

Homework Set 3, CPSC 8420, Fall 2024

Your Name

Due 11/11/2024, 11:59PM EST

Problem 1

Please download the image from [https://en.wikipedia.org/wiki/Lenna#/media/File:Lenna_\(test_image\).png](https://en.wikipedia.org/wiki/Lenna#/media/File:Lenna_(test_image).png) with dimension $512 \times 512 \times 3$. Assume for each RGB channel data X , we have $[U, \Sigma, V] = \text{svd}(X)$. Please show each compression ratio and reconstruction image if we choose first 2, 5, 20, 50, 80, 100 components respectively. Also please determine the best component number to obtain a good trade-off between data compression ratio and reconstruction image quality. (Open question, that is your solution will be accepted as long as it's reasonable.)

Problem 2

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 3

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 $\mathbf{AB}^T = \mathbf{Z}$. The new objective is:

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

- Assume $[\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}] = \text{svd}(\mathbf{Z})$, show that if $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}^{\frac{1}{2}}, \mathbf{B} = \mathbf{V}\mathbf{\Sigma}^{\frac{1}{2}}$, then Eq. (??) is equivalent to Eq. (??).
- The *Burer-Monteiro* method suggests if we can find $\mathbf{A}^*, \mathbf{B}^*$, then the optimal \mathbf{Z} to Eq. (??) can be recovered by $\mathbf{A}^* \mathbf{B}^{*T}$. It boils down to solve Eq. (??). 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 ← 100, i ← 1  % you can also set T to be other number instead of 100
if i ≤ T then
    update A row by row while fixing B
    update B row by row while fixing A
    i ← i + 1
end if

```