Master's degree in Computer Science - A.A. 21/22

OPTIMIZATION FOR MACHINE LEARNING - 6 CFU

Project N. 8

Given the dataset **dataset8.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 **Support Vector Machine (SVM) separation**, aimed at separating the sets $\mathcal{A}$ and $\mathcal{B}$, on the basis of the following guidelines:

1. Use the linear kernel.

2. Each time, solve the primal optimization problem.

3. 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).

4. Perform a bilevel 10-fold cross validation, using, for the model selection, a 5-fold cross validation.

5. For each first-level fold, draw a picture containing the following objects:

   - the first-level training set ($\mathcal{A}$ and $\mathcal{B}$);
   - the separating hyperplane $H(v, \gamma)$;
   - the hyperplane

   $$H^+(v, \gamma) \triangleq \{x \in R^n \,|\, v^T x = \gamma + 1\};$$

   - the hyperplane

   $$H^-(v, \gamma) \triangleq \{x \in R^n \,|\, v^T x = \gamma - 1\}.$$

6. 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.