

Homework Set 2, CPSC 8420, Fall 2023

Your Name

Due 10/26/2023, Thursday, 11:59PM EST

Problem 1

For PCA, from the perspective of maximizing variance, please show that the solution of ϕ to maximize $\|\mathbf{X}\phi\|_2^2$, *s.t.* $\|\phi\|_2 = 1$ is exactly the first column of \mathbf{U} , where $[\mathbf{U}, \mathbf{S}, \mathbf{U}] = \text{svd}(\mathbf{X}^T \mathbf{X})$. (Note: you need prove why it is optimal than any other reasonable combinations of \mathbf{U}_i , say $\hat{\phi} = 0.8 * \mathbf{U}(:, 1) + 0.6 * \mathbf{U}(:, 2)$ which also satisfies $\|\hat{\phi}\|_2 = 1$.)

Problem 2

Given matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ (assume each column is centered already), where n denotes sample size while p feature size. To conduct PCA, we need find eigenvectors to the largest eigenvalues of $\mathbf{X}^T \mathbf{X}$, where usually the complexity is $\mathcal{O}(p^3)$. Apparently when $n \ll p$, this is not economic when p is large. Please consider conducting PCA based on $\mathbf{X}\mathbf{X}^T$ and obtain the eigenvectors for $\mathbf{X}^T \mathbf{X}$ accordingly and use experiment to demonstrate the acceleration.

Problem 3

Let's revisit Least Squares Problem: minimize $\frac{1}{2}\|\mathbf{y} - \mathbf{A}\boldsymbol{\beta}\|_2^2$, where $\mathbf{A} \in \mathbb{R}^{n \times p}$.

1. Please show that if $p > n$, then vanilla solution $(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}$ is not applicable any more.
2. Let's assume $\mathbf{A} = [1, 2, 4; 1, 3, 5; 1, 7, 7; 1, 8, 9]$, $\mathbf{y} = [1; 2; 3; 4]$. Please show via experiment results that Gradient Descent method will obtain the optimal solution with Linear Convergence rate if the learning rate is fixed to be $\frac{1}{\sigma_{\max}(\mathbf{A}^T \mathbf{A})}$, and $\boldsymbol{\beta}_0 = [0; 0; 0]$.
3. Now let's consider ridge regression: minimize $\frac{1}{2}\|\mathbf{y} - \mathbf{A}\boldsymbol{\beta}\|_2^2 + \frac{\lambda}{2}\|\boldsymbol{\beta}\|_2^2$, where $\mathbf{A}, \mathbf{y}, \boldsymbol{\beta}_0$ remains the same as above while learning rate is fixed to be $\frac{1}{\lambda + \sigma_{\max}(\mathbf{A}^T \mathbf{A})}$ where λ varies from 0.1, 1, 10, 100, 200, please show that Gradient Descent method with larger λ converges faster.

Problem 4

We consider matrix completion problem. As we discussed in class, the main issue of *softImpute* (*Matrix Completion via Iterative Soft-Thresholded SVD*) is when the matrix size is large, conducting SVD is computational demanding. Let's recall the original problem where $\mathbf{X}, \mathbf{Z} \in \mathbb{R}^{n \times d}$:

$$\min_{\mathbf{Z}} \frac{1}{2} \|P_{\Omega}(\mathbf{X}) - P_{\Omega}(\mathbf{Z})\|_F^2 + \lambda \|\mathbf{Z}\|_* \quad (1)$$

People have found that instead of finding optimal \mathbf{Z} , it might be better to make use of *Burer-Monteiro* method to optimize two matrices $\mathbf{A} \in \mathbb{R}^{n \times r}$, $\mathbf{B} \in \mathbb{R}^{d \times r}$ ($r \geq \text{rank}(\mathbf{Z}^*)$) such that $\mathbf{A}\mathbf{B}^T = \mathbf{Z}$. The new objective is:

$$\min_{\mathbf{A}, \mathbf{B}} \frac{1}{2} \|P_{\Omega}(\mathbf{X} - \mathbf{A}\mathbf{B}^T)\|_F^2 + \frac{\lambda}{2} (\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2). \quad (2)$$

- Assume $[\mathbf{U}, \boldsymbol{\Sigma}, \mathbf{V}] = \text{svd}(\mathbf{Z})$, show that if $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}^{\frac{1}{2}}$, $\mathbf{B} = \mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}}$, then Eq. (2) is equivalent to Eq. (1).
- The *Burer-Monteiro* method suggests if we can find $\mathbf{A}^*, \mathbf{B}^*$, then the optimal \mathbf{Z} to Eq. (1) can be recovered by $\mathbf{A}^* \mathbf{B}^{*T}$. It boils down to solve Eq. (2). Show that we can make use of least squares with ridge regression to update \mathbf{A}, \mathbf{B} row by row in an alternating minimization manner as below. Assume $n = d = 2000$, $r = 200$, please write program to find \mathbf{Z}^* .

$T \leftarrow 100, i \leftarrow 1$ % you can also set T to be other number instead of 100

if $i \leq T$ **then**

update \mathbf{A} row by row while fixing \mathbf{B}

update \mathbf{B} row by row while fixing \mathbf{A}

$i \leftarrow i + 1$

end if