

# Advanced Machine Learning Project

## M2 GENIOMHE – 2015/2016

### Kernel/Multiple Kernel methods

The project will be carried out in groups of 3 students. The goal of this project is to get hands-on experience in machine learning tasks. The project grade will constitute 33% of the total grade.

January 8 (Friday): project due.

You will submit your source code, data, and the final report.

Do not underestimate the importance of writing the final report. Writing the report is just as important as coming up with algorithms, writing code and running the experiments. Even if you get very good results, they will have no value if you are unable to communicate them in writing.

You will choose one of the biological applications described below:

Protein Subcellular Localization: The datasets can be downloaded from

[http://www.cbs.dtu.dk/services/TargetP/data sets/data sets.php](http://www.cbs.dtu.dk/services/TargetP/data%20sets/data%20sets.php).

Protein fold recognition: The datasets can be downloaded from

<http://mkl.ucsd.edu/dataset/protein-fold-prediction>.

Gene expression time series (clustering): The datasets are available at <http://www-genome.stanford.edu/mec1>.

### Supervised kernel/multiple kernel methods:

#### Questions:

Write a script indicating the sequence of functions to be called for:

1. applying the svm algorithm to each dataset
2. Applying the svm to the datasets using a linear combination of the kernels (eg. The mean of kernels) using a fixed rule.
3. Applying svm to the datasets using a heuristic approach.
4. Using an optimization approach (eg. shogun toolbox, SPG-GMKL).
5. Comparing/analyzing the results of the different approaches using cross validation

### Unsupervised kernel/multiple kernel methods:

#### Questions:

Write a script indicating the sequence of functions to be called for:

1. Applying the spectral clustering algorithm to each dataset.
2. Applying the spectral clustering algorithm to the dataset using a linear combination of the kernels (eg. the mean of kernels) using a fixed rule.
3. Applying multiple kernel (spectral) clustering algorithm (eg. affinity aggregation spectral clustering (Matlab code), multiple maximum margin clustering)
4. Evaluating/visualizing/analyzing the clustering results using the data labels (if possible)

To choose the number of clusters  $k$ , you will use the “eigengap” criterion if you use the spectral clustering.