

HIGH DIMENSIONAL DATA CLASSIFICATION FOR PATTERN RECOGNITION

Charles BOUYEYRON, Stéphane GIRARD and Cordelia SCHMID

LMC-IMAG & INRIA Rhône-Alpes

This work was supported by the French department of research under the program "ACI Masse de données"



1. The problem and existing solutions

• The problem:

- Classification is very difficult in high-dimensional spaces because learning methods suffer from the *curse of dimensionality*.
- The data usually contain clusters hidden in different subspaces.

• Existing solutions:

- Global dimension reduction techniques (PCA, feature selection).
- Dimension reduction in a classification purpose (Fisher discriminant analysis).
- Parsimonious models which require the estimation of fewer parameters.

2. The main idea of the new model

In order to efficiently solve the problem of high dimensionality, we propose a new parametrization of the Gaussian model which combines:

- **a local linear approach:** The new parametrization takes into account the specific subspace and the intrinsic dimension of each class.
- **a parsimonious model:** Classes are assumed spherical in order to reduce the number of parameters to estimate.

This new parametrization gives a robust clustering method in high-dimensional spaces and its decision rule is easily understood.

3. The new Gaussian model $[a_i b_i Q_i d_i]$

- The class conditional densities are assumed Gaussian $\mathcal{N}(\mu_i, \Sigma_i)$ with means μ_i , covariance matrices Σ_i and *a priori* probabilities $\pi_i, \forall i = 1, \dots, k$.
- Let Q_i be the orthogonal matrix of eigenvectors of the covariance matrix Σ_i .
- The class conditional covariance matrix Δ_i is defined in the eigenspace of Σ_i by:

$$\forall i = 1, \dots, k, \Delta_i = Q_i^t \Sigma_i Q_i.$$

We assume in addition that Δ_i has only two different eigenvalues $a_i > b_i$. Thus, Δ_i has the following form in the eigenspace of the class C_i :

$$\Delta_i = \left(\begin{array}{ccc|ccc} a_i & & 0 & & & \\ & \dots & & & & \\ 0 & & a_i & & & \\ \hline & & & b_i & & \\ & & & & \dots & \\ 0 & & & & & b_i \end{array} \right) \left. \begin{array}{l} \\ \\ \\ \\ \\ \end{array} \right\} \begin{array}{l} d_i \\ \\ \\ (p - d_i) \end{array}$$

- Let \mathbb{E}_i be the affine space generated by eigenvectors associated to the eigenvalue a_i and such that $\mu_i \in \mathbb{E}_i$. We also define \mathbb{E}_i^\perp such that $\mathbb{E}_i \oplus \mathbb{E}_i^\perp = \mathbb{R}^p$ and $\mu_i \in \mathbb{E}_i^\perp$. Finally, let P_i and P_i^\perp be the projection operators on \mathbb{E}_i and \mathbb{E}_i^\perp . Figure 1 summarizes these notations.

