**Instructions**

For the final examination you have a choice:

**1. Build and evaluate support vector machine regression models for an appropriate regression data set, OR**

2. Select a classification data set and build SVM and RandomForest classifiers for this data set and compare them (see Part C below).

Please indicate your choice clearly. For data set selection the same rules apply as for the midterm except that it can be a multi-class problem in case of classification.

**[Part A]** Perform an exploratory data analysis using summary statistics and histograms. Briefly explain your findings.  You can use your analysis from your proposal but please include it here, do not just reference it.

**[Part B]** Build the best model possible for your data set:

1. Document your grid search/model evaluation process carefully, including the type of kernel you are using, the values of its free parameters, and the value of C.
2. For classification use cross-validated accuracy (or error) to find your best model. **For regression use the cross-validated mse, rmse, or R2 metric in order to determine your best model.**
3. Select the two best performing models.

**[Part C]**Investigate whether the difference in performance of your top two models is statistically significant or not using the bootstrap. You should use 95% confidence intervals for this investigation.

1. What are the 95% error confidence intervals for your two models?
2. Is the performance difference statistically significant? If yes, which model would you pick? if no, which model would you pick and why?

Write a brief report summarizing your findings from Parts A, B, and C.   
NOTE: All work has to be your own, no collaboration is permitted.