density estimation

Find a transformation T such that for $U = T(X) = (V, W)$ we can assume $W|V = v \sim \mathcal{N}(m(v), C(v))$.

Our density estimate is

$$\hat{f}_X(x) = |J_T(x)|\hat{f}_U(T(x)),$$

where $J_T(x)$ is the Jacobian of T.

Sketch of algorithm

1. Transform the data
2. Use cross-validation to choose

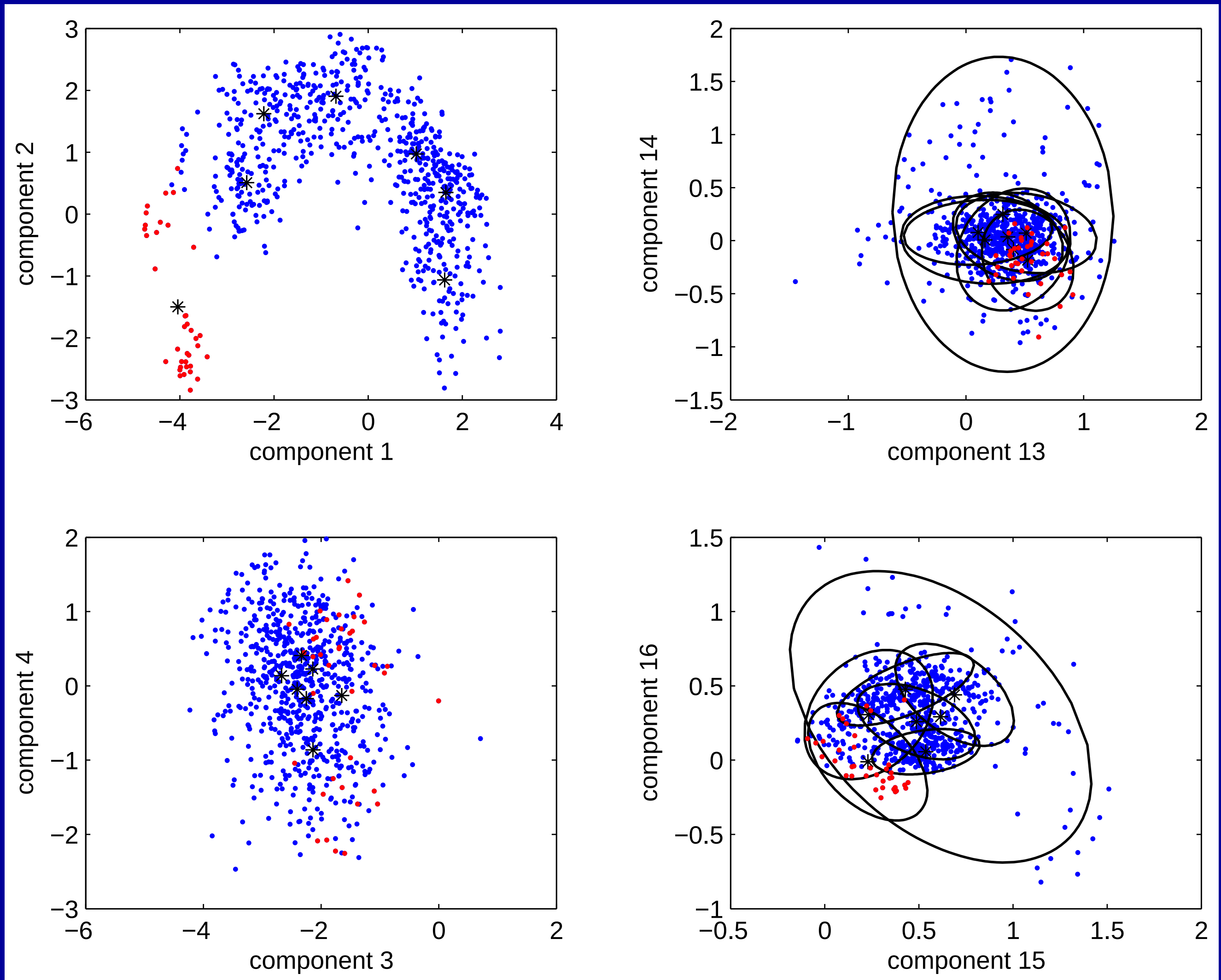
- total dimension d'
- splitting dimension s
- smoothing parameters h_1

