Homework 6

Classification methods

The aim of the homework is to find a good classification method for the variable `deciasion` (having values `yes` and `no`) in terms of other variables, when the cost of assinging `yes` to observations with actual value `no` is 3 times higher than assigning `no` to observations with actual value `yes`. The training data is in `hw7_train.csv`.

The conditions of the assignment:

1) Find the best GAM, random forest and SVM methods for the prediction problems. Describe how you searched for the best problem (what kind of feature engineering steps were used and why, what parameters were tuned and why you thing those were the most important parameters and so on).
2) Select the best model for predicting `decision`, fit it to the full training set by using a workflow (with command like `wf|>fit(training_data)`) and save resulting fitted workflow and the function `my_predictions(wf,new_data)` for using the result of the fitted workflow to make optimal predictions in the file "familyName_hw7.Rdata" by command `save(model_classification=fitted_workflow, prediction_function=my_predictions),file="familyName_hw7.Rdata"`. If you use any data transformations, then those should be included in the workflow.
3) Submit the .rmd file and .html (or .pdf) file with your solution (output file should show code with results and comments) and additionally the file with the best model together with prediction function.

I evaluate your best model by using test data and best 5 models (according to the loss function corresponding to weighted classification errors describe in the beginning of the probem) give the authors 1 to five extra points. If best 5 are not uniquely determined by the performance measure, then submission time (earlier is better) is used as a secondary ordering variable.

I remind you that you are expected to solve the homework problems by yourself, sharing solutions (and copying parts of other student's solutions) is not allowed.