CS/ECE/ME532 Activity 11

Estimated time: 45 mins for Q1 and 20 mins for Q2.

1. See `period_11.ipynb`.

2. Let a 4-by-2 matrix $\boldsymbol{X}$ have SVD $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$ where $\boldsymbol{U} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$, $\boldsymbol{S} =$

2×2

$\begin{bmatrix} 1 & 0 \\ 0 & \gamma \end{bmatrix}$, and $\boldsymbol{V} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. $X = \frac{1}{2}\begin{bmatrix} 1 & r \\ 1 & -r \\ 1 & -r \\ 1 & r \end{bmatrix} \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{2}\frac{1}{\sqrt{2}}\begin{bmatrix} 1+r & 1-r \\ 1-r & 1+r \\ 1-r & 1+r \\ 1+r & 1-r \end{bmatrix}$

a) Express the solution to the least-squares problem $\arg\min_{\boldsymbol{w}} ||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$ as a function of $\boldsymbol{U}$, $\boldsymbol{S}, \boldsymbol{V}$, and $\boldsymbol{y}$.

b) Let $\boldsymbol{y} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$. Find the weights $\boldsymbol{w}$ that minimize $||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$ as a function of γ. Calculate $||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$ and $||\boldsymbol{w}||_2^2$ as a function of γ for this value of $\boldsymbol{w}$. What happens to $||\boldsymbol{w}||_2^2$ as $\gamma \to 0$?

c) Now consider a "low-rank" inverse. Instead of writing

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T = \sum_{i=1}^{p} \frac{1}{\sigma_i}\boldsymbol{v}_i\boldsymbol{u}_i^T$$

where p is the number of columns of $\boldsymbol{X}$ (assumed less than the number of rows), we approximate

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \approx \sum_{i=1}^{r} \frac{1}{\sigma_i}\boldsymbol{v}_i\boldsymbol{u}_i^T$$

In this approximation we only invert the largest r singular values, and ignore all of them smaller than σ_r. If $r = 1$, use the low-rank inverse to find $\boldsymbol{w}$, $||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}||_2^2$, and $||\boldsymbol{w}||_2^2$ when $\boldsymbol{y} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$ as in part b). Compare $||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}||_2^2$, and $||\boldsymbol{w}||_2^2$ to the results for part b).

a) $w = (X^TX)^{-1}X^Ty = (VS^TU^TUSV^T)^{-1}VS^TU^Ty$

$= VS^{-1}U^Ty$

b) $V = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$ $S^{-1} = \begin{bmatrix}1 & 0\\ 0 & \frac{1}{r}\end{bmatrix}$ $U^T = \frac{1}{2}\begin{bmatrix}1 & 1 & 1 & 1\\ 1 & -1 & -1 & 1\end{bmatrix}$ $y = \begin{bmatrix}1\\0\\0\\1\end{bmatrix}$

$VS^{-1} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & \frac{1}{r}\\ 1 & -\frac{1}{r}\end{bmatrix}$ $VS^{-1}U^T = \frac{1}{\sqrt{2}}\frac{1}{2}\begin{bmatrix}1+\frac{1}{r} & 1-\frac{1}{r} & 1-\frac{1}{r} & 1+\frac{1}{r}\\ 1-\frac{1}{r} & 1+\frac{1}{r} & 1+\frac{1}{r} & 1-\frac{1}{r}\end{bmatrix}$

$w = \frac{1}{2\sqrt{2}}\begin{bmatrix}2+\frac{2}{r}\\ 2-\frac{2}{r}\end{bmatrix}$

$\|Xw-y\|_2^2 = \frac{1}{2\sqrt{2}}\begin{bmatrix}1+r & 1-r\\ 1-r & 1+r\\ 1-r & 1+r\\ 1+r & 1-r\end{bmatrix}\frac{1}{2\sqrt{2}}\begin{bmatrix}2+\frac{2}{r}\\ 2-\frac{2}{r}\end{bmatrix} - \begin{bmatrix}1\\0\\0\\1\end{bmatrix}$

