COMPSCI 514: Problem Set 3

Due: 4/10 by 11:59pm in Gradescope.

Instructions:

- You are allowed to work on this problem set in a group of up to three members.
- Each group should **submit a single solution set**: one member should upload a pdf to Gradescope, marking the other members as part of their group in Gradescope.
- The problem set is meant for your own practice. We strongly discourage the use of LLMs to directly solve problems.
- Each problem will be graded on the following scale:
  - ✓+: (2 points) Submitted work demonstrates a full understanding of the problem. There may be some errors, omissions, or unclear steps, but overall, a reader would be able to understand how to solve the problem by looking at the submitted work.
  - ✓-: (1 point) Submitted work demonstrates partial understanding of the concepts, but contains significant omissions or errors.
  - X: (0 points) Not completed or submitted work doesn't not provide enough information to determine whether there is understanding of the problem.

1. Alternative Approaches to Dimensionality Reduction

The Johnson-Lindenstrauss lemma gives that, for any set of points $\vec{x}_1, \ldots, \vec{x}_n$, letting $\mathbf{\Pi} \in \mathbb{R}^{m \times d}$ be a random projection matrix with each entry chosen independently from $\mathcal{N}(0, \frac{1}{m})$ and $m = O\left(\frac{\log(n/\delta)}{\epsilon^2}\right)$, with probability $\geq 1 - \delta$, for all $i, j \in [n]$:

$$(1-\epsilon)\|\vec{x}_i - \vec{x}_j\|_2^2 \leq \|\mathbf{\Pi}\vec{x}_i - \mathbf{\Pi}\vec{x}_j\|_2^2 \leq (1+\epsilon)\|\vec{x}_i - \vec{x}_j\|_2^2.$$

1. Given $\mathbf{\Pi} \in \mathbb{R}^{m \times d}$ distributed as above, what is the runtime of computing the compressed vectors $\tilde{x}_i = \mathbf{\Pi}\vec{x}_i$ for all $\vec{x}_i$? Assume a single basic arithmetic operation (addition, multiplication, etc.) on real numbers takes $O(1)$ time.

2. One very fast alternative to using a random projection matrix $\mathbf{\Pi}$ is to just perform dimensionality reduction by sampling. Let $\mathbf{W} \in \mathbb{R}^{m \times d}$ be a matrix that samples m rows of a vector uniformly at random and re-weights them. Specifically, the j^{th} row of $\mathbf{W}$ is equal to $\sqrt{d/m} \cdot \vec{e}_i$ with probability $1/d$ where $\vec{e}_i$ is the i^{th} standard basis vector for any $i \in [d]$. Show that if we let $\tilde{x}_i = \mathbf{W}\vec{x}_i$ then $\mathbb{E}[\|\tilde{x}_i - \tilde{x}_j\|_2^2] = \|\vec{x}_i - \vec{x}_j\|_2^2$. That is, this simple dimensionality reduction method preserves the distance between vectors in expectation.

3. When might the above sampling method perform poorly in comparison to random projection even though it preserves distances in expectation?

Now suppose that the entries of $\mathbf{\Pi}$ are chosen independently and uniformly in the set $\{-1/\sqrt{m}, 1/\sqrt{m}\}$.

4. Prove that for any $x \in \mathbb{R}^d$, $\mathbb{E}[\|\mathbf{\Pi}x\|_2^2] = \|x\|_2^2$.
5. Prove that for any $x \in \{-1, 1\}^d$, if we let $\mathbf{y} = \mathbf{\Pi}x$, $\mathbb{E}[\mathbf{y}(i)^4] \leq 3d^2/m^2$.
6. Prove that if we set $m = \frac{c}{\epsilon^2 \delta}$ for a large enough constant c, then for any $x \in \{-1, 1\}^d$ we will have
$$(1 - \epsilon)\|x\|_2^2 \leq \|\mathbf{\Pi}x\|_2^2 \leq (1 + \epsilon)\|x\|_2^2$$
with probability at least $1 - \delta$.

2. Linear Algebra Practice

1. Verify that for any two matrices $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $\mathbf{B} \in \mathbb{R}^{d \times k}$, we have $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$. **Hint**: Use the definition of matrix multiplication and the fact that, by definition, for any matrix $\mathbf{M}$, $\mathbf{M}_{ij} = (\mathbf{M}^T)_{ji}$.
2. Use part (1) to conclude that for any set of z matrices, $\mathbf{A}_1 \in \mathbb{R}^{d_0 \times d_1}, \mathbf{A}_2 \in \mathbb{R}^{d_1 \times d_2}, \ldots, \mathbf{A}_z \in \mathbb{R}^{d_{z-1} \times d_z}$ that $(\mathbf{A}_1 \mathbf{A}_2 \ldots \mathbf{A}_z)^T = \mathbf{A}_z^T \mathbf{A}_{z-1}^T \ldots \mathbf{A}_1^T$.
3. Verify that for any two matrices $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $\mathbf{B} \in \mathbb{R}^{d \times n}$ we have $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA}) = \sum_{i=1}^{n} \sum_{j=1}^{d} \mathbf{A}_{ij} \mathbf{B}_{ji}$.
4. Verify that for any $\mathbf{A} \in \mathbb{R}^{n \times d}$, $\|\mathbf{A}\|_F^2 = \text{tr}(\mathbf{A}^T \mathbf{A}) = \text{tr}(\mathbf{A}\mathbf{A}^T)$.
5. Consider any two matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $\mathbf{B} \in \mathbb{R}^{d \times n}$. Let $\mathbf{v} \in \mathbb{R}^n$ be an eigenvector of the matrix $\mathbf{AB}$ with corresponding eigenvalue λ. Show that there is an eigenvector of $\mathbf{BA}$ that also has eigenvalue λ.

