CMSC 25300 / 35300
Practice Exam 2 – Fall 2019

Name: ______________________________

(Each problem is worth 25 points)

1. ____________

2. ____________

3. ____________

4. ____________

____________ **(total score)**

1. The SVD: The matrix

$$X = \begin{bmatrix} 2.63 & 2.78 & 3.03 & 3.10 \\ 2.23 & 1.87 & 3.46 & 3.67 \\ 2.63 & 2.78 & 3.03 & 3.10 \end{bmatrix}$$

can be factored as $X = U\Sigma V^\top$ where

$$U = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & -2/\sqrt{6} & 0 \\ 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad V = \begin{bmatrix} .43 & .32 & -.84 & .03 \\ .43 & .74 & .51 & -.08 \\ .55 & .36 & .18 & .73 \\ .57 & -.46 & .09 & -.67 \end{bmatrix}$$

a) Think of each column of X as a point in $\mathbb{R}^3$. Based on this factorization, what is a basis find best 1d subspace (line though origin) fit to these four points?

b) Denote the subspace found above B. Write an expression (not numerical values) for calculating the subspace coefficients for each column of X? That is, if we project each point (column of X) onto the subspace spanned by B, then that projection is a weighted sum of the column(s) of B; what are the weight(s)?

c) What is a basis for the best 2d subspace fit to these four points?

d) Below are the SVDs of two different matrices. What is the rank of each matrix? Is a 2d subspace a better fit to the first or second matrix? How can you tell? What is the error of the 2d subspace fit in each case?

```
>> X1

X1 =

    3.0552    6.9651    3.2542    2.5675
   -1.0067    2.1208    1.4278    2.3550
    4.2499    0.3938    5.6056   -1.4958

>> [U1,S1,V1] = svd(X1)

U1 =

   -0.8249   -0.4220   -0.3760
   -0.2188   -0.3751    0.9008
   -0.5212    0.8254    0.2171


S1 =

   10.0000         0         0         0
         0    6.0000         0         0
         0         0    2.0000         0


V1 =

   -0.4515    0.4327   -0.5665    0.5367
   -0.6415   -0.5683   -0.3116   -0.4105
   -0.5918    0.4530    0.6397   -0.1878
   -0.1854   -0.5336    0.4156    0.7129
```

```
>> X2

X2 =

    8.1197    3.7301    7.2735    6.4122
    6.6171    5.6319    3.4323   -0.2696
    1.7773   -1.9314    2.0396    6.9807

>> [U2,S2,V2] = svd(X2)

U2 =

   -0.8217    0.1055   -0.5601
   -0.4878   -0.6383    0.5954
   -0.2947    0.7625    0.5760


S2 =

   16.0000         0         0         0
         0    8.0000         0         0
         0         0    1.0000         0


V2 =

   -0.6515   -0.2515    0.4157   -0.5827
   -0.3277   -0.5843    0.1516    0.7268
   -0.5157    0.0164   -0.8556   -0.0409
   -0.4497    0.7714    0.2686    0.3614
```

2. Logistic regression. Recall the least squares loss $(1 - y_i \boldsymbol{x}_i^\top \boldsymbol{w})^2$ and hinge loss $(1 - y_i \boldsymbol{x}_i^T \boldsymbol{w})_+$. Another option is the logistic loss function defined as

$$L_i(\boldsymbol{w}) = \frac{1}{\log(2)} \log\left(1 + e^{-y_i \boldsymbol{x}_i^\top \boldsymbol{w}}\right) .$$

a) Carefully sketch all three loss functions and explain why the logistic loss function is perhaps a desirable choice for binary classification.

b) Minimizing the sum of logistic loss functions is called *logistic regression*: $\min_{\boldsymbol{w}} \sum_{i=1}^m L_i(\boldsymbol{w})$. Derive a gradient descent algorithm for logistic regression. Explicitly give the computations required at each step.

3. Support vector machines. Consider a classification problem with the following training data.

$$y_1 = +1, \qquad y_2 = +1, \qquad y_3 = +1 \qquad : \qquad \boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \qquad \boldsymbol{x}_2 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \qquad \boldsymbol{x}_3 = \begin{bmatrix} 3 \\ 4 \end{bmatrix},$$

$$y_4 = -1, \qquad y_5 = -1, \qquad y_6 = -1 \qquad : \qquad \boldsymbol{x}_4 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \boldsymbol{x}_5 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \qquad \boldsymbol{x}_6 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

a) Plot these 6 training points in 2-D (use different symbols for the different labels) and compute by inspection the weight vector for the linear classifier (with constant offset) that minimizes the standard hinge loss. Plot the decision boundary for this classifier.

b) Suppose you are given a new data vector $\boldsymbol{x} = \begin{bmatrix} a \\ b \end{bmatrix}$ and you must classify it using your classifier from part **a)**. Write pseudo-code that computes the label ($+1$ or -1) as a function of a and b.

4. Matrix completion

a) What is the "economy" SVD of the following matrix? (You do not need to have your matrix columns normalized. That is, find matrices U, Σ, and V so that $X = U\Sigma V^\top$ and Σ is a diagonal matrix.)

$$X = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{bmatrix}.$$

b) You are given the following matrix with missing entries:

$$X = \begin{bmatrix} 1 & \cdot & 3 & \cdot \\ \cdot & -4 & \cdot & -6 \\ 1 & \cdot & \cdot & 3 \end{bmatrix}.$$

You are told that the columns of X lie in a 1d subspace spanned by the column

$$U = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}.$$

Fill in the missing entries of X.

c) Write pseudocode for performing low-rank matrix completion given a matrix with missing entries (like the one above) where `Omega` indicates the observed entries in `X` and where you do NOT know U. Assume the rank of X is r.