$\|w\|_2^2 = \left(\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{2}r}\right)^2 + \left(\frac{1}{\sqrt{2}}-\frac{1}{\sqrt{2}r}\right)^2 = \frac{1}{2}+r+\frac{1}{2r^2}+\frac{1}{2}-r+\frac{1}{2r^2}$

$= 1+\frac{1}{r^2}$

As lambda approaches $\emptyset$,
$\|w\|_2^2$ increases/approaches ∞.

c) $w = 1\left(\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}\right)\frac{1}{2}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}\begin{bmatrix}1\\0\\0\\1\end{bmatrix}$

$= \begin{bmatrix}\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}}\\ \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}}\end{bmatrix}\begin{bmatrix}1\\0\\0\\1\end{bmatrix}$

$= \begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}$

$\|w\|_2^2 = \left(\frac{1}{\sqrt{2}}\right)^2 + \left(\frac{1}{\sqrt{2}}\right)^2 = 1$

This norm is less than what it was in part b.

CS/ECE/ME532 Period 11 Activity

Preambles

```
%matplotlib inline
# to enable 3D plot interaction
import numpy as np # numpy
from pprint import pprint as pprint # pretty print
from scipy.io import loadmat # load & save data
from scipy.io import savemat
import matplotlib.pyplot as plt # plot
from mpl_toolkits import mplot3d
np.set_printoptions(formatter={'float': lambda x: "{0:0.2f}".format(x)})
```

K-means has some 'random' components in it. You will get different results depending on your luck. Even when you run an identical code, you will see some different results from your peers. So... we need the following line of code to start with:

```
np.random.seed(2)
```

Indeed, one may be tempted to try so many random seeds until you get a good performance!

Don't do that :-)... Some subfields in ML are suffering from "reproduction crisis" partially due to this: See these for more details https://arxiv.org/abs/1709.06560 https://www.nature.com/articles/d41586-019-03895-5 https://www.wired.com/story/artificial-intelligence-confronts-reproducibility-crisis/

And see the following figure from the attached paper:

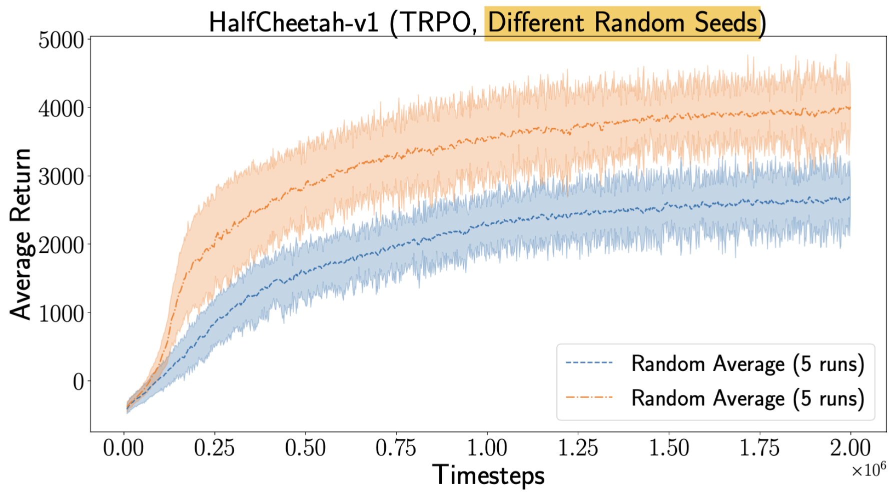


Figure 5: TRPO on HalfCheetah-v1 using the same hyperparameter configurations averaged over two sets of 5 different random seeds each. The average 2-sample t-test across entire training distribution resulted in $t = -9.0916$, $p = 0.0016$.

