**Assignment #4**

**Due:**

Wednesday noon (Pacific time) on Week 10 (6/7)

Late submissions will be penalized and the penalty will grow with time.

**How to submit:**

Upload Rmarkdown (code) and html (results) files to a github repository. Send the URL/link via email to @szilard with subject STATS-418 HW-4.

Please do not attach files in email. Please use the email subject exactly like above. You must use github for submissions, no other file repository (e.g. dropbox, domino etc). Any deviations will be penalized.

**Task:**

Use the same dataset as for HW-3.

Try neural networks on your data. Try various architectures, tricks (initialization, regularization, momentum, adaptive learning rate etc.). Use early stopping.

Try hyperparameter optimization for GBMs with random search.

Try ensembling various models.

Do an overall evaluation of all the models for HW-3 and HW-4 above and do a cost-benefit analysis of accuracy vs work and training time required.

As with HW-3, for all of the above use a proper split of train, validation and test set. If you use cross-validation be careful how you do model selection.

Use AUC as a metric of goodness. Also plot the ROC curve in several cases and discuss the tradeoff of true positives (TP) vs false positives (FP) in your particular problem.

Document your findings. Submit a report that's not a mere dump of code and results, but contains an explanation and discussion of your results. In real life you must "sell" your results across the business (your boss, other departments etc.)