4. High Dimensional Discriminant Analysis

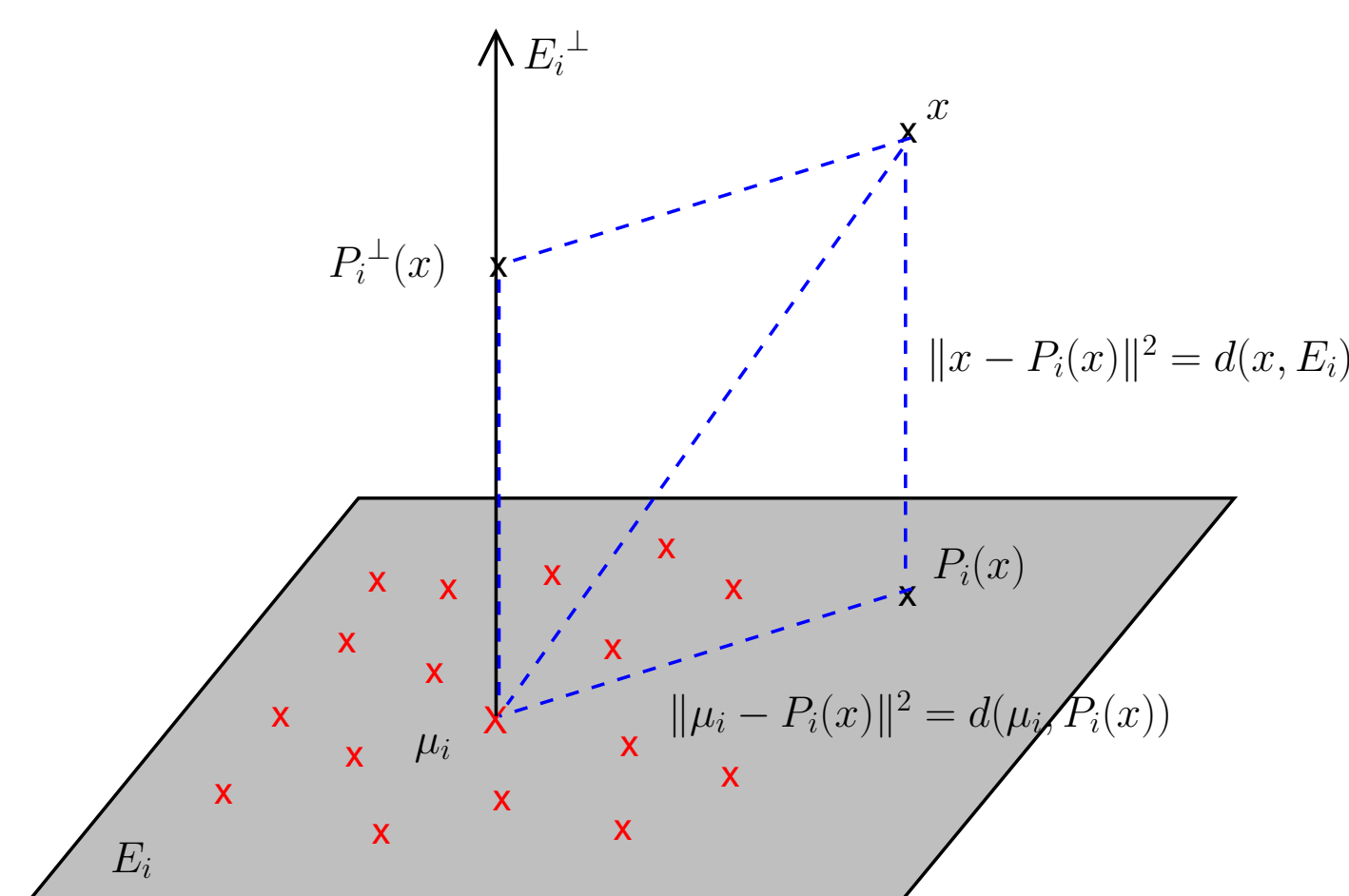


FIG 1: The subspaces \mathbb{E}_i and \mathbb{E}_i^\perp of the class C_i .

Theorem 1. Using the new model $[a_i b_i Q_i d_i]$ in the supervised framework, the Bayes rule yields the decision rule of HDDA which assigns x to the class minimizing the following cost function $K_i(x)$:

$$K_i(x) = \frac{1}{a_i} \|\mu_i - P_i(x)\|^2 + \frac{1}{b_i} \|x - P_i(x)\|^2 + d_i \log(a_i) + (p - d_i) \log(b_i) - 2 \log(\pi_i).$$

By allowing some parameters to be common between classes, HDDA gives particular rules which correspond to different types of regularization.

5. High Dimensional Data Clustering

If we use the model $[a_i b_i Q_i d_i]$ in the Gaussian mixture model framework, the EM algorithm for estimating the parameters $\theta_i = \{\mu_i, a_i, b_i, Q_i, d_i\}$, can be written as follows.

Theorem 2. The E step computes the conditional probabilities $t_{ij}^{(q)}$ with:

$$t_{ij}^{(q)} = 1 / \sum_{l=1}^k \exp \left(\frac{1}{2} (K_l^{(q)}(x_j) - K_i^{(q)}(x_j)) \right),$$

where K_i is defined in Theorem 1. The M step maximizes the conditional likelihood and the updating formulas are given in Section 6.

