CS/ECE/ME 532

Unit 3 Practice Problems

1. In (a) - (e), let the SVD of a matrix be given as $\boldsymbol{A} = \boldsymbol{U}\Sigma\boldsymbol{V}^T$, with $\Sigma_{1,1} = \sigma_1$ denoting the first singular value.

a) σ_1 is the largest value of $||\boldsymbol{A}\boldsymbol{x}||_2$ for any unit norm vector $\boldsymbol{x}$.
True False

b) σ_1 is the largest value of $||\boldsymbol{x}^T\boldsymbol{A}||_2$ for any unit norm vector $\boldsymbol{x}$.
True False

c) σ_1 is ℓ_2 norm of the vector $\boldsymbol{x}$ that maximizes $||\boldsymbol{x}^T\boldsymbol{A}||_2$.
True False

d) $\sigma_1 = ||\boldsymbol{A}||_{op}$.
True False

e) $\sigma_1^2 = \sum_{i,j} \boldsymbol{A}_{i,j}^2$.
True False

SOLUTION:

(a) True

(b) True

(c) False

(d) True

(e) False

2. You collect eight, four dimensional data points that you store as columns in a matrix $\boldsymbol{X}$:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_8 \end{bmatrix} \tag{1}$$

You cluster the 8 data points by running the k-means algorithm with $k = 3$, which produces cluster centers, $\boldsymbol{t}_1, \boldsymbol{t}_2, \boldsymbol{t}_3$:

$$\boldsymbol{T} = [\boldsymbol{t}_1\ \boldsymbol{t}_2\ \boldsymbol{t}_3] = \begin{bmatrix} 1 & 1 & -2 \\ -4 & -2 & 0 \\ 7 & -6 & 2 \\ 7 & -6 & 9 \end{bmatrix} \tag{2}$$

a) The data points $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ are assigned to cluster $\boldsymbol{t}_1$, while $\boldsymbol{x}_4$ is assigned to $\boldsymbol{t}_2$, and the remaining data points are assigned to $\boldsymbol{t}_3$. Specify the cluster assignment matrix $\boldsymbol{W}$, so that $\boldsymbol{X} \approx \boldsymbol{T}\boldsymbol{W}^T$.

SOLUTION:

$$\boldsymbol{W}^T = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

b) What is the rank of $\boldsymbol{T}\boldsymbol{W}^T$? Why?

SOLUTION: $\boldsymbol{T}\boldsymbol{W}^T$ is rank 3. Clearly $\boldsymbol{W}$ is rank 3, as is $\boldsymbol{T}$ since both have three linearly independent columns. Since both $\boldsymbol{W}$ and $\boldsymbol{T}$ are full rank, the product is equal to the rank of each matrix.

3. You are told that a 3-by-4 matrix $\boldsymbol{X} = \begin{bmatrix} 4 & . & . & . \\ . & -2 & . & . \\ . & . & . & 2 \end{bmatrix}$ has singular-value decomposition $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$ where $\boldsymbol{S} = \begin{bmatrix} 4\sqrt{6} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ and $\boldsymbol{V} = \frac{1}{2}\begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 \end{bmatrix}$. Find the matrix $\boldsymbol{X}$. Be sure to explain how you obtained your answer.

SOLUTION:

$\boldsymbol{X}$ is a rank-1 matrix since there is only one nonzero singular value. Hence each row of $\boldsymbol{X}$ must lie in the space spanned by the first row of $\boldsymbol{V}^T$, that is, must be a scalar multiple of the first row of $\frac{1}{2}\begin{bmatrix} 1 & 1 & -1 & 1 \end{bmatrix}$. Thus, to agree with the given entries, we have

$$\boldsymbol{X} = \begin{bmatrix} 4 & 4 & -4 & -4 \\ -2 & -2 & 2 & 2 \\ -2 & -2 & 2 & 2 \end{bmatrix}$$

4. Your are given n data points in $\mathbb{R}^7$, and you want to cluster them using k means, $k = 2$, with initial clusters centers $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$. Write pseudo-code for the k means algorithm. Use matrix notation: i.e, approximate $\boldsymbol{X} \approx \boldsymbol{T}_i\boldsymbol{W}_i^T$, where $\boldsymbol{T}_i$ is the matrix of cluster centers on iteration i, and $\boldsymbol{W}_i$ is the cluster assignment matrix.

SOLUTION:
set $\mathbf{T}_0 = [\mathbf{t}_1\ \mathbf{t}_2]$.
for $j = 1, \ldots n$, set column j of $\mathbf{W}_0^T$ as $[0\ 1]^T$ if $||\mathbf{x}_j - \mathbf{t}_1|| > ||\mathbf{x}_j - \mathbf{t}_2||$, else set as $[1\ 0]^T$
set $\mathbf{X}_0 = \mathbf{T}_0\mathbf{W}_0^T$
set $i = 0$
while $||\mathbf{X}_i - \mathbf{X}_{i-1}||_F > 0$ (i.e, while not converged)
    Update cluster centers: $\mathbf{t}_1 = \frac{1}{n_1}\mathbf{X}\mathbf{w_1}$ and $\mathbf{t}_2 = \frac{1}{n_2}\mathbf{X}\mathbf{w_2}$ (where $n_j = ||\mathbf{w}_j||_1$)
    for $j = 1, \ldots n$, set column j of $\mathbf{W}_i^T$ as $[0\ 1]^T$ if $||\mathbf{x}_j - \mathbf{t}_1|| > ||\mathbf{x}_j - \mathbf{t}_2||$, else set as $[1\ 0]^T$
    $i = i + 1$
    set $\mathbf{X}_i = \mathbf{T}_i\mathbf{W}_0^T$

