

## PROBLEM SET 5

*Assigned: November 11, 2020**Due: Friday, December 04 at midnight*

Opportunities for bonus points applied to your homework score. Any that you do must be turned in by the last data of finals. Be warned: I will run your code with an automated tool that compares it to publically available code on webpages and github. It is not fooled by simply renaming variables or functions.

1. Sign up for `kaggle.com`, pick any public dataset, and perform some kind of analysis of it. It could just be data cleaning and visualization, it could be a more in-depth exploratory analysis, or it could be actually doing classification or regression. Worth up to 10% boost in your *total* homework score, depending on how thorough it is. To turn this in just setup a time to show me what you have done.
2. Pick any publicly available dataset and design a machine learning assignment for the next time I offer the course. You can get some points for an interesting assignment about topics we have already covered, but to maximize the points choose a topic we didn't have a homework assignment on — either deep learning or some topic we covered briefly or not at all. Worth up to 10% boost in your *total* homework score, depending on how thorough it is. Email me your submission, either as a python file, Jupyter notebook, or zip files.
3. Complete part or all of Andreas Müller's deep learning homework assignment here: <https://raw.githubusercontent.com/amueller/COMS4995-s20/master/homework/homework5-aml-2020.pdf>. Worth up to 20% boost in your *total* homework score, depending on how thorough it is. Email me your submission, either as a python file, Jupyter notebook, or zip files.