---

1. K-means and SVD for rating prediction

We return to the movies rating problem considered previously. The movies and ratings from your friends on a scale of 1-10 are:

Movie	Jake	Jennifer	Jada	Theo	Ioan	Bo	Juanita
Star Trek	4	7	2	8	7	4	2
Pride and Prejudice	9	3	5	6	10	5	5
The Martian	4	8	3	7	6	4	1
Sense and Sensibility	9	2	6	5	9	5	4
Star Wars: Empire Strikes	4	9	2	8	7	4	1

Run the following code block to create a numpy array X

```
X = np.array([
    [4,7,2,8,7,4,2],
    [9,3,5,6,10,5,5],
    [4,8,3,7,6,4,1],
    [9,2,6,5,9,5,4],
    [4,9,2,8,7,4,1],
    ], float)
print(X)
```

```
    [[4.00 7.00 2.00 8.00 7.00 4.00 2.00]
```

```
 [9.00 3.00 5.00 6.00 10.00 5.00 5.00]
 [4.00 8.00 3.00 7.00 6.00 4.00 1.00]
 [9.00 2.00 6.00 5.00 9.00 5.00 4.00]
 [4.00 9.00 2.00 8.00 7.00 4.00 1.00]]
```

> `float` is necessary as the array will only hold integers otherwise

Also, we load the K-mean algorithm we implemented in the last activity.

```
def dist(x, y):
    return (x-y).T@(x-y)

def kMeans(X, K, maxIters = 20):
    X_transpose = X.transpose()
    centroids = X_transpose[np.random.choice(X.shape[0], K)]
    for i in range(maxIters):
        # Cluster Assignment step
        C = np.array([np.argmin([dist(x_i, y_k) for y_k in centroids]) for x_i in X_transpose]
        # Update centroids step
        for k in range(K):
            if (C == k).any():
                centroids[k] = X_transpose[C == k].mean(axis = 0)
            else: # if there are no data points assigned to this certain centroid
                centroids[k] = X_transpose[np.random.choice(len(X))]
    return centroids.transpose() , C
```

Note that `(x-y).T@(x-y)` is the squared L^2 norm of $x - y$: since x and y are 1-d numpy arrays, the .T does not actually impact the code.

1 a) Use the K-means algorithm to represent the columns of X with two clusters.

```
centroids_2, C_2 = kMeans(X, 2) ## Fill in the blank: call the "kMeans" algorithm with proper
print('centroids = \n', centroids_2)
print('centroid assignment = \n', C_2)
```

```
centroids =
 [[6.33 3.75]
 [8.33 4.50]
 [5.67 4.00]
 [7.67 4.25]
 [6.33 4.00]]
centroid assignment =
 [0 1 1 0 0 1 1]
```

1 b) Express the rank-2 approximation to X based on this cluster as TW^T where the columns of T contains the cluster centers and W is a vector of ones and zeros. Compare the rank-2 clustering

approximation to the original matrix.

