

Project 1: Unsupervised Learning

Instructor: René Vidal

Due Date: 11/24/2017

INSTRUCTIONS: This project is to be done in MATLAB or Python and in groups of three students. Please report which student did which part. Each student should also email me a grade (0-100%=best) for the work of his/her mates.

1. **Low Rank Matrix Completion (LRMC).** Implement Algorithm 3.2 according to the following format.

Function `A=lrmc(X,W,tau,beta)`

Parameters

X $D \times N$ data matrix.

W $D \times N$ binary matrix denoting known (1) or missing (0) entries.

τ Parameter of the optimization problem in (3.30).

β Step size of the dual gradient ascent step in (3.32).

Returned values

A Low-rank completion of the matrix X .

Description

Finds a low-rank matrix A whose entries in W coincide with those of X by using the SVT algorithm.

2. **Face Completion.** In this exercise you will use a small subset of the Yale B dataset, which consists of images of 28 individuals under 9 poses and 64 illumination conditions. Specifically, you will use frontal images from the first individual under ten different illumination conditions, i.e. 10 images. Download the file [YaleB-Dataset.zip](#), which contains faces from 3 individuals as well as 5 non-face images (accessible as images 1-5 from individual 4). This file also the MATLAB function `loadimage.m`. Decompress the file and type `help loadimage` at the MATLAB prompt to see how to use this function. The function operates as follows.

Function `img=loadimage(individual,condition)`

Parameters

`individual` Number of the individual.

`condition` Number of the image for that individual.

Returned values

`img` The pixel image loaded from the database.

Description

Read an image from the dataset. The dataset (directory `images`) must be in the same directory as this file.

Remove uniformly at random 0, 20, or 40 % of the entries of all images of individual 1. Apply the Low Rank Matrix Completion (LRMC) algorithm to fill in the missing entries. Plot the completed faces and comment on the quality of completion as a function of the percentage of missing entries by visually comparing the original images (before removing the missing entries) to the completed ones. Compare your results to those of Fig. 3.2. Plot also the mean-squared-error (MSE) between the original images and the completed ones as a function of the percentage of missing entries and comment your results. Repeat for different choices of τ with $\beta = \min(2, \frac{DN}{M})$.

3. **Movie Recommendation Grand Challenge.** In this exercise you will develop a new recommender system for movies. Given a matrix of incomplete user ratings for different movies, you will use low-rank matrix completion to predict the ratings for other movies. Specifically, you will do the following: Download the [Movielens](#) dataset. Study the README files to get familiarized with the database content. Extract all movies corresponding to the genres Horror and Romance with at least 4 ratings per user. This should give you 231 users and 1615 movies with 13,472 ratings. Alternatively, you can download the specific dataset from here [Movielens-Romance-Horror](#). Divide the ratings into training and test sets. Apply Algorithm 3.2 to the Horror, Romance, and Horror+Romance training sets and report the mean-squared-error (MSE) on the held out test set ratings.