3. Projection Matrix Practice

Throughout the following questions, let $\mathbf{V} \in \mathbb{R}^{d \times k}$ be an orthonormal matrix (i.e., its columns $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathbb{R}^d$ all have unit norm and are orthogonal to each other). As discussed in class, $\mathbf{V}\mathbf{V}^T$ is the projection matrix onto the subspace $\mathcal{V} \subset \mathbb{R}^d$ spanned by $\mathbf{V}$'s columns.

1. Prove that $\mathbf{V}\mathbf{V}^T$ projects vectors orthogonally to the subspace $\mathcal{V}$. Formally, show that for any $\mathbf{y} \in \mathbb{R}^d$ and any $\mathbf{x} \in \mathcal{V}$ we have $\langle \mathbf{x}, (\mathbf{y} - \mathbf{V}\mathbf{V}^T\mathbf{y}) \rangle = \mathbf{x}^T(\mathbf{y} - \mathbf{V}\mathbf{V}^T\mathbf{y}) = 0$. **Hint:** Start by using that we can write $\mathbf{x} = \mathbf{Vc}$ for some coefficient vector $\mathbf{c} \in \mathbb{R}^k$.
2. Prove formally that $\mathbf{V}\mathbf{V}^T$ projects any point to the nearest point in the subspace $\mathcal{V}$. That is, prove that for any $\mathbf{y} \in \mathbb{R}^d$:
$$\mathbf{V}\mathbf{V}^T\mathbf{y} = \arg\min_{\mathbf{z} \in \mathcal{V}} \|\mathbf{y} - \mathbf{z}\|_2^2.$$
Hint: Use the Pythagorean theorem: For any $\mathbf{x} \in \mathbb{R}^d$, $\|\mathbf{x}\|_2^2 = \|\mathbf{V}\mathbf{V}^T\mathbf{x}\|_2^2 + \|\mathbf{x} - \mathbf{V}\mathbf{V}^T\mathbf{x}\|_2^2$. Try a proof by contradiction. You may also want to use part (1) and perhaps draw a diagram to help your intuition.
3. Use part (2) to prove that, when $k = d$, $\mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}$, where $\mathbf{I}$ is the $d \times d$ identity matrix. That is – a square matrix with orthonormal columns has orthonormal rows. Or in other words, is its own inverse.
4. Use the Courant-Fischer Principal to prove that $\mathbf{V}\mathbf{V}^T$ has exactly k eigenvalues equal to 1 and exactly $d - k$ eigenvalues equal to 0. **Hint:** You might want to first show that $\|\mathbf{V}\mathbf{V}^T\mathbf{x}\|_2^2 \leq \|\mathbf{x}\|_2^2$ for any vector $\mathbf{x} \in \mathbb{R}^d$.

4. Random Projection for Faster Matrix Multiplication

Let $\boldsymbol{\pi} \in \mathbb{R}^n$ be a random vector with each entry set independently to 1 with probability 1/2 and -1 with probability 1/2. Let $\mathbf{A} \in \mathbb{R}^{n \times d}$ be any matrix.

1. Show that $\mathbb{E}[\mathbf{A}^T \boldsymbol{\pi}\boldsymbol{\pi}^T \mathbf{A}] = \mathbf{A}^T \mathbf{A}$.

   Hint: Fix $i, j \in [d]$ and show that $\mathbb{E}[(\mathbf{A}^T \boldsymbol{\pi}\boldsymbol{\pi}^T \mathbf{A})_{ij}] = (\mathbf{A}^T \mathbf{A})_{ij}$.

2. Show that $\mathbb{E}[\|\mathbf{A}^T \boldsymbol{\pi}\boldsymbol{\pi}^T \mathbf{A} - \mathbf{A}^T \mathbf{A}\|_F^2] \leq 2\|\mathbf{A}\|_F^4$.

