CS 303-01: Introduction to Machine Learning

HW 1 (Given August 31, 2018; Due Sept. 6, 2018)

Email your answers to the TA before midnight on the day it is due

---

Numbers in the parentheses indicate points allocated to the question.

1. Assume that you are a loan officer in a bank and you want to develop an application that will help you score an applicant on the likelihood that he or she will be a defaulter on the loan. While a typical loan officer considers many many variables, assume for simplicity that the training data has only the following variables that have been looked at: For each Client ID we have the Gender, Marital Status, Income, Education, Existing loans and amounts. We also know if a given Client ID has prior defaults. How would you go about designing a Bayes Classifier that will help score an applicant on the likelihood that he or she will default. Be very specific. **(20 points)**

2. Do you think uniformly quantizing a variable from the perspective of characterizing its probaiblity distribution is optimal? If not, Identify at least one scenario where it can be less than optimal. **(20 points)**