Master's degree in Artificial Intelligence and Computer Science

OPTIMIZATION FOR MACHINE LEARNING - 6 CFU

Project N. 23

Given the dataset **dataset3.mat**, where X is the matrix whose rows contain the points to be classified and y is the array of the corresponding class labels, perform a **spherical separation**, aimed at separating the set $\mathcal{A}$ from the set $\mathcal{B}$, on the basis of the following guidelines:

1. Use the Gaussian kernel with $\sigma = 1$.
2. Choose the sets $\mathcal{A}$ and $\mathcal{B}$ at your convenience, between the two sets of points (positive with label +1 and negative with label -1).
3. Perform a bilevel 10-fold cross validation, using, for the model selection, a 5-fold cross validation. In fixing the grid of C, consider only the cases such that
$$C \geq \frac{1}{2r},$$
with $r = \min\{m, k\}$, where $m = |\mathcal{A}|$ and $k = |\mathcal{B}|$.
4. Each time, fix the center x_0 of the separating sphere
$$S(x_0, R) \triangleq \{x \in R^n \mid \|x - x_0\|^2 = R^2\}$$
as the bariycenter of the set $\mathcal{A}$.
5. Compute the following performance indexes:
   - the average training correctness;
   - the average training sensitivity;
   - the average training specificity;
   - the average training F-score;
   - the average testing correctness (accuracy);
   - the average testing sensitivity;
   - the average testing specificity;
   - the average testing F-score.