   Hint: Fix $i, j \in [d]$ and show that $\mathbb{E}[(\mathbf{A}^T \boldsymbol{\pi}\boldsymbol{\pi}^T \mathbf{A} - \mathbf{A}^T \mathbf{A})_{ij}^2] = \text{Var}((\mathbf{A}^T \boldsymbol{\pi}\boldsymbol{\pi}^T \mathbf{A})_{ij}) \leq 2\|\mathbf{a}_i\|_2^2 \cdot \|\mathbf{a}_j\|_2^2$ where $\mathbf{a}_i, \mathbf{a}_j \in \mathbb{R}^n$ are the i^{th} and j^{th} columns of $\mathbf{A}$. Then sum over all $i, j \in [d]$

3. Let $\boldsymbol{\Pi} \in \mathbb{R}^{n \times m}$ be a random matrix with each entry set independently to $1/\sqrt{m}$ with probability 1/2 and $-1/\sqrt{m}$ with probability 1/2. Show that $\mathbb{E}[\|\mathbf{A}^T \boldsymbol{\Pi}\boldsymbol{\Pi}^T \mathbf{A} - \mathbf{A}^T \mathbf{A}\|_F^2] \leq \frac{2\|\mathbf{A}\|_F^4}{m}$.

   Hint: Show that $\mathbf{A}^T \boldsymbol{\Pi}\boldsymbol{\Pi}^T \mathbf{A} = \frac{1}{m}\sum_{t=1}^m \mathbf{A}^T \boldsymbol{\pi}_t \boldsymbol{\pi}_t^T \mathbf{A}$, where $\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_t \in \mathbb{R}^n$ are independent random vectors distributed as in parts (1) and (2). Then leverage your work from part (2).

4. Show that if $m = \frac{20}{\epsilon^2}$, then with probability at least 9/10, $\|\mathbf{A}^T \boldsymbol{\Pi}\boldsymbol{\Pi}^T \mathbf{A} - \mathbf{A}^T \mathbf{A}\|_F \leq \epsilon \|\mathbf{A}\|_F^2$.
   Note: Here we are looking at the Frobenius norm of $\mathbf{A}^T\ \boldsymbol{\Pi}\boldsymbol{\Pi}^T \mathbf{A} - \mathbf{A}^T \mathbf{A}$, not the squared Frobenius norm. But you may want to apply a concentration inequality to the squared Frobenius norm as part of your proof.

5. In terms of n, d, m, what is the runtime required to compute the approximate matrix product $\mathbf{A}^T \boldsymbol{\Pi}\boldsymbol{\Pi}^T \mathbf{A}$ as compared to the exact product $\mathbf{A}^T \mathbf{A}$.

5. Optimal Low-Rank Approximation From Scratch

In class we used the Courant-Fischer theorem to prove that the best low-rank approximation to any matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ is given by $\mathbf{X}\mathbf{V}_k\mathbf{V}_k^T$ where $\mathbf{V}_k \in \mathbb{R}^{d \times k}$ contains the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$ (i.e., the top k singular vectors of $\mathbf{X}$). Here you will prove this from scratch, using just the basic properties of projection matrices and eigenvectors.

1. Let $\mathbf{X} \in \mathbb{R}^{n \times d}$ be any matrix and $\mathbf{B} \in \mathbb{R}^{n \times d}$ be any rank-k matrix with SVD $\mathbf{B} = \mathbf{W}\mathbf{S}\mathbf{Z}^T$ for orthonormal $\mathbf{W} \in \mathbb{R}^{n \times k}$, $\mathbf{Z} \in \mathbb{R}^{d \times k}$, and diagonal $\mathbf{S} \in \mathbb{R}^{k \times k}$. Prove that $\|\mathbf{X} - \mathbf{B}\|_F^2 = \|\mathbf{X}\mathbf{Z}\mathbf{Z}^T - \mathbf{B}\|_F^2 + \|\mathbf{X} - \mathbf{X}\mathbf{Z}\mathbf{Z}^T\|_F^2$. **Hint:** Use the Pythagorean theorem.

2. Use part (1) to show that if $\mathbf{B} = \underset{\mathbf{M}:\text{rank}(\mathbf{M})=k}{\arg\min} \|\mathbf{X} - \mathbf{M}\|_F^2$ then we have $\mathbf{X}\mathbf{Z}\mathbf{Z}^T = \mathbf{B}$.

3. Using a similar argument as above, one can show that if $\mathbf{B}$ is an optimal rank-k approximation of $\mathbf{X}$ then $\mathbf{W}\mathbf{W}^T\mathbf{X} = \mathbf{B}$. Use this and part (2) to show that: $\mathbf{X}\mathbf{Z} = \mathbf{W}\mathbf{S}$ and $\mathbf{W}^T\mathbf{X} = \mathbf{S}\mathbf{Z}^T$.

4. Use part (3) to show that if $\mathbf{B}$ is an optimal rank-k approximation of $\mathbf{X}$ then $\mathbf{X}^T\mathbf{X}\mathbf{Z} = \mathbf{Z}\mathbf{S}^2$ and use this to argue that each column of $\mathbf{Z}$ is an eigenvector of $\mathbf{X}^T\mathbf{X}$.

5. Complete the proof, showing that the best low-rank approximation of $\mathbf{X}$ is given by $\mathbf{X}\mathbf{V}_k\mathbf{V}_k^T$ where $\mathbf{V}_k$ contains the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$.