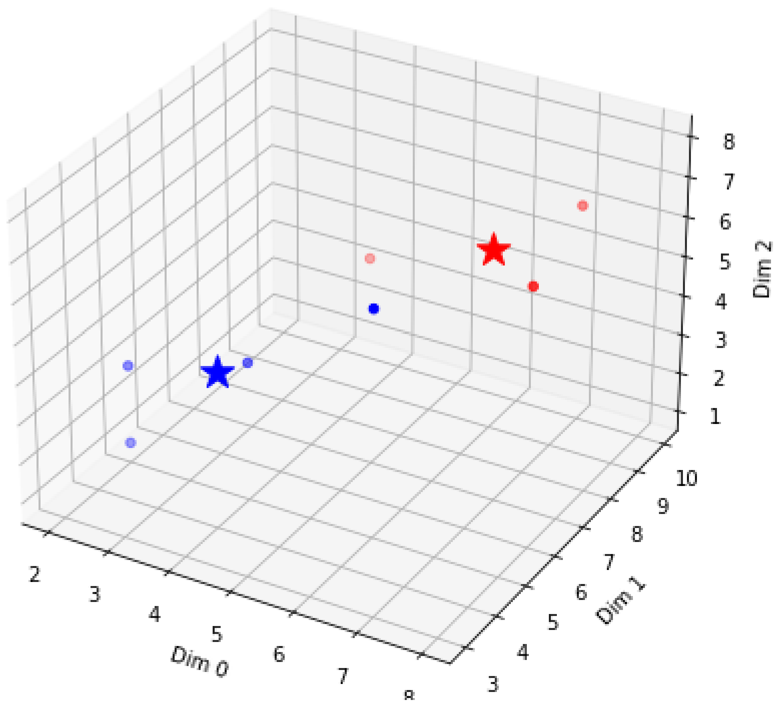
```
# Construct rank-2 approximation using cluster
centroids_transposed_2 = centroids_2.transpose()
W_2 = np.array([[1,0,0,1,1,0,0],
                [0,1,1,0,0,1,1]])
X_hat_2 = centroids_2@W_2 ## Fill in the blank
print('Rank-2 Approximation = \n', X_hat_2)
print('Difference between rank 2 approximation and original: \n',X-X_hat_2)
```

```
Rank-2 Approximation =
 [[6.33 3.75 3.75 6.33 6.33 3.75 3.75]
 [8.33 4.50 4.50 8.33 8.33 4.50 4.50]
 [5.67 4.00 4.00 5.67 5.67 4.00 4.00]
 [7.67 4.25 4.25 7.67 7.67 4.25 4.25]
 [6.33 4.00 4.00 6.33 6.33 4.00 4.00]]
Difference between rank 2 approximation and original:
 [[-2.33 3.25 -1.75 1.67 0.67 0.25 -1.75]
 [0.67 -1.50 0.50 -2.33 1.67 0.50 0.50]
 [-1.67 4.00 -1.00 1.33 0.33 0.00 -3.00]
 [1.33 -2.25 1.75 -2.67 1.33 0.75 -0.25]
 [-2.33 5.00 -2.00 1.67 0.67 0.00 -3.00]]
```

1 c) Play with the following code! You can pick three dimensions to look at by modifying `coordinates_to_plot`. Just have fun with it.

```
fig = plt.figure(figsize = (10, 7))
ax = plt.axes(projection ="3d")
coordinates_to_plot = [0,1,2]
color_array = np.array(['red', 'blue'])
ax.scatter3D(
            X[coordinates_to_plot[0],:], # x
            X[coordinates_to_plot[1],:], # y
            X[coordinates_to_plot[2],:], # y
            color=color_array[C_2] # color depends on cluster idx
            )
for i in range(2):
    ax.scatter3D(
            centroids_2[coordinates_to_plot[0],i], # x
            centroids_2[coordinates_to_plot[1],i], # y
            centroids_2[coordinates_to_plot[2],i], # y
            marker='*', # star instead of circle
            s=300, # size
            color=color_array[i] # color
            )

ax.set_xlabel('Dim %d'%coordinates_to_plot[0])
ax.set_ylabel('Dim %d'%coordinates_to_plot[1])
ax.set_zlabel('Dim %d'%coordinates_to_plot[2])
```

```
    Text(0.5, 0, 'Dim 2')
```



