

# Assignment 0

Machine Learning COMS 4771

Spring 2017, Itsik Pe'er

Assigned: Jan 18<sup>th</sup>

Due: Class time, Jan 30<sup>th</sup>

Submission: Your submission folder on Courseworks

0. Disclose any collaboration[mandatory, ungraded]
1. Explain the academic honesty policy for the class and state your commitment to that policy.  
[10 points]
2. In a couple of sentences, state your goals for this class. What are you planning to achieve?  
[No wrong answers, but we would like to hear from you so encourage that with 5 points credit]
3. Sign up to Khan Academy with your Columbia email. Set [ml4771tas@lists.cs.columbia.edu](mailto:ml4771tas@lists.cs.columbia.edu) as your coach. Make sure you are comfortable with the material in *High School Statistics* and the under-development *AP Statistics*. Either state that you are, or become comfortable by solving all tasks in those sections.  
[20 points]
4. Make sure you can effectively use Python and submit code: Write a Python class called `LinearRegressionSimulator` whose constructor accepts a nonempty numpy array `Theta[0, ..., D]` of coefficients and a non-negative float `StdDev`. The class has a public method `SimData(XInput)` where `XInput` is a pandas data table with columns indexed `1, ..., D` and rows indexed `0, ..., N-1`. `SimData` returns a numpy array whose `i`-th entry is drawn from a normally distributed random variable with standard deviation `StdDev`, and mean  $\text{Theta}[0] + \sum_{d=1 \dots D} \text{Theta}[d] \times \text{XInput}[i, d]$   
Please submit `LinearRegressionSimulator.py` in a subfolder called `Assignment00.Problem04` of your zipped CourseWorks submission.  
Write efficient, pretty code [20 points]

Good luck!