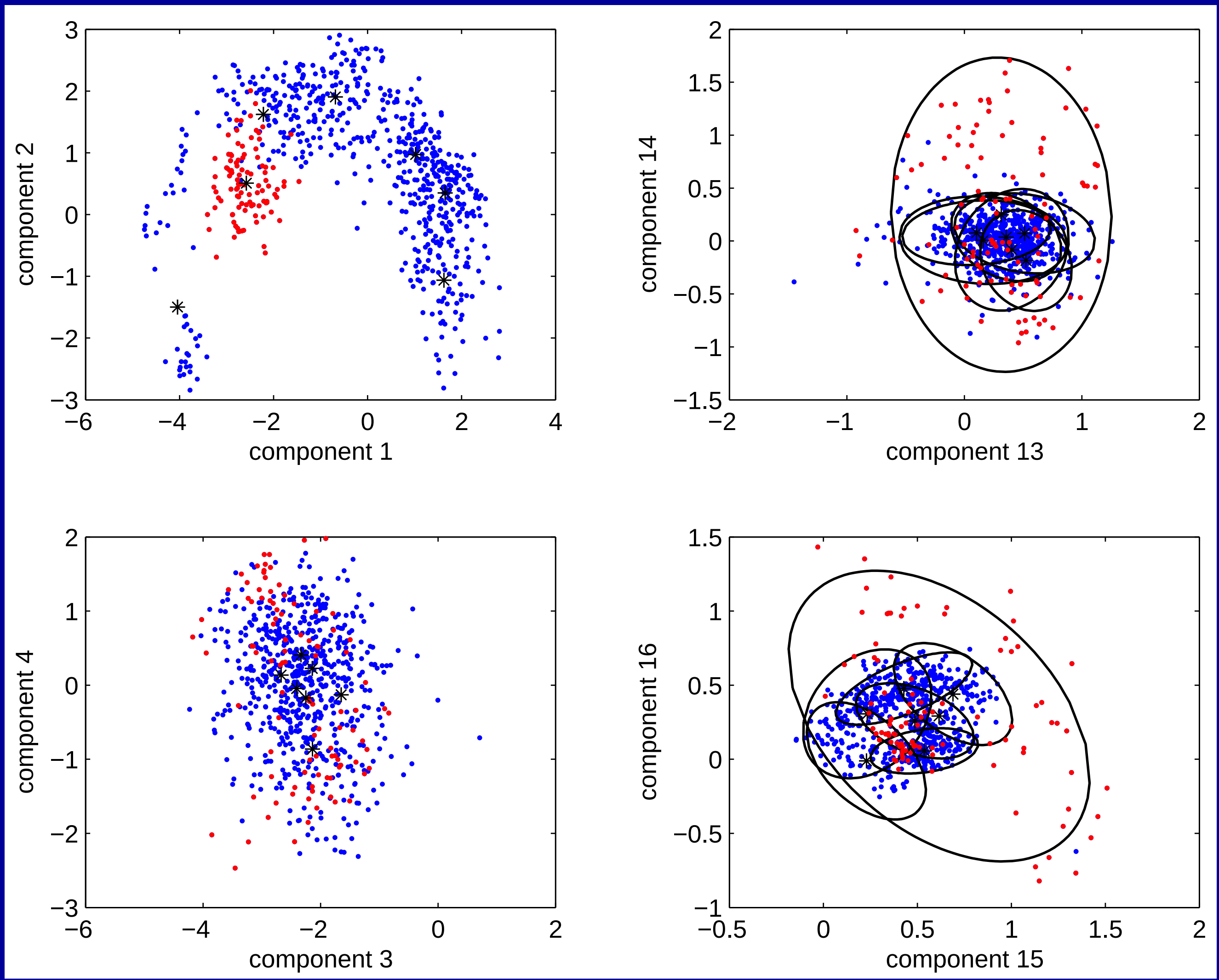
3. construct classifier

Classification Results

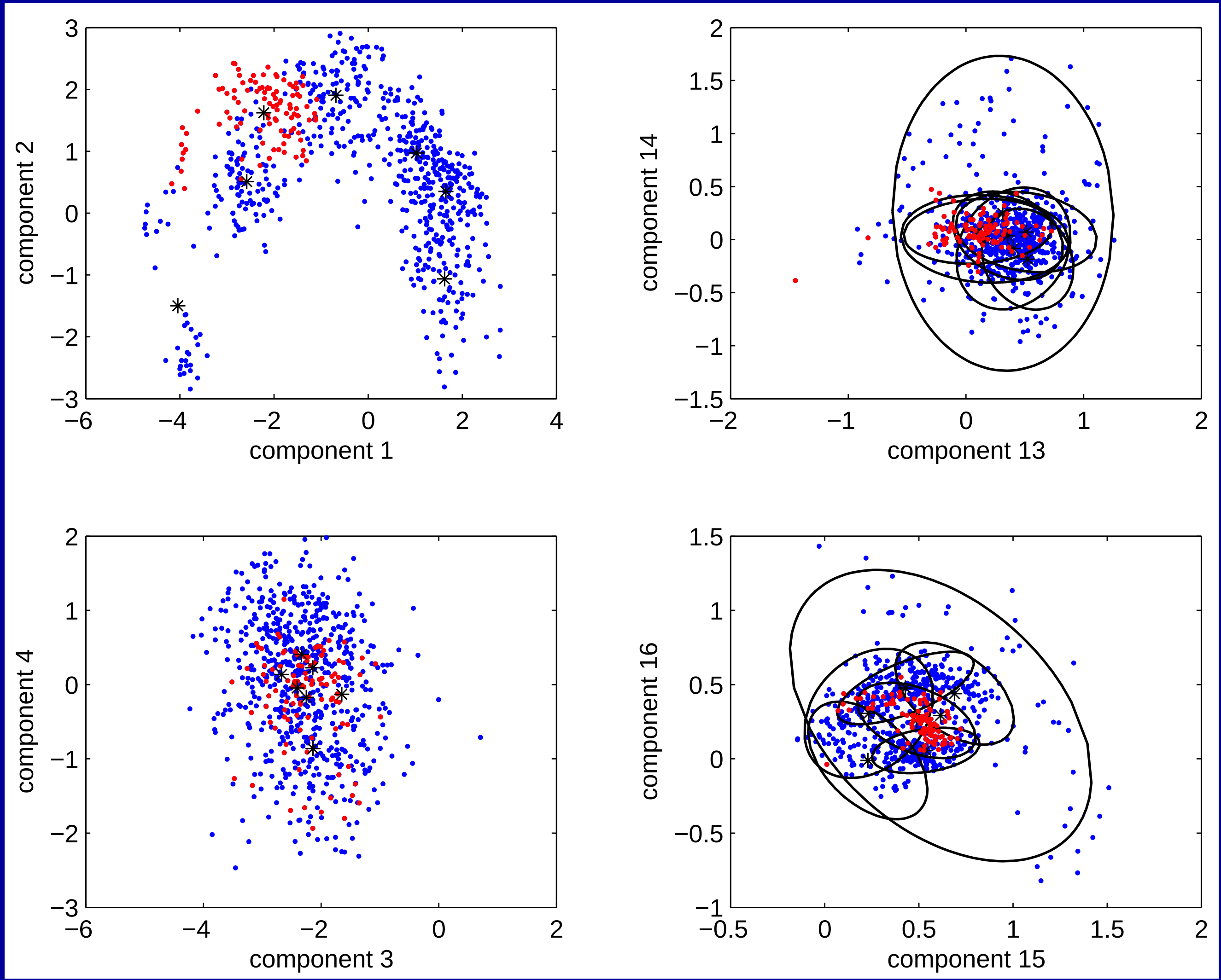
Data	method	error(%)	cv parameters
Digits 1	sKDA	**2.37**	$d' = 16, s = 12$
	KDA	2.60	$d' = 16$
	QDA	4.09	$d' = 15$
Digits 2	sKDA	**3.06**	$d' = 40, s = 25$
	KDA	3.39	$d' = 48$
	QDA	4.40	$d' = 40$
Satellite	sKDA	**8.35**	$d' = 18, s = 9$
	KDA	9.15	$d' = 16$
	QDA	14.55	$d' = 14$

