

**Course:**

Intro to Data Science – DS-GA-1001 /

Data Mining for Business Analytics - INFO-GB.3336.11

Fall 2014

**Instructor:**

Brian Dalessandro

**Homework 5 – Due 12/03/2014 at 5 pm**

**(Submissions must be a WORD or PDF document)**

Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Grade Total: \_\_\_\_\_\_\_\_\_\_ out of 10**

**1. What does it mean for collaborative filtering methods to be "domain free?" (1pt)**

**2. Why do neighborhood methods in collaborative filtering suffer from the cold start problem, and how is this addressed with matrix factorization methods? (2pts)**

**3. In equation 2, how does the parameter lambda affect the estimates of the matrix factorization? How might one find the best lambda to choose? (2pts)**

**4. How do Collaborative Filtering methods account for the fact that certain items (i.e., songs or movies) are more popular, or widely liked, than others? How do neighborhood methods differ from matrix factorization methods on this? (2pts)**

**5. How might one deal with the fact that users tend to change their preferences over time when making predictions off of a matrix factorization model? (2pts)**

**6. What is the difference between implicit and explicit feedback in the data used to develop recommender systems (1 pt)?**