We discussed in this module a simple linear model mimicking how a centralized data matrix could be factorized in order to retrieve important information. This homework asks you to think about some data in our mundane lives where interesting information can be mined.

1. Describe two possible data sets where "interesting information" might be mined. It will be most fitting if the linear model we described can be applied, but your data sets need be of or for this model.

    (a) (20 pts) If the data sets are available over the network but are too large to be downloaded, list their complete URL's. Make sure that you give credits to the original sources by giving references, if the data set is not your own.

    (b) (20 pts) Provide a short description for each data set. For example, if your data is a matrix, then describe what each dimension or entry represents. You may use a shortened/reduced data set to demonstrate your point.

    (c) (10 pts) Describe what information you want to retrieve. For example, in the linear model, what factors are you looking for.

2. What to submit:

    • If possible, please typeset your report. You may prepare your report in whatever format, but your report should be a PDF file for submission.

3. Some possible repositories:

    • https://www.nature.com/sdata/policies/repositories

    • https://archive.ics.uci.edu/ml/datasets.php

    • http://oad.simmons.edu/oadwiki/Data_repositories