6. Parameters estimation

- **Estimation of Q_i, a_i and b_i :** Estimation of these parameters is done by the Maximum Likelihood method. The d_i first columns of Q_i are the eigenvectors associated to the d_i largest eigenvalues of $\hat{\Sigma}_i$. The estimator of a_i is $\hat{a}_i = \sum_{l=1}^{d_i} \lambda_{il} / d_i$ and the estimator of b_i is $\hat{b}_i = [\text{tr}(\hat{\Sigma}_i) - \sum_{l=1}^{d_i} \lambda_{il}] / (p - d_i)$, where λ_{il} is the l th largest eigenvalues of $\hat{\Sigma}_i$.
- **Intrinsic dimension estimation:** The estimation of intrinsic dimensions d_i is based on *scree-test* of Catell which looks for a break in the eigenvalues scree.

7. Application to object recognition

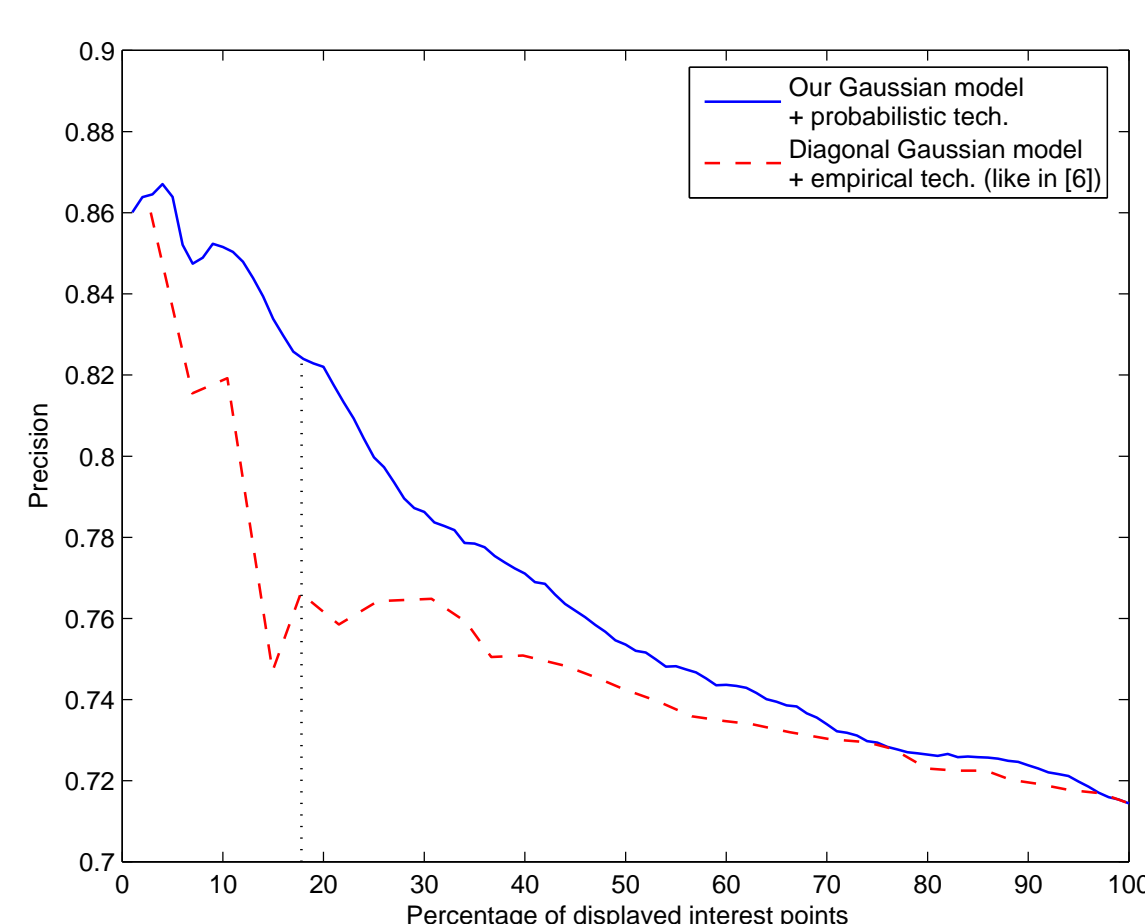


FIG 2: Object localization results.