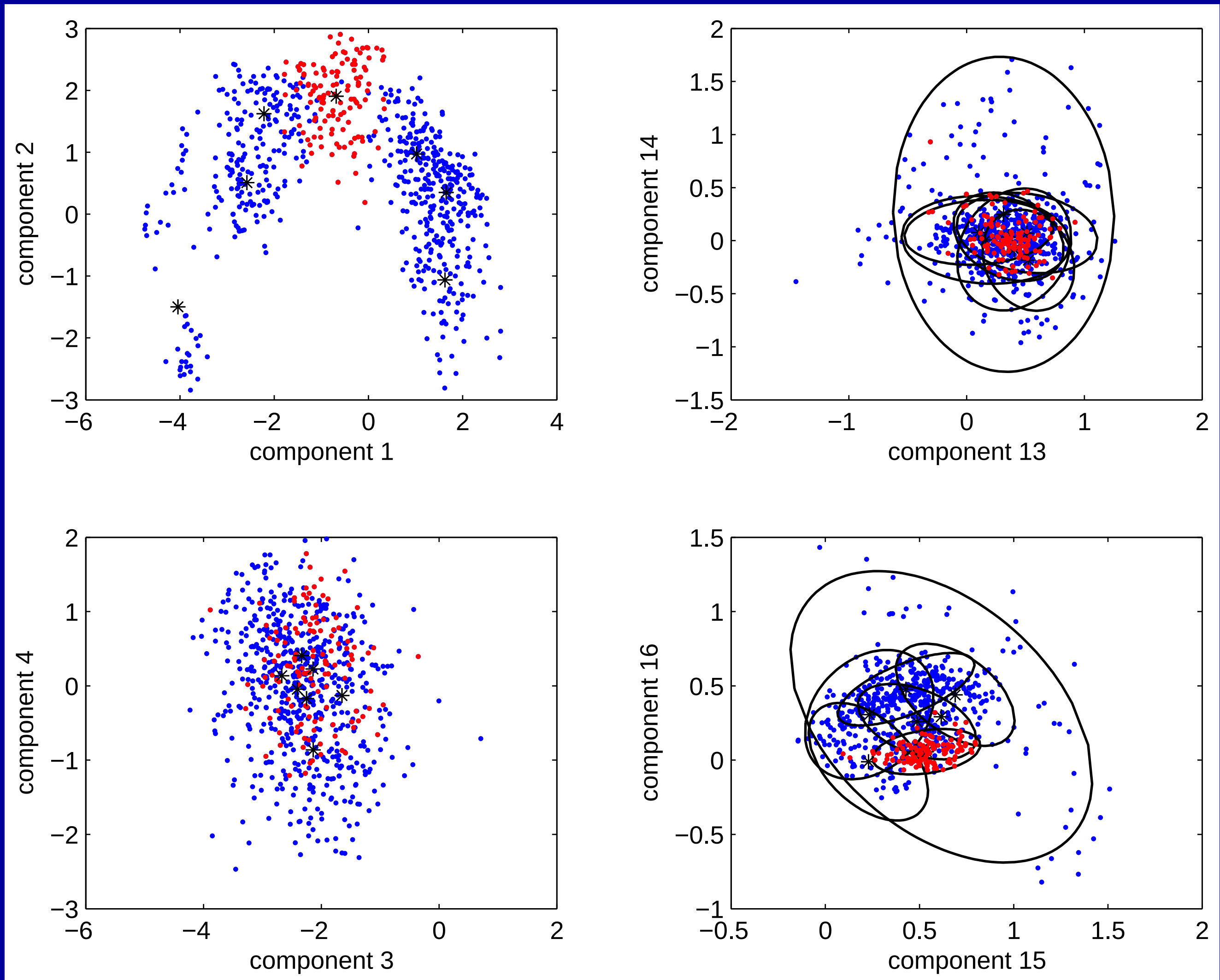
Exploratory visualization

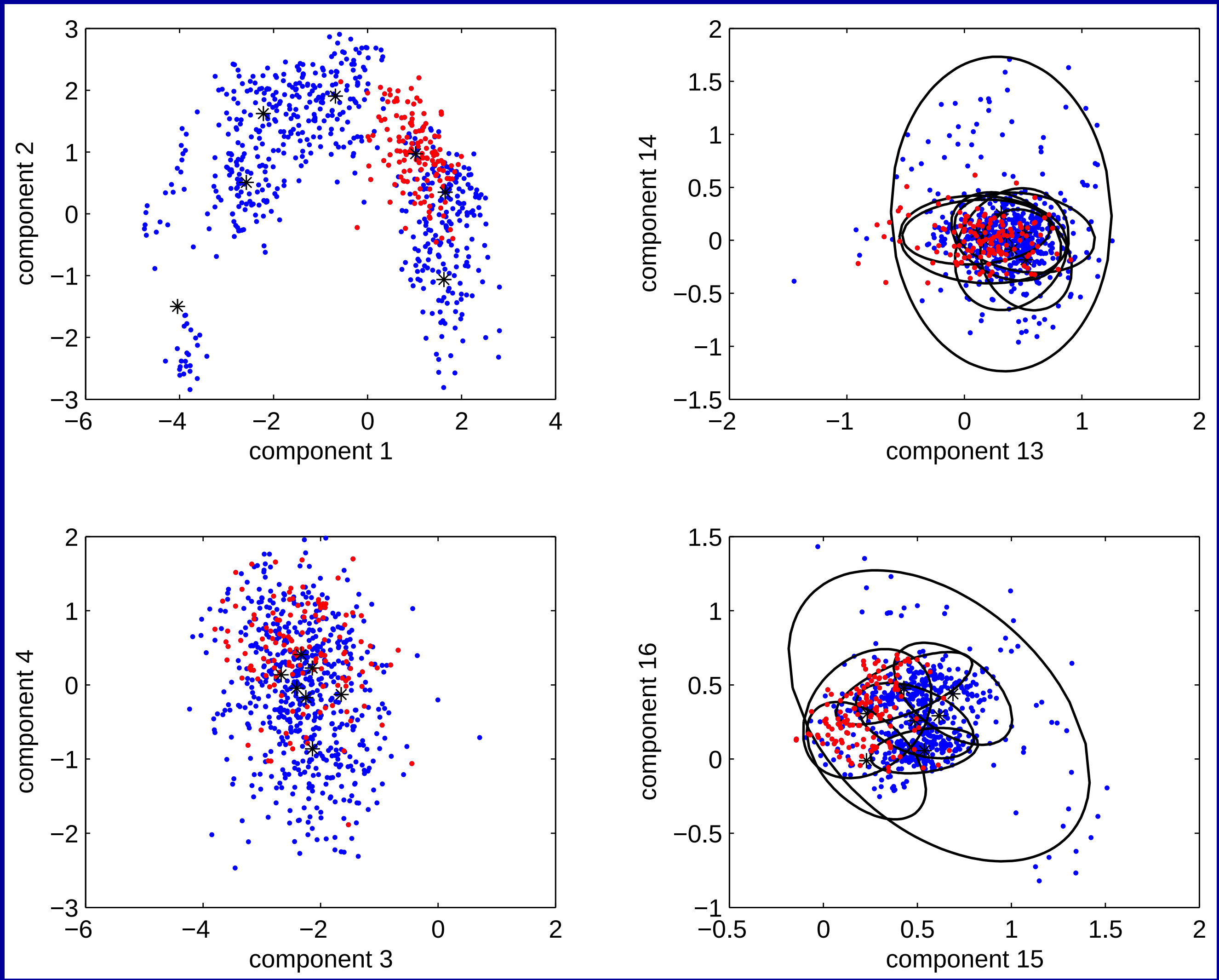

component 1
component 2
component 13
component 14
component 3
component 4
component 15
component 16
3
2
1
0
−1
−2
−3
−6
−4
−2
0
2
4
1.5
0.5
−0.5
−1.5
−1
−0.5
0.5
1.5



component 1
component 2
component 13
component 14
component 3
component 4
component 15
component 16


component 1
component 2
component 13
component 14
component 3
component 4
component 15
component 16


component 1
component 2
component 13
component 14
component 3
component 4
component 15
component 16


component 1
component 2
component 13
component 14
component 3
component 4
component 15
component 16

more on this topic..

F. Hoti and L. Holmström.
A semiparametric density estimation approach to
pattern classification
Pattern Recognition, 37:409-419, 2004