Choose one of the databases from public sources:

<https://archive.ics.uci.edu/ml/index.php>

<https://www.kaggle.com>

… any other resource of your choice …

(not necessary on biology or medicine)

**HW2 – Build regression models for the chosen data set.**

For every model perform parameter tuning.

1. Linear Regression
2. Ridge Regression
3. Elastic Net
4. Lasso
5. SVM, tuning parameters: type of kernel and cost value
6. Random Forest, tuning parameters: number of trees in the forest, number of predictors randomly chosen at each split. Report variable importance.
7. Gradient Boosting: tuning parameters: take any 3 parameters of your choice (the full description of all parameters in GB can be found here - <https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/> ). Report variable importance.
8. Neural Network. (It is easier to run it in Google Colab). Tuning parameters: number of epoch and learning rate. Here the results are not that important, you must demonstrate that you can run and tune the model. Metrics of performance are not important for evaluation.

Make the final table where you put all the models and their metrics of performance – R2 and RMSE.

**HW3 – Build classification models for the chosen data set.**

1. Logistic Regression
2. SVM, tuning parameters: type of kernel and cost value
3. Random Forest, tuning parameters: number of trees in the forest, number of predictors randomly chosen at each split. Report variable importance.
4. Gradient Boosting: tuning parameters: take any 3 parameters of your choice (the full description of all parameters in GB can be found here - <https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/> ). Report variable importance.
5. Neural Network. (It is easier to run it in Google Colab). Tuning parameters: number of epoch and learning rate. Here the results are not that important, you must demonstrate that you can run and tune the model. Metrics of performance are not important for evaluation.
6. Naïve Bayes

For every model present ROC-curve.

Make the final table where you put all the models and their metrics of performance – Accuracy, Precision, Recall, ROC AUC.

**Deadlines:**

HW2: December 5th 23:59

HW3: December 15th 23:59

Please send the python notebooks with the data sets to Jin Seungmin:

Dry Jin <dryjins@gmail.com>