

Introduction

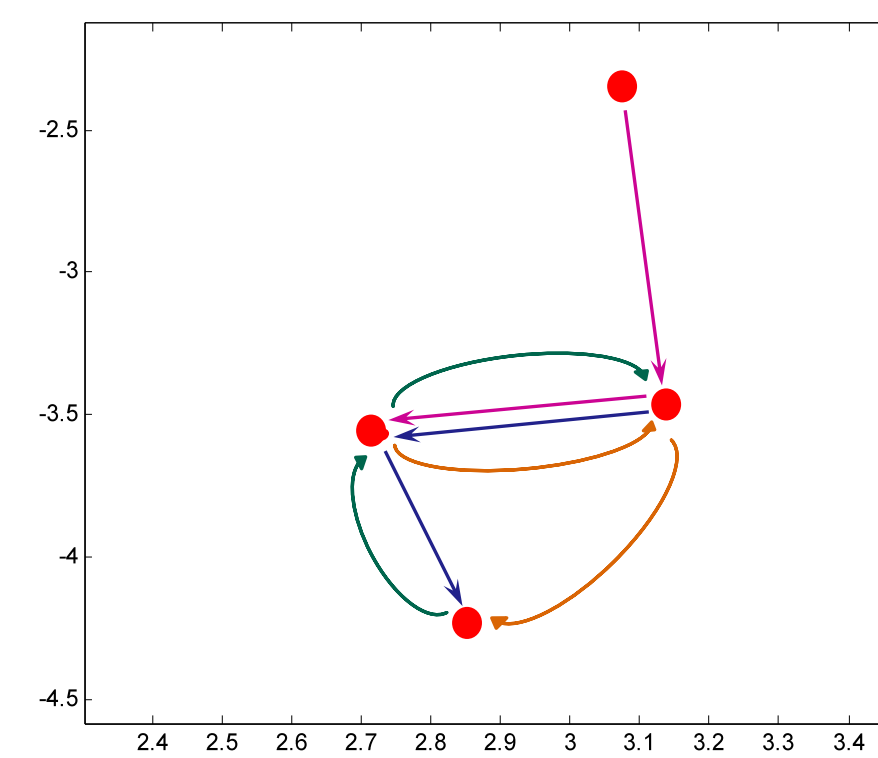
- **Dissimilarity Increments** is a high order dissimilarity exploring triplets of points.
- We propose a novel classifier (**MAP-GMDID**) that combines a maximum *a posteriori* (MAP) approach using Gaussian Mixture Models (GMM) and the **Dissimilarity Increments Distribution (DID)**.
- In this work, objects are described using **dissimilarities** between pairs of objects.
- We build several **pseudo-Euclidean feature spaces** based on the relations given by the dissimilarity matrix. We preserve only the largest eigenvalues.

Pseudo-Euclidean Spaces

- **Pseudo-Euclidean Spaces:**
 - **PES:** $p+q$ eigenvectors (p largest positive and q largest negative)
 - **PPES:** p positive eigenvectors
 - **NPES:** q negative eigenvectors
 - **CES:** add $2|a|$ ($|a|$ largest negative eigenvalue) to all eigenvalues
- **Dissimilarity Spaces:** we compute pairwise Euclidean distances between data points of the previous spaces.

Dissimilarity Increments Distribution (DID)

- $(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k)$ – triplet of nearest neighbors
 - \mathbf{x}_j is the nearest neighbor of \mathbf{x}_i
 - \mathbf{x}_k is the nearest neighbor of \mathbf{x}_j (different from \mathbf{x}_i)



