

CSC 576 - 2015 Fall: Homework 2

Submission before the class on Oct. 6

Requirements

L^AT_EX generated Homework is preferred. *One additional point can be earned if your homework is created from L^AT_EX.* Due to the request from some students, the homework is posted online right now, but would be updated probably every week until it is formally released. (*), (**), or (***) indicates the difficulty of each question. Please submit your homework *before* our class on Oct. 6. Any late submission would not be accepted no matter what reasons.

(**) Question 1: 2 points

Define $P_k(X)$ be the polynomial form of matrix X with fixed coefficients $\{a_i\}_{i=1}^k$:

$$P_k(X) = \sum_{i=1}^k a_i X^i.$$

Let (λ, v) be an eigenpairs of the symmetric matrix A . Provide an eigenpair for $[I + P_k(A)A]^T A [I + P_k(A)A]$.

(**) Question 2: 2 points

Given three square matrices $A, B, C \in \mathbb{R}^{n \times n}$. Apply Randomized Kaczmarz algorithm to find a matrix $X \in \mathbb{R}^{n \times n}$ satisfying the following linear equation:

$$AX + XB = C. \tag{1}$$

Assume that there exists a least solution. Write down the algorithm procedure, that is, how to update X_{k+1} from X_k . (Hint: you can think about first how many linear equations defined in (1).)

(**) Question 3: 12 points

This is a programming homework about Netflix recommendation problem. Please first read the problem background on https://en.wikipedia.org/wiki/Netflix_Prize. The original dataset is too large, so here we only use a small data set with 100,000 ratings (1-5) from 943 users on 1682 movies.

I will provide a training data set and a testing data set confident for you guys. You are asked to design an algorithm to estimate missing elements. You need to provide following things:

- (2 points) Upload your runnable code for your algorithm to blackboard (make sure that TA can repeat your result). You will lose all points in this Question if you fail to do this;
- (2 points) Upload your prediction for elements required in the testing dataset in the form of testing data to blackboard. You will lose all points in Question 2 if you fail to do this;
- (4 points) In your hardcopy report, calculate your best RSME values on the training dataset and the testing dataset respectively

$$\text{RSME} = \sqrt{\frac{\sum_{(i,j) \in \text{training (or testing) set}} [\text{your prediction on } (i,j) - \text{true value on } (i,j)]^2}{\# \text{ training (or testing) samples}}}$$

These two values will significantly affect your grade in this homework. An ideal scenario is that the RSME score on the testing dataset is slightly higher than this score on the training dataset, and both should be as small as possible;

- (4 points) In your hardcopy report, describe your algorithm, parameters you chose, and all tricks you used in details. Simply speaking, you need to guarantee that readers can repeat your result based on your algorithm description.

Note that you are not allowed to download the code from anywhere. It would be considered as cheating if you do so. Your score will be evaluated based on all aspects above.