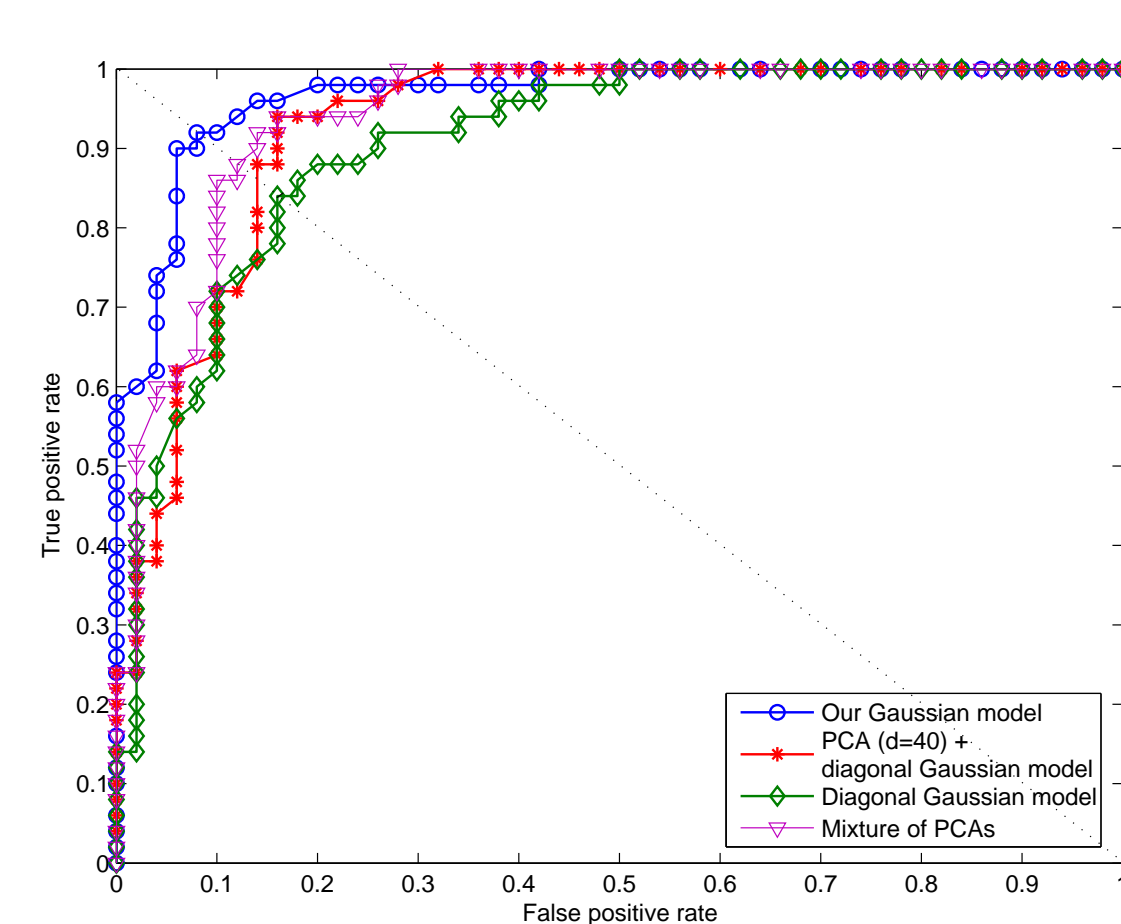


FIG 3: Image classification results.