- The **dissimilarity increments** between neighboring patterns is defined as

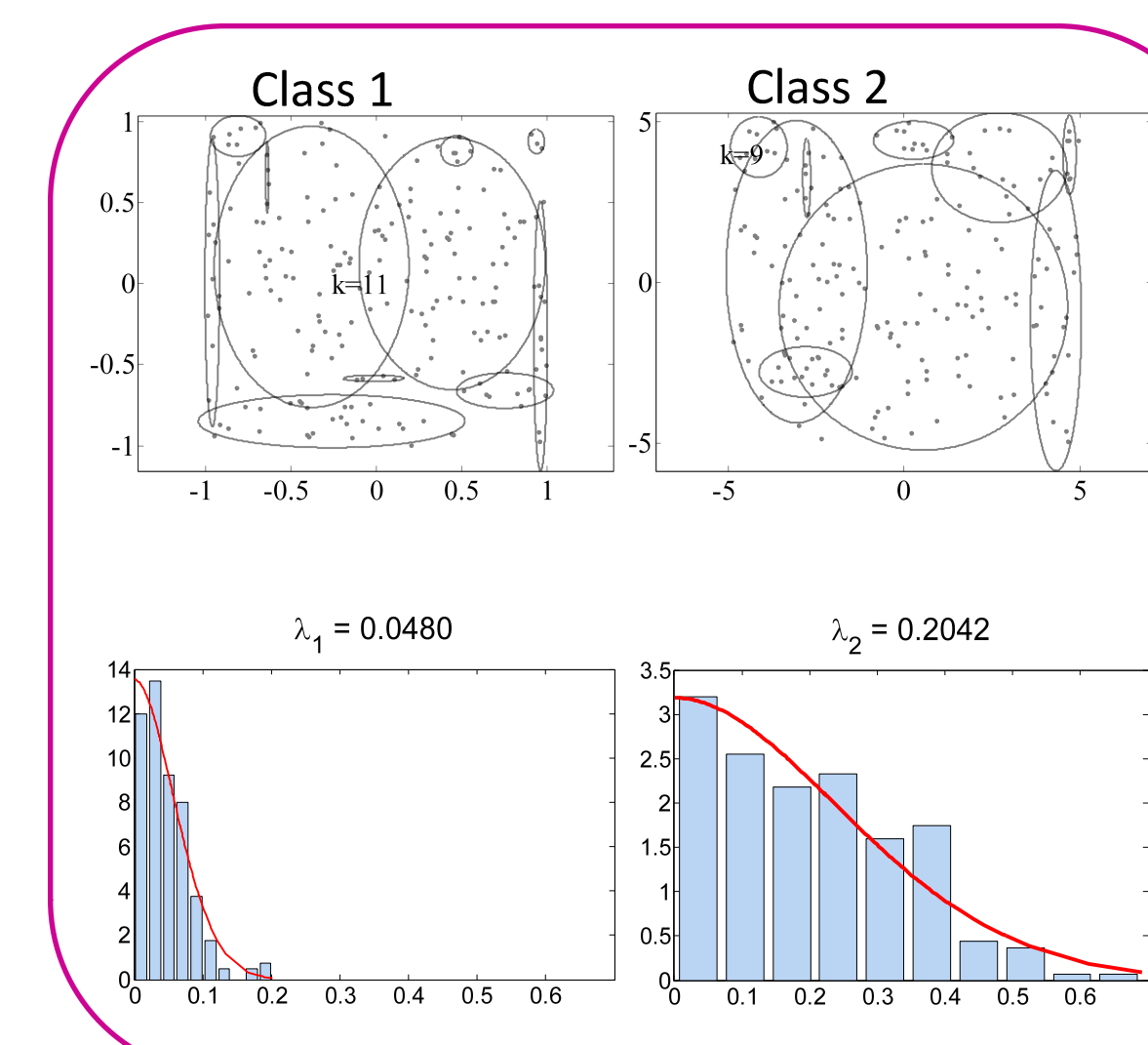
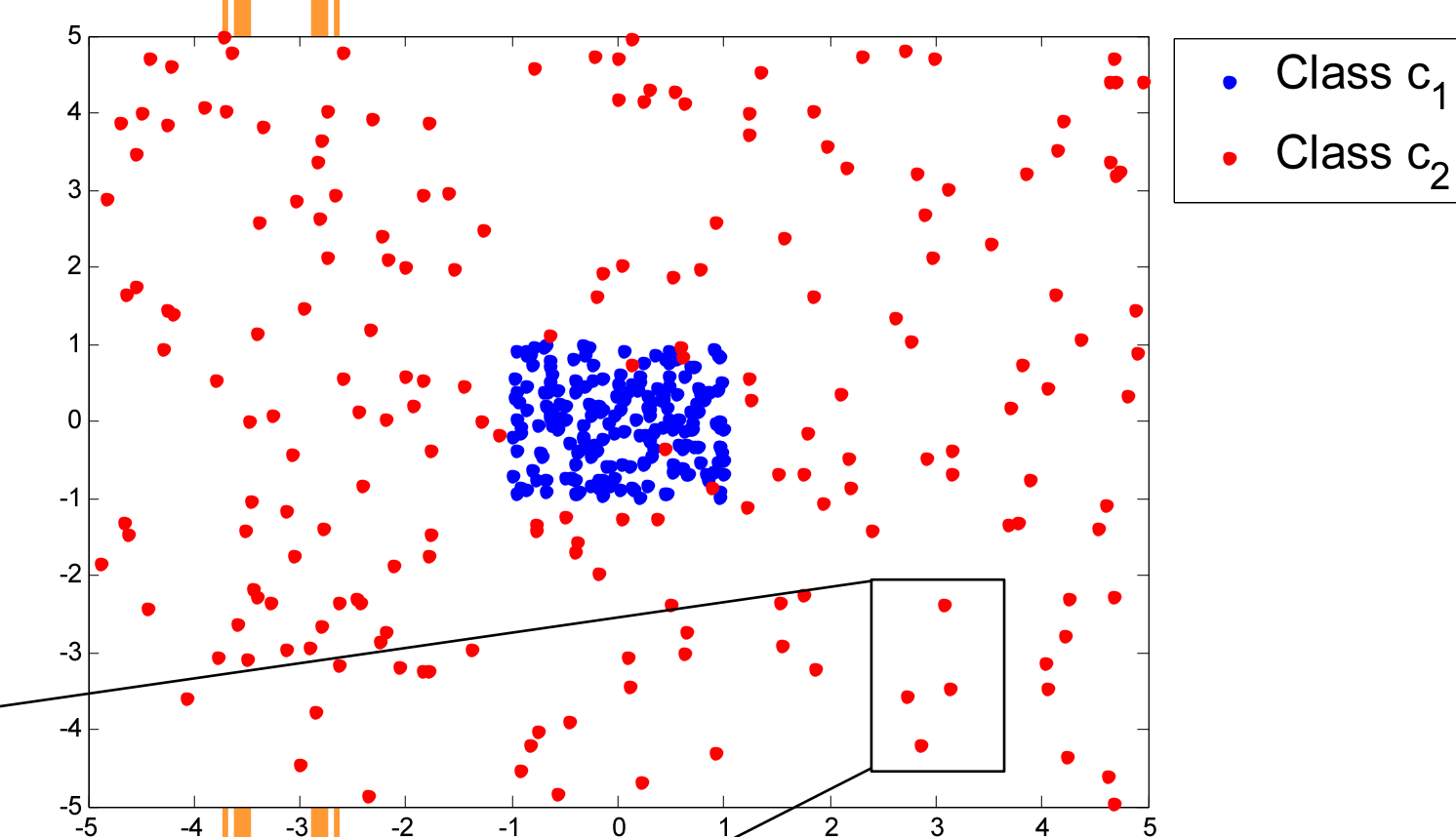
$$d_{inc}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k) = |d(\mathbf{x}_i, \mathbf{x}_j) - d(\mathbf{x}_j, \mathbf{x}_k)|$$