5. **The informative SVD.** Consider a data matrix $\boldsymbol{X}$, where the m rows correspond to different training examples and the n columns correspond to different features. Let $\boldsymbol{y}$ be an $m \times 1$ vector with the labels for each example. Suppose the full SVD of $\boldsymbol{X}$ is given by $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$, where:

$$\boldsymbol{U} = \frac{1}{7}\begin{bmatrix} 4 & 2 & 2 & -5 \\ 1 & 4 & 4 & 4 \\ 4 & -5 & 2 & 2 \\ 4 & 2 & -5 & 2 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \boldsymbol{V} = \frac{1}{3}\begin{bmatrix} 2 & -2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & -2 \end{bmatrix}$$

Answer the following questions.

a) For a weight vector $\boldsymbol{w} \in \mathbb{R}^n$, the vector $\boldsymbol{X}\boldsymbol{w}$ is its prediction of the labels. Give a basis for the set of all such prediction vectors.

SOLUTION: $\boldsymbol{X}\boldsymbol{w}$ lies in the space spanned by the first two columns of $\boldsymbol{U}$, since $\boldsymbol{X}$ has rank 2 (there are only two nonzero entries in $\boldsymbol{\Sigma}$.)

b) If we restrict the weight vector to satisfy $\|\boldsymbol{w}\|_2 \leq 1$, what is the largest possible prediction $\boldsymbol{X}\boldsymbol{w}$ (as measured in terms of its 2-norm)?

SOLUTION: The largest possible prediction is obtained when we use $\boldsymbol{w} = \boldsymbol{v}_1$. The prediction with the largest possible norm is:

$$\boldsymbol{X}\boldsymbol{v}_1 = \sigma_1 \boldsymbol{u}_1 = \frac{4}{7}\begin{bmatrix} 4 \\ 1 \\ 4 \\ 4 \end{bmatrix}$$

and this vector has norm $\sigma_1 = 4$. This is the operator norm of $\boldsymbol{X}$ or the largest singular value.

c) Are there weight vectors such that $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{0}$? If so, find a basis for the set of all such vectors.

SOLUTION: Since $\boldsymbol{X}$ has rank 2 columns of $\boldsymbol{X}$ are linearly dependent and thus there is a $\boldsymbol{w}$ satisfying $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{0}$. The set of such $\boldsymbol{w}$ must lie in the space spanned by the last column of $\boldsymbol{V}$.

d) Write an expression for the pseudo-inverse $\boldsymbol{X}^\dagger$ satisfying $\boldsymbol{X}^\dagger \boldsymbol{X} = \boldsymbol{I}$ (you may leave it in factored form)

SOLUTION: The pseudoinverse of $\boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1^\top$ is $\boldsymbol{V}_1\boldsymbol{\Sigma}_1^{-1}\boldsymbol{U}_1^\top$. In this case, this is:

$$\boldsymbol{X}^\dagger = \frac{1}{3}\begin{bmatrix} 2 & -2 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}\frac{1}{7}\begin{bmatrix} 4 & 2 \\ 1 & 4 \\ 4 & -5 \\ 4 & 2 \end{bmatrix}^\top$$

e) Suppose $\boldsymbol{y} = \begin{bmatrix} 7 \\ 0 \\ 0 \\ 0 \end{bmatrix}$. Compute the value of $\min_{\boldsymbol{w}} \|\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}\|_2^2$.

SOLUTION: The minimum error is the projection of $\boldsymbol{y}$ onto the space orthogonal to the space spanned by the columns of $\boldsymbol{X}$. The first two columns of $\boldsymbol{U}$ are a basis for the space spanned by the columns of $\boldsymbol{X}$, so the last two columns of $\boldsymbol{U}$ are a basis for the space orthogonal to the columns of $\boldsymbol{X}$. We can project $\boldsymbol{y}$ onto this subspace easily, since we have an orthonormal basis, and we can find the norm of the residual easily as well:

$$\begin{aligned}
\|\boldsymbol{X}\widehat{\boldsymbol{w}} - \boldsymbol{y}\|^2 &= \|\boldsymbol{U}_2\boldsymbol{U}_2^{\mathsf{T}}\boldsymbol{y}\|^2 \\
&= \|\boldsymbol{U}_2^{\mathsf{T}}\boldsymbol{y}\|^2 \\
&= \left\| \frac{1}{7}\begin{bmatrix} 2 & -5 \\ 4 & 4 \\ 2 & 2 \\ -5 & 2 \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} 7 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\|^2 \\
&= \left\| \begin{bmatrix} 2 \\ -5 \end{bmatrix} \right\|^2 \\
&= 29
\end{aligned}$$