COMS 4772 Fall 2016 Homework 3
Due Monday, November 21

Instructions:

- The required number of points for this assignment is 100. Any points you earn beyond this is extra credit.
- The usual homework policies (`http://www.cs.columbia.edu/~djhsu/coms4772-f16/about.html`) are, of course, in effect.
- Using this LaTeX template will be helpful for grading purposes.

Problem 1 (25 points)**.** Let $\boldsymbol{A}$ be a symmetric psd $n \times n$ matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ and corresponding orthonormal eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$. Also let $\widetilde{\boldsymbol{A}}$ be a symmetric psd $n \times n$ matrix with $\hat{\boldsymbol{v}}_1 \in \arg\max_{\boldsymbol{v} \in S^{n-1}} \boldsymbol{v}^\top \widetilde{\boldsymbol{A}} \boldsymbol{v}$. Prove that

$$\langle \hat{\boldsymbol{v}}_1, \boldsymbol{v}_1 \rangle^2 \ \geq \ 1 - \left(2\epsilon/\gamma\right)^2 ,$$

where $\gamma := \lambda_1 - \lambda_2$ and $\epsilon := \|\widetilde{\boldsymbol{A}} - \boldsymbol{A}\|_2$. Try to do this from first principles (i.e., do not invoke Davis-Kahan or Wedin's theorem).

Solution. □

Problem 2 (25 points)**.** Let $\boldsymbol{A}$ be the adjacency matrix in $\{0,1\}^{n\times n}$ for a random undirected graph over n vertices, where the edges appear independently, each with probability at most p. Use Matrix Bernstein (Theorem 1, below) to prove that with probability at least 0.99,

$$\left\|\boldsymbol{A} - \mathbb{E}(\boldsymbol{A})\right\|_2 \;\leq\; O\left(\sqrt{pn\log n} + \log n\right) .$$

Solution. □

Problem 3 (10+65=75 points)**.** In a crowdsourcing problem, there are m images that need to be labeled with either $+1$ or -1, and there are n workers available to do the labeling. Each worker provides $\{\pm 1\}$ labels for all images: the label provided by worker j on image i is $X_{i,j}$. The correct $\{\pm 1\}$ label for image i is Z_i.

Assume the following generative process for the correct labels and worker-provided labels. The process is governed by parameters $\boldsymbol{\gamma} := (\gamma_1, \gamma_2, \dots, \gamma_m) \in [-1,+1]^m$ and $\boldsymbol{\delta} := (\delta_1, \delta_2, \dots, \delta_n) \in [-1,+1]^n$. The data for images $\{(X_{i,1}, X_{i,2}, \dots, X_{i,n}, Z_i)\}_{i=1}^m$ are independent. For each image i,

- the distribution of the correct label is given by

$$\mathbb{P}(Z_i = +1) \;=\; 1 - \mathbb{P}(Z_i = -1) \;=\; \frac{1+\gamma_i}{2}\,;$$

- the worker-provided labels $X_{i,1}, X_{i,2}, \dots, X_{i,n}$ are conditionally independent given Z_i;

- worker j provides the correct label with probability $\frac{1+\delta_j}{2}$: for each $z \in \{\pm 1\}$,

$$\mathbb{P}(X_{i,j} = z \mid Z_i = z) \;=\; 1 - \mathbb{P}(X_{i,j} \neq z \mid Z_i = z) \;=\; \frac{1+\delta_j}{2}\,.$$

Suppose the random matrix $\boldsymbol{X}$ (whose (i,j)-th entry is $X_{i,j}$) is observed, and the correct labels $\boldsymbol{Z} := (Z_1, Z_2, \dots, Z_m)$ are hidden.

(a) Write expressions for the largest singular value of $\mathbb{E}(\boldsymbol{X})$ and also for the corresponding (unit length) left and right singular vectors.

(b) Assume that $\boldsymbol{\gamma} \in \{\pm 1\}^m$ and $\delta_1 \geq 0.1$. Write a procedure for estimating $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ based on the singular value decomposition of $\boldsymbol{X}$. Prove bounds on the Euclidean norm errors of your estimates that hold with probability at least 0.99.

Hint: you may find (some of) the theorems given below to be useful.

Solution.

(a)

(b)

□

Some theorems

Theorem 1 (Matrix Bernstein)**.** *Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N$ be independent, random symmetric matrices in $\mathbb{R}^{d \times d}$. Assume each $\boldsymbol{X}_i$ satisfies $\mathbb{E}(\boldsymbol{X}_i) = \boldsymbol{0}$ and $\lambda_{\max}(\boldsymbol{X}_i) \leq R$ almost surely. For all $t \geq 0$,*

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{i=1}^{N} \boldsymbol{X}_i\right) \geq t\right) \;\leq\; d \cdot \exp\left(-\frac{t^2}{2(\sigma^2 + Rt/3)}\right) \quad \textit{where} \quad \sigma^2 \;:=\; \left\|\sum_{i=1}^{N} \mathbb{E}\boldsymbol{X}_i^2\right\|_2 .$$

Theorem 2 (Weyl)**.** *For any symmetric $n \times n$ matrices $\boldsymbol{A}$ and $\boldsymbol{H}$,*

$$\lambda_i(\boldsymbol{A}) + \lambda_n(\boldsymbol{H}) \;\leq\; \lambda_i(\boldsymbol{A} + \boldsymbol{H}) \;\leq\; \lambda_i(\boldsymbol{A}) + \lambda_1(\boldsymbol{H}), \quad 1 \leq i \leq n,$$

where $\lambda_i(\cdot)$ denotes the i-th largest eigenvalue of its argument.