1 d) Repeat a)--c) with $K = 3$.

```
centroids_3, C_3 = kMeans(X, 3) ## Fill in the blank
print('centroids = \n', centroids_3)
print('centroid assignment = \n', C_3)
```

```
    centroids =
     [[7.00 7.50 3.00]
     [3.00 8.00 6.00]
     [8.00 6.50 3.00]
     [2.00 7.00 6.00]
     [9.00 7.50 2.75]]
    centroid assignment =
     [2 0 2 1 1 2 2]
```

```
# Construct rank-3 approximation using cluster
centroids_transposed_3 = centroids_3.transpose()
W_3 = np.array([[0,1,0,0,0,0,0],
                [0,0,0,1,1,0,0],
                [1,0,1,0,0,1,1]])
X_hat_3 = centroids_3@W_3 ## Fill in the blank
print('Rank-3 Approximation = \n', X_hat_3)
print('Difference between rank 3 approximation and original: \n',X-X_hat_3)
```

```
    Rank-3 Approximation =
     [[3.00 7.00 3.00 7.50 7.50 3.00 3.00]
     [6.00 3.00 6.00 8.00 8.00 6.00 6.00]
     [3.00 8.00 3.00 6.50 6.50 3.00 3.00]
     [6.00 2.00 6.00 7.00 7.00 6.00 6.00]
     [2.75 9.00 2.75 7.50 7.50 2.75 2.75]]
    Difference between rank 3 approximation and original:
     [[1.00 0.00 -1.00 0.50 -0.50 1.00 -1.00]
     [3.00 0.00 -1.00 -2.00 2.00 -1.00 -1.00]
     [1.00 0.00 0.00 0.50 -0.50 1.00 -2.00]
     [3.00 0.00 0.00 -2.00 2.00 -1.00 -2.00]
     [1.25 0.00 -0.75 0.50 -0.50 1.25 -1.75]]
```

```
fig = plt.figure(figsize = (10, 7))
ax = plt.axes(projection ="3d")
coordinates_to_plot = [0,1,2]
color_array = np.array(['red', 'blue', 'green'])
ax.scatter3D(
            X[coordinates_to_plot[0],:], # x
            X[coordinates_to_plot[1],:], # y
            X[coordinates_to_plot[2],:], # y
            color=color_array[C_3] # color depends on cluster idx
            )
for i in range(3):
    ax.scatter3D(
            centroids_3[coordinates_to_plot[0],i], # x
            centroids_3[coordinates_to_plot[1],i], # y
            centroids_3[coordinates_to_plot[2],i], # y
            marker='*', # star instead of circle
            s=300, # size
            color=color_array[i] # color
            )

ax.set_xlabel('Dim %d'%coordinates_to_plot[0])
ax.set_ylabel('Dim %d'%coordinates_to_plot[1])
ax.set_zlabel('Dim %d'%coordinates_to_plot[2])
```

```
Text(0.5, 0, 'Dim 2')
```

1 e) SVD can be also used to find T and W such that $X \approx TW$. Assume that you are given the SVD of X, i.e., $X = USV^T$. Find SVD-based T and W as a function of U, S, V (In an equation form, not numbers.) Recall that T is a 5-by-r matrix with orthonormal columns.

Your answer goes here:

$T = U.$

$W = SV^T.$

The matrix U has the singular vectors, which are an orthonormal basis for the columns of X. Since T is orthonormal, the matrix U will work out. Then W has singular values which are the weights for the orthonormal basis, which then will multiply out to create an approximation of X.

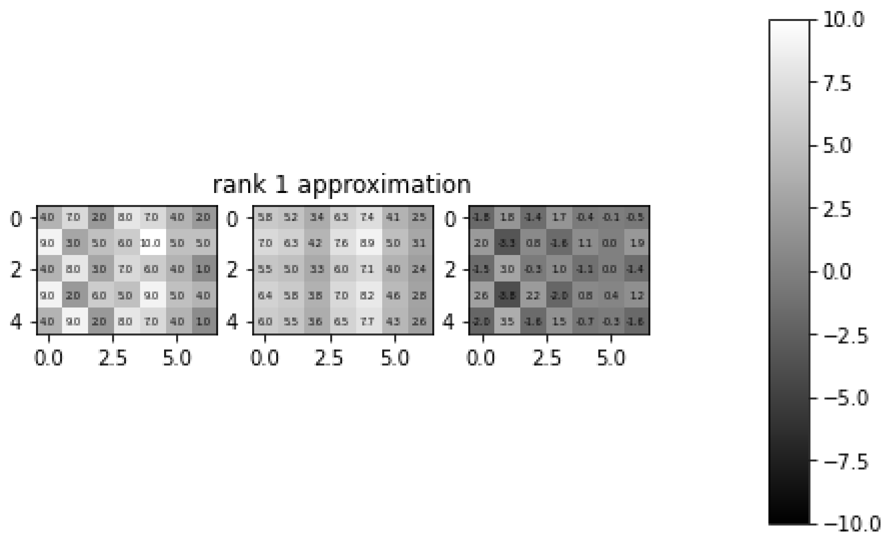
1 f) Find T, W and the rank-r approximation to X for $r = 2$. What aspects of the ratings does the first taste vector capture? What about the second taste vector?

