Kernel-Based Learning & Multivariate Modeling
MIRI Master

Lluís A. Belanche
belanche@cs.upc.edu

Problem Set #3, Oct 1, 2019

Use the `R` language with the package `kernlab`. You have to download the package, install it (once) and load it into `R` (every time you use it):

```
> install.packages('kernlab')
> library(kernlab)
```

Notes:

1. For hyper-parameter selection (e.g., C, ε, ...), you should use standard resampling methods, like cross-validation.

2. Many times it is helpful to standardize the data prior to doing anything. I suggest to do it as a preprocess, using the `scale` function in `R`) –then you have to "deactivate" automatic scaling in the `ksvm` method by doing `ksvm (..., scaled=c(), ...)`.

Problem 1 The SVM for regression in action

Use the `ksvm` method to perform SVM regression on some data set of your choice (as typical examples, use the Concrete Compressive Strength data set[1] or the Yacht Hydrodynamics data set[2]).

If you wish, you can also use standard multivariate regression methods, like (ridge) regression (`lm.ridge` in `R`); you can also use kernel (ridge) regression for a comparison against linear methods.

The SVM regression is found in the `ksvm` method (be sure to specify `eps-svr`) for the 'type' parameter). Draw conclusions on your results.

.....

Problem 2 The SVM for classification in action

Use the `ksvm` method to perform SVM classification on some data set of your choice (as typical examples, use the Ionosphere data set[3] or the Breast Cancer Wisconsin (Diagnostic) data set[4]).

[1] http://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength
[2] http://archive.ics.uci.edu/ml/datasets/Yacht+Hydrodynamics
[3] http://archive.ics.uci.edu/ml/datasets/Ionosphere
[4] http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)

As less typical (and more challenging) examples, you can use the Pen-Based Recognition of Handwritten Digits data set[5] or the Splice (DNA sequences) data set[6].

If you wish, you can also use standard multivariate classification methods, like discriminant analysis (`lda/qda` methods in `R`); you can also use logistic regression (`glm` method in `R`, for two classes) or multinomial regression (`multinom (nnet)` method, for more than two) for a comparison against linear methods. Draw conclusions on your results.

. . . . . .

[5] http://archive.ics.uci.edu/ml/datasets/Pen-Based+Recognition+of+Handwritten+Digits
[6] http://www.cs.toronto.edu/~delve/data/splice/desc.html