1. SVD Decomposition. Given an \( m \times n \) matrix \( A \) and its SVD decomposition \( A = U \Sigma V' \) where \( U \) and \( V \) are left and right singular vector matrices, and \( \Sigma \) is the diagonal singular value matrix. Here, \( U = [u_1 \ u_2 \ ... \ u_k] \), \( V = [v_1 \ v_2 \ ... \ v_k] \) where \( u_i \)'s and \( v_i \)'s are the paired left and right singular vectors (columns), and \( \Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k) \). \([20 \text{ points}]

I. Please write down the reconstruction of \( A \) using the top-2 left and right singular-vectors and the singular values. What is the rank of this reconstructed matrix? \([7 \text{ points}]

II. Let \( m = 1000 \), and \( n = 500 \). How many left and right singular vectors should one use in order to best reconstruct \( A \) with (at most) half of the memory needed to store the original matrix \( A \)? \([5 \text{ points}]

III. In practice, the singular values \( \sigma_i \)'s are sorted in a descending order, which is called the spectrum of \( A \). What kind of spectrum makes \( A \) easier to compress by SVD? What kind of matrices will have such kind of spectrum? \([8 \text{ points}]

2. Linear Regression. The mean squared error loss function (MSE) can be written as \( \|Xw-Y\|^2 \) where \( X \) is the data matrix and \( Y \) is the label. \([25 \text{ points}]

I. Compute the gradient of the loss with regard to model parameter vector \( w \), set it to zero, and obtain a closed form solution of \( w \). \([15 \text{ points}]

II. Suppose we add an L2-norm regularization term \( \lambda \|w\|^2 \) to the loss function. In this case, if we follow the same procedure as in I, can we still obtain a closed-form solution? \([10 \text{ points}]

3. Decision function. \([20 \text{ points}]

I. For a binary classification problem illustrated below, write down the decision function, that can correctly predict the labels of samples (+ for positive class, - for negative class) \([10 \text{ points}]

Hint: use the line that passes through (0,0) and (-2,1) as the boundary of the two classes.
II. Predict the label of the two new points A and B, using the decision function you just provided. How certain are you in making the predictions and why? [10]

4. Loss function and Regularization. [35 points]
   I. Given training data \( \{x_1, x_2, \ldots, x_n\} \) and label \( \{y_1, y_2, \ldots, y_n\} \). Write down the exponential loss for linear classification model defined on this data set. [5]
   II. Derive the gradient update rule for the exponential loss, and compare it with the update rule of the Perceptron learning algorithm and explain which might be better? [20]
   III. If we add a regularization term using the L1-norm of model parameter, what will the gradient update rule look like? [5]
   IV. What is the specific advantage using L1-norm regularization? [5]