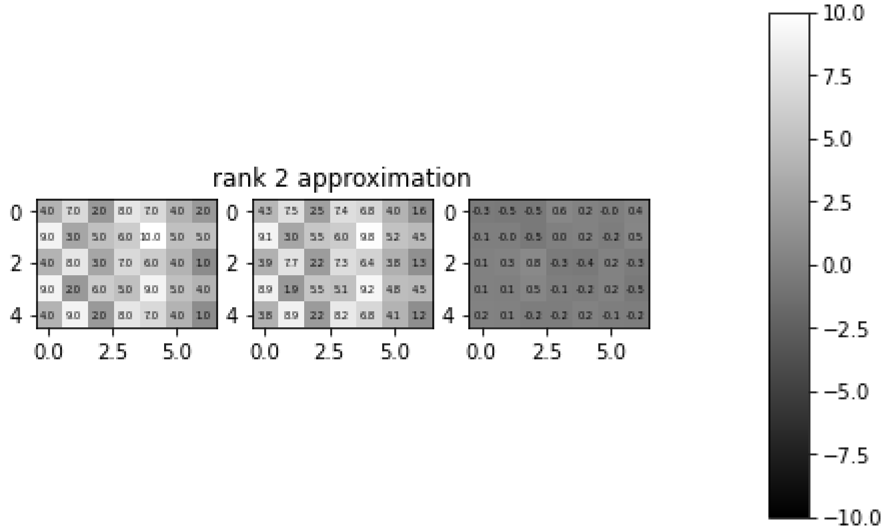
```
U, s, VT = np.linalg.svd(X, full_matrices=True)
S_matrix = np.zeros_like(X)
np.fill_diagonal(S_matrix, s)

## Fill in the blank using U, S_matrix, and VT
T = U
W = S_matrix@VT

for r in range(0,2):
    T_r = T[:,0:r+1] ## Choose the first r columns of T
    W_r = W[0:r+1,:] ## Choose the first r rows of W
    print(T_r)
    print(W_r)
```

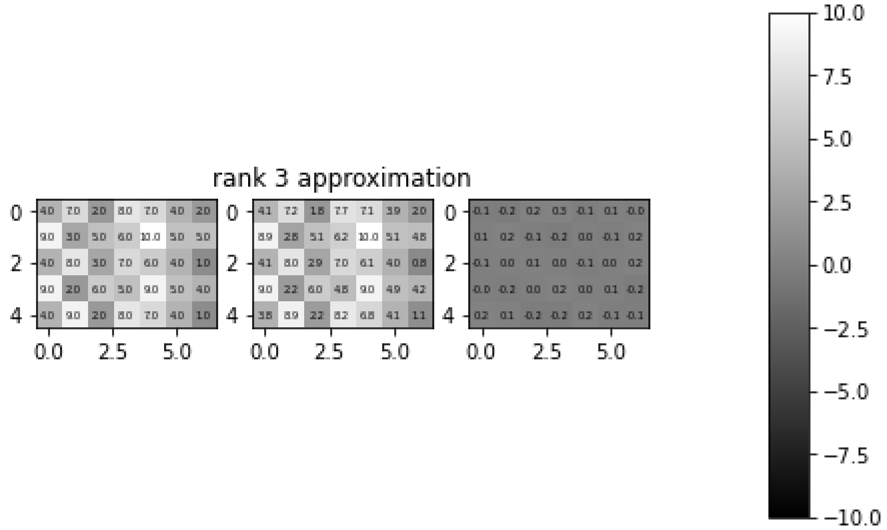
```
[[-0.42]
 [-0.51]
 [-0.40]
 [-0.47]
 [-0.44]]
[[-13.77 -12.52 -8.24 -15.02 -17.65 -9.89 -6.07]]
[[-0.42 -0.32]
 [-0.51 0.47]
 [-0.40 -0.37]
 [-0.47 0.55]
 [-0.44 -0.48]]
[[-13.77 -12.52 -8.24 -15.02 -17.65 -9.89 -6.07]
 [4.49 -7.06 2.93 -3.46 1.80 0.40 3.06]]
```