- The **DID** in a class is given by

$$p_{d_{inc}}(w; \lambda) = \frac{\pi\beta^2}{4\lambda^2} w \exp\left(-\frac{\pi\beta^2}{4\lambda^2} w^2\right) + \frac{\pi^2\beta^3}{8\sqrt{2}\lambda^3} \times \left(\frac{4\lambda^2}{\pi\beta^2} - w^2\right) \exp\left(-\frac{\pi\beta^2}{8\lambda^2} w^2\right) \operatorname{erfc}\left(\frac{\sqrt{\pi}\beta}{2\sqrt{2}\lambda} w\right)$$

- $\operatorname{erfc}(\cdot)$ is the complementary error function
- $\beta = 2 - \sqrt{2}$
- $\lambda = \mathbb{E}[w]$

MAP-GMDID Classifier



MAP-GMM
Error = 3.25%

MAP-DID
Error = 3.75%

MAP-GMDID
Error = 2.25%

- $\{\mathbf{x}_i, c_i, inc_i\}_{i=1}^N$ is labeled dataset
 - \mathbf{x}_i is a feature vector in \mathbb{R}^d
 - c_i is the class label
 - inc_i is the set of increments yielded by all the triplets of points containing \mathbf{x}_i

- MAP rule:
$$\max_{c_j} p(c_j | \mathbf{x}_i, inc_i) = \max_{c_j} p(\mathbf{x}_i, inc_i | c_j) p(c_j)$$

- MAP-GMDID combines the GMM and DID assuming that \mathbf{x}_i and inc_i are conditionally independent:

$$\begin{aligned} p(\mathbf{x}_i, inc_i | c_j) &= \left(\sum_{l=1}^K p(\mathbf{x}_i | g_l) \right) p(inc_i | c_j) \\ &= \left(\sum_{l=1}^K \alpha_l p(\mathbf{x}_i | \Sigma_l, \mu_l) \right) \left(\frac{\sum_{n=1}^M p(inc_i^n | \lambda_l)}{M} \right) \end{aligned}$$

- $p(c_j) = |c_j|/N$

Conclusions

- MAP-GMDID can be interpreted as a GMM with an operator that forces a class to have a common increment structure.
- MAP-GMDID outperforms other classifiers, especially in NPES and PES.
- CES and PPES have the lowest error rates for most of the classifiers.
- MAP-GMDID as similar performance in PES, PPES and CES.