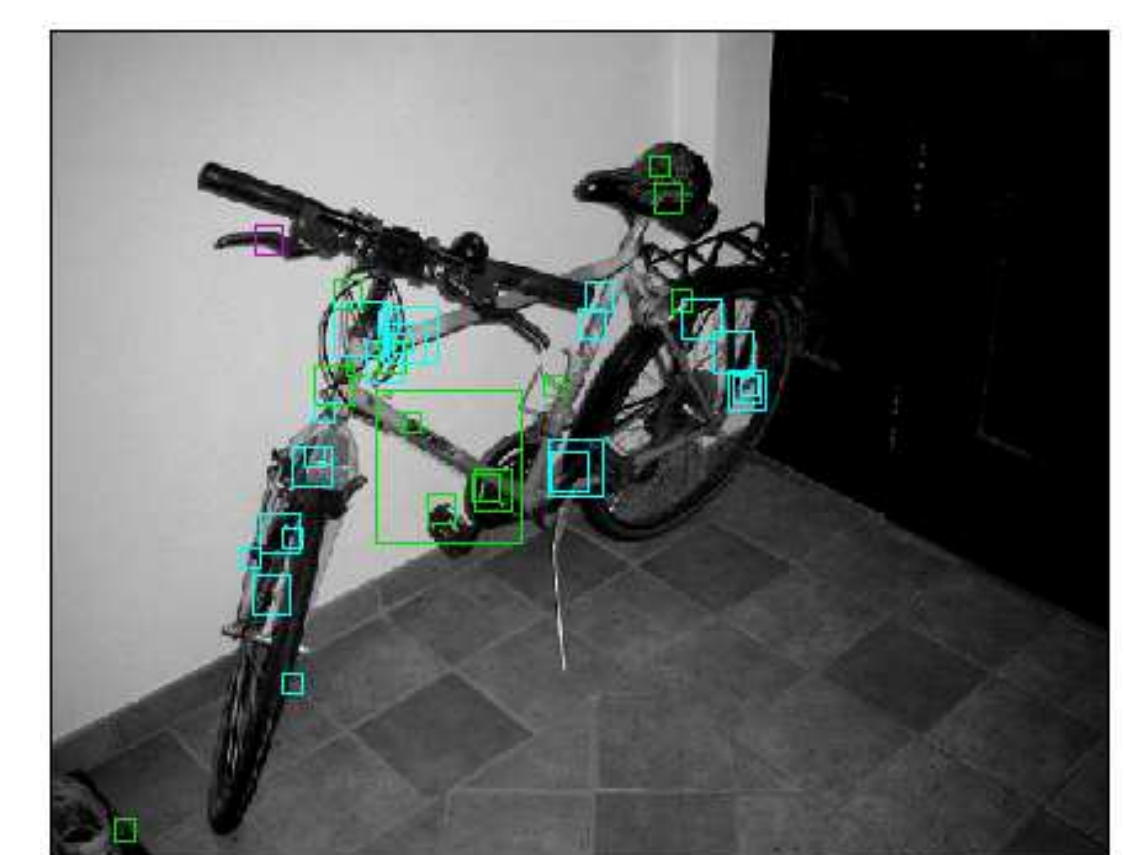


FIG 4: Localization of the object "bicycle" on real images with our statistical approach.

• Our object recognition approach:

- Our approach is based on local descriptors. The detection of interest points is done using the Harris-Laplace operator. Each interest point is then converted to a 128-dimensional descriptor by the SIFT operator.
- We use our classification methods in a weakly-supervised framework, *i.e.* objects in training images are not segmented.

• Learning of discriminant parts:

- The learning of discriminant parts of the object consists in organizing data in k classes using HDDC and then computing the discriminant capacity $P(C_i \in O|C_i)$ of each class.

• Object localization:

- HDDA classifies each interest point x_j and provides the *a posteriori* probability $P(x_j \in C_i|x_j)$ that the interest point x_j belongs to the class C_i .
- Then, we compute the probability $P(x_j \in O|x_j) = \sum_{i=1}^k P(C_i \in O|C_i)P(x_j \in C_i|x_j)$ that x_j belongs to the object.

• Image classification:

- We compute the probabilistic score $S = \frac{1}{m} \sum_{x_j \in I} \sum_{i=1}^k P(C_i \in O|C_i)P(x_j \in C_i|x)$ where m is the number of interest points on I .
- We decide that a test image contains the object if the score S is larger than a given threshold.