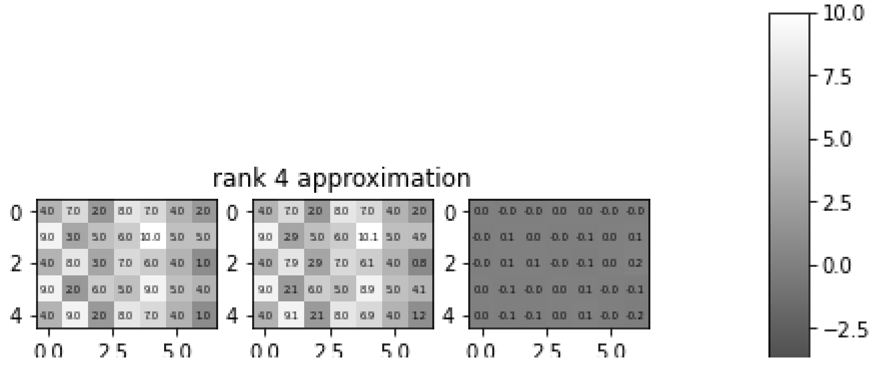
1 g) The following code visualizes the rank-r approximation for an increasing value of r. When does the approximation become exact? Why?

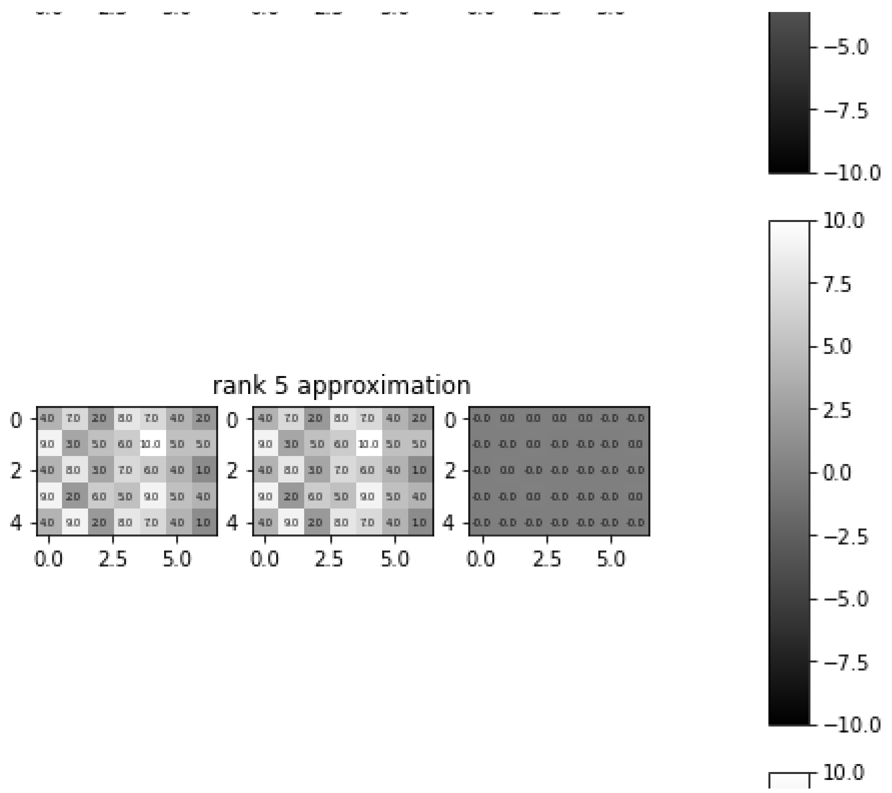
```
for r in range(0,7):
    T_r = T[:,0:r+1] ## Choose the first r columns of T
    W_r = W[0:r+1,:] ## Choose the first r rows of W
    X_rank_r_approx = T_r@W_r
    fig, ax = plt.subplots(1,3,figsize=(5.5, 5))
    for (j,i),label in np.ndenumerate(X):
        ax[0].text(i,j,np.round(label,1),ha='center',va='center', size=5)
    im = ax[0].imshow(X, vmin=-10, vmax=10, interpolation='none', cmap='gray')
    for (j,i),label in np.ndenumerate(X_rank_r_approx):
        ax[1].text(i,j,np.round(label,1),ha='center',va='center', size=5)
    im = ax[1].imshow(X_rank_r_approx, vmin=-10, vmax=10, interpolation='none', cmap='gray')
    for (j,i),label in np.ndenumerate(X-X_rank_r_approx):
        ax[2].text(i,j,np.round(label,1),ha='center',va='center', size=5)
    im = ax[2].imshow(X-X_rank_r_approx, vmin=-10, vmax=10, interpolation='none', cmap='gray')

    cbar_ax = fig.add_axes([1.05, 0.15, 0.05, 0.7])
    fig.colorbar(im, cax=cbar_ax)
    ax[1].set_title("rank %d approximation" % (r+1))
```