Theorem 3 (Weyl (again))**.** *For any $m \times n$ matrices $\boldsymbol{A}$ and $\boldsymbol{E}$,*

$$\left|\sigma_i(\boldsymbol{A}) - \sigma_i(\boldsymbol{A} + \boldsymbol{E})\right| \;\leq\; \|\boldsymbol{E}\|_2 , \quad 1 \leq i \leq \min\{m, n\},$$

where $\sigma_i(\cdot)$ denotes the i-th largest singular value of its argument.

Theorem 4 (Davis-Kahan)**.** *Let $\boldsymbol{A} = \boldsymbol{E}_0 \boldsymbol{A}_0 \boldsymbol{E}_0{}^\top + \boldsymbol{E}_1 \boldsymbol{A}_1 \boldsymbol{E}_1{}^\top$ and $\boldsymbol{A} + \boldsymbol{H} = \boldsymbol{F}_0 \boldsymbol{\Lambda}_0 \boldsymbol{F}_0{}^\top + \boldsymbol{F}_1 \boldsymbol{\Lambda}_1 \boldsymbol{F}_1{}^\top$ be symmetric matrices with $[\boldsymbol{E}_0, \boldsymbol{E}_1]$ and $[\boldsymbol{F}_0, \boldsymbol{F}_1]$ orthogonal. If the eigenvalues of $\boldsymbol{A}_0$ are contained in an interval (a, b), and the eigenvalues of $\boldsymbol{\Lambda}_1$ are excluded from the interval $(a - \delta, b + \delta)$ for some $\delta > 0$, then*

$$\left\|\boldsymbol{F}_1{}^\top \boldsymbol{E}_0\right\|_2 \;\leq\; \frac{\left\|\boldsymbol{F}_1{}^\top \boldsymbol{H} \boldsymbol{E}_0\right\|_2}{\delta} .$$

Theorem 5 (Wedin)**.** *Suppose matrices $\boldsymbol{A}, \widetilde{\boldsymbol{A}} \in \mathbb{R}^{m \times n}$ may be written as*

$$\boldsymbol{A} \;=\; \boldsymbol{U}_1 \boldsymbol{S}_1 \boldsymbol{V}_1^\top + \boldsymbol{U}_2 \boldsymbol{S}_2 \boldsymbol{V}_2^\top , \qquad \widetilde{\boldsymbol{A}} \;=\; \widetilde{\boldsymbol{U}}_1 \widetilde{\boldsymbol{S}}_1 \widetilde{\boldsymbol{V}}_1^\top + \widetilde{\boldsymbol{U}}_2 \widetilde{\boldsymbol{S}}_2 \widetilde{\boldsymbol{V}}_2^\top ,$$

where $\boldsymbol{U}_1^\top \boldsymbol{U}_1 = \boldsymbol{V}_1^\top \boldsymbol{V}_1 = \boldsymbol{I}$, $\boldsymbol{U}_2^\top \boldsymbol{U}_2 = \boldsymbol{V}_2^\top \boldsymbol{V}_2 = \boldsymbol{I}$, $\widetilde{\boldsymbol{U}}_1^\top \widetilde{\boldsymbol{U}}_1 = \widetilde{\boldsymbol{V}}_1^\top \widetilde{\boldsymbol{V}}_1 = \boldsymbol{I}$, and $\widetilde{\boldsymbol{U}}_2^\top \widetilde{\boldsymbol{U}}_2 = \widetilde{\boldsymbol{V}}_2^\top \widetilde{\boldsymbol{V}}_2 = \boldsymbol{I}$; and $\boldsymbol{S}_1, \boldsymbol{S}_2, \widetilde{\boldsymbol{S}}_1, \widetilde{\boldsymbol{S}}_2$ are diagonal and non-negative. If there exists $\alpha > 0$ and $\delta > 0$ such that the smallest singular value in $\boldsymbol{S}_1$ is at least $\alpha + \delta$, and the largest singular value in $\widetilde{\boldsymbol{S}}_2$ is at most α, then

$$\max\left\{\left\|\boldsymbol{U}_2^\top \tilde{\boldsymbol{U}}_1\right\|_2 , \left\|\boldsymbol{V}_2^\top \tilde{\boldsymbol{V}}_1\right\|_2\right\} \;\leq\; \frac{\max\left\{\left\|(\widetilde{\boldsymbol{A}} - \boldsymbol{A})\boldsymbol{V}_1\right\|_2 , \left\|\boldsymbol{U}_1^\top (\widetilde{\boldsymbol{A}} - \boldsymbol{A})\right\|_2\right\}}{\delta} .$$