Take-home Final Exam

Ishwar Bhat Kattur

IFT598: Analyzing Big Data

Prof. Robert Rucker

05/04/2017

**Question 1**

Before starting this regression question, it would help to be clear about the components and their meaning and uses: Using the components of Q1 to illustrate your answers:

0. entities ( these may be implicit as in the array below), observations, features

1. Independent and dependent variables

2. Categorical versus numerical features

3. Labels

4. Labeled versus unlabeled observations

5. Describe the connection between a regression *algorithm* and a regression *model*, and illustrate this with the *classes* used in the analysis below..

**Entities** in Machine Learning are the crux of the problem for which we are trying to find a solution or build a model for. For example, in the Spam-Ham classification problem, Emails are the entities. Also, each entity is associated with real-valued label or a binary label (depending on whether we are working on a linear regression or logistic regression problem).