rank 1 approximation


rank 2 approximation


rank 3 approximation


rank 4 approximation



At rank 5, the approximation becomes exact. Since there are only 5 movies to estimate, there only needs to be 5 columns before the correct weights are found. The rank of the matrix A is 5, and the other two columns will be combinations of the 5 because there is only 5 features.

1 g) Your friend Jon rates Star Trek 6 and Pride and Prejudice 4. Assume a two-column taste matrix T. Formulate a system of equations that can find Jon's weight vector. Write down the least square solution.

Your answer goes here:

$X = Tw$

We only know 2 ratings, so take out the first 2 ratings columns from the taste matrix and then run least squares on that. X then also only has the two known ratings.

least squares: $w = (T_2^T T_2)^{-1} T_2^T X_2$

1 h) Using this weight vector, how can we predict Jon's ratings for all five movies, including the remaining three movies?

Your answer goes here:

Multiply the weights by the old taste matrix to get the predicted ratings for Jon.

1 i) Predict Jon's ratings for all the five movies with different choices of the taste matrix.

- Choice 1: T is the two centroids of the K-means result with $K = 2$
- Choice 2: T is the first two centroids of the K-means result with $K = 3$
- Choice 3: T is the first two SVD-based taste vectors

```
y = np.array([[6],[4]])

## Choice 1: K-means (K=2) based taste matrix T
K2,c = kMeans(X, 2)
T = K2[:,0:2] # fill in the blank
T_12 = T[0:2,:]
print(T@np.linalg.inv(T_12.T@T_12)@T_12.T@y)

## Choice 2: K-means (K=3) based taste matrix T
K3,c = kMeans(X, 3)
T = K3[:,0:2] # fill in the blank
T_12 = T[0:2,:]
print(T@np.linalg.inv(T_12.T@T_12)@T_12.T@y)

## Choice 3: SVD based taste matrix T
TSVD = U
T = TSVD[:,0:2] # fill in the blank
T_12 = T[0:2,:]
print(T@np.linalg.inv(T_12.T@T_12)@T_12.T@y)
```

```
[[6.00]
 [4.00]
 [5.98]
 [3.23]
 [6.74]]
[[6.00]
 [4.00]
 [4.64]
 [2.64]
 [6.87]]
[[6.00]
 [4.00]
 [6.01]
 [3.23]
 [6.83]]
```