Download and get to know the "Adult" census database.

<http://archive.ics.uci.edu/ml/datasets/Adult>

There are 14 attributes in this data set.

1. For each attribute design and program a similarity metric. This is a function that takes in two values attribute values and returns a result in the interval [0,1]. "0" means ' not similar at all, "1" means 'similar'.

2. Write a function that takes two entire records (two rows from the database) and returns a result in the interval [0,1]. "0" means ' not similar at all, "1" means 'similar'.

3. You should hand in a pdf that includes 15 subsections - one for each attribute, and one for the overall function. Each section should contain a short (no more than one paragraph) explanation of how the function works why your strategy makes sense for that attribute and the code for the function itself (not the entire program).

Due Feb 6th end of the day.

Collaboration encouraged. Groups can be no more than 2 people, but you are highly encouraged to talk among groups. Don't share electronic files, but it's ok to post equations and links to online sites that have looked at this problem. (A LOT of people have worked on this). In your hand in, give references for all the help you have gotten. **CLASS DISCUSSION ENCOURAGED.**