

Online Learning Exercises

Roberto Paredes

Pattern Recognition and Human Language Technologies
Instituto Tecnológico de Informática
Universidad Politécnica de Valencia

The Workspace

- In the directory the students will find:
 - Several matlab files for loading and processing the data
 - 7 different data sets:
 - Each data set is a binary-classification problem
 - The students have to use the file “training” as the training data
 - The students have to use the file “val” to tune the different parameters
 - The students have to use the file “test” to get an error estimation over the test set
- The students will use this datasets to check their implementations

Load data

- Each set is represented as a file with n rows and $d + 1$ columns.
- Each row is the representation of a vector $\mathbf{x} \in \mathbb{R}^d$
- The last column is the class-label $\{-1, +1\}$
- To load the data you can do:

$[X, y] = \text{load_data}('path/filename')$

Online Passive-Aggressive implementation

- The students have to complete the implementation appearing in the file “online.m”
- The students have to implement the linear PA in the file “pa.m”
- Check the results using the training, val and test data as described before
- To this end the students have to implement some method “test.m” in order to test the results
- Plot results w.r.t the parameter C

Online Passive-Aggressive Kernel implementation

- The students have to complete the implementation appearing in the file “onlinek.m”
- The students have to implement the kernel PA in the file “pak.m”
- Check the results using the training, val and test data as described before
- To this end the students have to implement some method “testk.m” in order to test the results
- Plot results w.r.t the parameters C and kernel parameter (σ for RBF)

Online on a Budget

- Implement the RBP method for linear and kernel
- Use the best parameters C and kernel parameter (in case of kernel)
- Plot results w.r.t the budget parameter B

- Implement the BPA-Simple method for linear and kernel
- Use the best parameters C and kernel parameter (in case of kernel)
- Plot results w.r.t the budget parameter B