HW III

EE 546: Mathematics of High-dimensional Data

University of Southern California

Assigned on: October 2, 2018 Due date: 3PM November 2, 2020

The purpose of this homework is to help you get aquatinted with concentration of random matrices and also get some experience with randomized numerical linear algebra schemes.

1- **Randomized matrix multiplication.** Suppose we are interested in multiplying two matrices $\boldsymbol{B} \in \mathbb{R}^{m\times N}$ and $\boldsymbol{C} \in \mathbb{R}^{N\times n}$ to arrive at their multiplication $\boldsymbol{A} = \boldsymbol{BC} \in \mathbb{R}^{m\times n}$. However, N is very large and this matrix multiplication takes a long time. Let $\boldsymbol{b}_i$ and $\boldsymbol{c}_i^T$ denote the columns of $\boldsymbol{B}$ and rows of $\boldsymbol{C}^T$ respectively. Let $\mathcal{I}$ be a random variable that takes the value $i \in \{1, 2, \ldots, N\}$ with probability p_i. We shall estimate $\boldsymbol{A}$ by drawing $\mathcal{I}(t) \sim \mathcal{I}$ i.i.d. at random for $t = 1, 2, \ldots, r$ and calculating

$$\hat{\boldsymbol{A}} = \frac{1}{r}\sum_{t=1}^{r}\frac{1}{p_{\mathcal{I}(t)}}\boldsymbol{b}_{\mathcal{I}(t)}\boldsymbol{c}_{\mathcal{I}(t)}^T.$$

(i) Provide an upper-bound on

$$\frac{\mathbb{E}[\|\hat{\boldsymbol{A}} - \boldsymbol{A}\|]}{\|\boldsymbol{B}\|\|\boldsymbol{C}\|}.$$

Hint: This applies for all parts of this problem. Read Section 6.4 of An Introduction to Matrix Concentration Inequalities

(ii) What is a good choice for p_i?

(iii) If we want

$$\frac{\mathbb{E}[\|\hat{\boldsymbol{A}} - \boldsymbol{A}\|]}{\|\boldsymbol{B}\|\|\boldsymbol{C}\|} \le \epsilon,$$

how large do we have to pick r?

(iv) What is the computational cost of forming this approximation as a function of ϵ, m, and n?

2- **Randomized SVD.** Draw $\boldsymbol{X} \in \mathbb{R}^{m\times m}$ and $\boldsymbol{Y} \in \mathbb{R}^{n\times m}$ with $m \ge n$ uniformly at random from the space of orthonormal matrices with $m = 1000$ and $n = 100000$. That is, $\boldsymbol{X}$ and $\boldsymbol{Y}$ uniformly at random from $\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{I}$ and $\boldsymbol{Y}^T\boldsymbol{Y} = \boldsymbol{I}$ e.g. in matlab X=orth(randn(m,m)) and Y=orth(randn(n,m)). Also set the diagonal matrix $\boldsymbol{D} \in \mathbb{R}^{m\times m}$ with diagonal entries

$$D_{ii} = \begin{cases} r - i + 1 & \text{if } i \le r \\ 4\times 10^{-3} & \text{if } i > r \end{cases}$$

with $r = 10$. Set $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ equal to $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{D}\boldsymbol{Y}^T$. Use the randomized SVD of Drineas et al. to calculate the $r = 10$ top eigenvectors of $\boldsymbol{A}$. Let $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ with $\boldsymbol{U} \in \mathbb{R}^{m\times m}$, $\boldsymbol{V} \in \mathbb{R}^{n\times n}$, and $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$ be the singular value decomposition of $\boldsymbol{A}$. Also let $\boldsymbol{U}_r$ and $\boldsymbol{V}_r$ be the singular vectors corresponding to the top r singular values (if the singular values are ordered in decreasing order this corresponds to the first r columns of $\boldsymbol{U}$ and $\boldsymbol{V}$).

(i) How large does c (number of drawn columns/rows) has to be so we can calculate the eigenvectors with relative errors of $\epsilon = 0.01, 0.05, 0.1$. That is,

$$\frac{\|\hat{\boldsymbol{U}}_r\hat{\boldsymbol{U}}_r^T - \boldsymbol{U}_r\boldsymbol{U}_r^T\|}{\|\boldsymbol{U}_r\boldsymbol{U}_r^T\|} \leq \epsilon \quad \text{and} \quad \frac{\|\hat{\boldsymbol{V}}_r\hat{\boldsymbol{V}}_r^T - \boldsymbol{V}_r\boldsymbol{V}_r^T\|}{\|\boldsymbol{V}_r\boldsymbol{V}_r^T\|} \leq \epsilon.$$

Report this number based on the average of 10 random draws from the columns of $\boldsymbol{A}$. That is for the matrix $\boldsymbol{A}$ draw c columns 10 independent times. Compare the average run time (over the 10 random draws) with that of using an SVD on $\boldsymbol{A}$. Would you use the randomized SVD algorithm?

Clarification: In the algorithm of Drineas et. al. to estimate $\boldsymbol{U}_r$ you draw columns at random and to estimate $\boldsymbol{V}_r$ you draw rows at random.

Hint: Note that $\|\boldsymbol{U}_r\boldsymbol{U}_r^T\| = \|\boldsymbol{V}_r\boldsymbol{V}_r^T\| = 1$. Also note that to calculate $\|\hat{\boldsymbol{U}}_r\hat{\boldsymbol{U}}_r^T - \boldsymbol{U}_r\boldsymbol{U}_r^T\|$ and $\|\hat{\boldsymbol{V}}_r\hat{\boldsymbol{V}}_r^T - \boldsymbol{V}_r\boldsymbol{V}_r^T\|$ you don't need to build the matrices explicitly (in which case you probably get an out of memory error). You can calculate the spectral norm of $\hat{\boldsymbol{U}}_r\hat{\boldsymbol{U}}_r^T - \boldsymbol{U}_r\boldsymbol{U}_r^T$ via the power method by calculating matrix-vector products of the form $\left(\hat{\boldsymbol{U}}_r\hat{\boldsymbol{U}}_r^T - \boldsymbol{U}_r\boldsymbol{U}_r^T\right)\boldsymbol{z}$ which can be carried out efficiently. However, note that to calculate the spectrum of the matrix $\hat{\boldsymbol{U}}_r\hat{\boldsymbol{U}}_r^T - \boldsymbol{U}_r\boldsymbol{U}_r^T$ you should probably apply the power method using $\left(\hat{\boldsymbol{U}}_r\hat{\boldsymbol{U}}_r^T - \boldsymbol{U}_r\boldsymbol{U}_r^T\right)^2$ as the power method directly on $\hat{\boldsymbol{U}}_r\hat{\boldsymbol{U}}_r^T - \boldsymbol{U}_r\boldsymbol{U}_r^T$ will fail. Do you know why?

(ii) Repeat the experiment of the previous part with $r = 2, 5, 15, 20$. In these experiments use $\epsilon = 0.05$ and calculate the minimal number of required columns c to get to a relative error of ϵ (again averaged over the 10 trials). Draw c as a function of r. How does c depend on r? Is this consistent with the theorems proved in the class?