Homework 2

A bonus of 20% of the full score will be awarded for the assignment submitted two days before the due date, December 19, 2023. The total score will not exceed 20 points.

1. (10 points) Consider a subset of the MNIST hand-written images which consists of the first 1000 images for digits 0 and 2. Each image is of size 28×28, and is transformed by Gabor wavelets to extract features at different orientations and scales. This results in an array of size $28 \times 28 \times 8 \times 4$ for each image, or equivalently, a matrix $\mathbf{F}$ of size 784×32. Each column of $\mathbf{F}$ corresponds to a different view of the image.

   The goal of this assignment is to perform linear regression on the MNIST images, and classify them as either 0 or 2. The input variables are the features extracted by the Gabor wavelets, and the output variable is the label of the image.

   - (2 points) Part 1: Basic linear regression. Let $\mathbf{x}_k$ be the m-th view features of the k-th image, i.e., the m-th column of $\mathbf{F}_k$. Let y_k be the label of the k-th image. Solve the following linear regression problem:

     $$\min_{\mathbf{w}} \sum_{k=1}^{K} (y_k - \mathbf{x}_k^T \mathbf{w})^2 \tag{1}$$

     where K is the number of images in the subset. Report the classification accuracy for 10-fold cross-validation, where the subset is randomly divided into 10 equal parts, and each part is used as a test set once, while the remaining parts are used as a training set.
   - (6 points) Part 2: Robust linear regression. Consider a variant of linear regression in which the input variables $\mathbf{x}_k$ are perturbed by some noise $\mathbf{e}_k$ with $||\mathbf{e}_k||_2 \leq \delta$, where δ is a given parameter. The noise $\mathbf{e}_k$ represents the uncertainty or variability in the features extracted by the Gabor wavelets.

     The task is to seek a weight vector $\mathbf{w}$ that minimizes the worst-case error due to the noise $\mathbf{e}_k$, i.e.,

     $$\begin{aligned} \min_{\mathbf{w}} \max_{\mathbf{e}_k} \quad & \sum_{k=1}^{K} (y_k - (\mathbf{x}_k + \mathbf{e}_k)^T \mathbf{w})^2 && (1) \\ \text{s.t.} \quad & ||\mathbf{e}_k||_2^2 \leq \delta^2, \quad k = 1, \ldots, K. && (2) \end{aligned}$$

     Simplify the above problem to an unconstrained optimization problem. Solve the simplified problem and report the classification accuracy for 10-fold cross-validation.
   - Add Gaussian noise with zero mean and standard deviation of 0.01 to $\mathbf{x}_k$. Compare perform of the two linear regression methods for the noisy input data, $\boldsymbol{x}_k$.
   - Report classification accuracy for all views.

2. (10 points) For the same data in Prob 1, find a bi-linear system which maps the two view inputs $\boldsymbol{x}_k^{(m)} = \mathbf{F}_k(:, m)$ and $\boldsymbol{x}_k^{(n)} = \mathbf{F}_k(:, n)$ where $n \neq m$ to its labels such that

$$\min_{\mathbf{W}} \max_{\boldsymbol{v}_k, \boldsymbol{u}_k} \quad \frac{1}{K} \sum_{k=1}^{K} (y_k - (\boldsymbol{x}_k^{(m)} + \boldsymbol{v}_k)^T \mathbf{W} (\boldsymbol{x}_k^{(n)} + \boldsymbol{u}_k))_2^2 \tag{3}$$

$$\text{s.t.} \quad \|\boldsymbol{u}_k\|_2^2 \leq \delta^2, \quad \|\boldsymbol{v}_k\|_2^2 \leq \delta^2, \quad k = 1, \ldots, K. \tag{4}$$

where $\mathbf{W}$ is a weight matrix.

Derive update rules to estimate $\mathbf{W}$ and report classification accuracy for 10-fold cross-validation.

3. (10 points) A data matrix $\mathbf{Y}$ is made up of two submatrices, $\mathbf{Y}_1$ and $\mathbf{Y}_2$ of size $I \times J_n$. The task is to solve a distributed low-rank approximation by two agents, such that each agent has only one part of the data $\mathbf{Y}$, and does not have access to the other parts of the data.

Derive the algorithms to update $\{\mathbf{U}_n, \mathbf{V}_n\}$ by the n-th agent for the following problem

$$\min \quad \frac{1}{2}\|\mathbf{Y}_1 - \mathbf{U}_1\mathbf{V}_1^T\|_F^2 + \frac{1}{2}\|\mathbf{Y}_2 - \mathbf{U}_2\mathbf{V}_2^T\|_F^2 \tag{5}$$

$$\text{s.t.} \quad \mathbf{U}_1 = \mathbf{U}_2 \tag{6}$$

Give an example of two agents fitting an image $\mathbf{Y}$.

Report the approximation error

$$err_1 = \|\mathbf{Y}_1 - \mathbf{U}_1\mathbf{V}_1^T\|_F^2 + \|\mathbf{Y}_2 - \mathbf{U}_2\mathbf{V}_2^T\|_F^2$$

and the difference $\|\mathbf{U}_1 - \mathbf{U}_2\|_F^2$ between $\mathbf{U}_1$ and $\mathbf{U}_2$.

Check the approximation error after exchanging $\mathbf{U}_1$ and $\mathbf{U}_2$ in the two agents, i.e.,

$$err_2 = \|\mathbf{Y}_1 - \mathbf{U}_2\mathbf{V}_1^T\|_F^2 + \|\mathbf{Y}_2 - \mathbf{U}_1\mathbf{V}_2^T\|_F^2$$

Are the two approximation errors are significantly different? If so, explain why the algorithm fails and reformulate the problem if possible.

4. (10 points) Consider images from the MNIST dataset, $\mathbf{X}_k$, with missing pixels, $k = 1, 2, \ldots, K$. Denote by $\boldsymbol{\Omega}_k$ binary matrices which indicate the missing elements in $\mathbf{X}_k$, $\omega(i, j) = 0$ for the missing elements $x(i, j)$, otherwise $\omega(i, j) = 1$.

Solve an SVM problem which finds a hyperplane $\{\mathbf{X} \mid \text{tr}(\mathbf{W}^T\mathbf{X}) + b = 0\}$ such that

$$\min_{\mathbf{W}, b} \quad \frac{1}{2}\|\mathbf{W}\|_F^2 \tag{7}$$

$$\text{s.t.} \quad y_k(\text{tr}(\mathbf{W}^T\mathbf{X}_k) + b) \geq 1, \quad k = 1, 2, \ldots, K.$$

where $y_k \in \{-1, 1\}$ are labels for $\mathbf{X}_k$.

- Develop and implement an ADMM algorithm to solve the above problem
- Illustrate performance of the developed algorithms for training a classifier from 11774 handwritten images for digits 0 and 8. 50% of pixels of $\mathbf{X}_k$ are randomly removed. Check the extracted features for 1954 test images.
- Treat the missing elements in $\mathbf{X}_k$ as zeros, and solve the SVM problem (7) again. Compare performance of the two methods.