

EECS 545

Homework One

1) Linear Algebra

(i) True. Let A denotes the symmetric matrix mentioned.

$$\therefore A^T = A$$

$$\Rightarrow (A^{-1})^T = (A^T)^{-1} = A^{-1}$$

A^{-1} is a symmetric matrix.

(ii) True. Let $A = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$ denotes the 2×2 orthogonal matrix

$$\Rightarrow A^T A = A A^T = I$$

$$\therefore \begin{bmatrix} a & c \\ b & d \end{bmatrix} \cdot \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\therefore \begin{cases} a^2 + b^2 = 1 \\ ac + bd = 0 \\ c^2 + d^2 = 1 \end{cases}$$

And there are two solutions of the equation above:

$$\textcircled{1} \quad \begin{cases} a = \cos \alpha \\ b = \sin \alpha \\ c = \sin \alpha \\ d = -\cos \alpha \end{cases}$$

$$\textcircled{2} \quad \begin{cases} a = \cos \alpha \\ b = \sin \alpha \\ c = -\sin \alpha \\ d = \cos \alpha \end{cases}$$

$$\text{i.e.} \quad \begin{bmatrix} \cos \alpha & \sin \alpha \\ \sin \alpha & -\cos \alpha \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix}$$

(iii) False.

$$C = \begin{bmatrix} \vdots & c_1^T & \vdots \\ \vdots & c_2^T & \vdots \\ \vdots & c_3^T & \vdots \end{bmatrix} \quad C^T = \begin{bmatrix} \vdots & \vdots & \vdots \\ c_1 & c_2 & c_3 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

$$CC^T = \begin{bmatrix} c_1^T c_1 & \vdots & \vdots \\ \vdots & c_2^T c_2 & \vdots \\ \vdots & \vdots & c_3^T c_3 \end{bmatrix}$$

$$c_i^T c_i = c_i^2 \geq 0.$$

But in $A = \begin{bmatrix} -8 & -1 & -6 \\ -3 & -5 & -7 \\ -4 & -1 & -2 \end{bmatrix}$

which diagonal cannot be all non-negative.

So, A can't be written as $A = CC^T$.

(iii)

2) Probability

$$\begin{aligned} (i) \quad E_Y[E_X(X|Y)] &= E_Y\left[\sum_x x \cdot P(X=x|Y)\right] \\ &= \sum_y \left[\sum_x x \cdot P(X=x|Y=y)\right] P(Y=y) \\ &= \sum_x x \sum_y P(X=x, Y=y) \\ &= \sum_x x \cdot P(X=x) \\ &= E[X] \end{aligned}$$

$$(ii) \quad I(X \in C) = \begin{cases} 1 & \text{if } X \in C \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} E[I(X \in C)] &= \sum_{i=0}^1 i \cdot P(I(X \in C) = i) \\ &= 0 \cdot P(I(X \in C) = 0) + 1 \cdot P(I(X \in C) = 1) \\ &= P(I(X \in C) = 1) \\ &= P(X \in C) \end{aligned}$$

$$(iii) \quad \begin{aligned} \text{Var}(X) &= E(X^2) - (E(X))^2 \\ \text{Var}(X|Y) &= E(X^2|Y) - (E(X|Y))^2 \end{aligned}$$

$$\begin{aligned} E_Y[\text{Var}_X(X|Y)] + \text{Var}_Y[E_X(X|Y)] &= E_Y[E_X(X^2|Y) - E_X(X|Y) \cdot E_X(X|Y)] + [E(E(X|Y)^2) - (E(E(X|Y)))^2] \\ &= E(X^2) - E(X)^2 + (E(X)^2 - (E(X))^2) \\ &= E(X^2) - (E(X))^2 = \text{Var}(X) \end{aligned}$$

(iv) X and Y are independent i.e. $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$

$$\begin{aligned} E(XY) &= \iint f_{XY}(x,y) dx dy \\ &= \iint f_X(x) f_Y(y) dx dy \\ &= \int f_X(x) dx \cdot \int f_Y(y) dy \\ &= E(X) \cdot E(Y) \end{aligned}$$

$$IV) E(X) = 0 \cdot P(X=0) + 1 \cdot P(X=1) = P(X=1)$$

$$\text{So, } E(Y) = P(Y=1)$$

$$\text{Fix } E(XY) = P(X=1, Y=1) = P(X=1) \cdot P(Y=1)$$

$$P(X=1) = P(X=1, Y=0) + P(X=1, Y=1) = P(X=1, Y=0) + P(X=1) \cdot P(Y=1)$$

$$\text{Let } p \text{ denotes } P(X=1), \text{ } q \text{ denotes } P(Y=1)$$

$$\begin{aligned} P(X=1, Y=0) &= P(X=1) - P(X=1)P(Y=1) \\ &= p - pq = p(1-q) = P(X=1) \cdot P(Y=0) \end{aligned}$$

$$\begin{aligned} P(X=0, Y=0) &= 1 - pq - p(1-q) - q(1-p) \\ &= 1 - pq - p + pq - q + pq \\ &= (1-p)(1-q) \\ &= P(X=0)P(Y=0) \end{aligned}$$

$$P(X=0, Y=1) = P(X=0)P(Y=1)$$

$$1b) \quad i) \leq . P(H=h, D=d) = \frac{P(H=h \cap D=d)}{P(H=h)} \leq P(H=h)$$

$$ii) ? \text{ depend } P(H=h | D=d) = \frac{P(H=h, D=d)}{P(D=d)}$$

$$iii) \geq . LHS = \frac{P(H=h, D=d)}{P(D=d)}$$

$$RHS = \frac{P(D=d, H=h)}{P(H=h)} \cdot P(H=h) = P(H=h, D=d)$$

$$P(D=d) \leq 1.$$

$$P(H=h | D=d) \geq P(D=d | H=h) P(H=h)$$

(b) (a) ① A is PSD $\Rightarrow \lambda_i \geq 0$.

$$A = \sum_{i=1}^d \lambda_i u_i u_i^T \quad x^T A x \geq 0.$$

$$\begin{aligned} x^T A x &= \sum_{i=1}^d \lambda_i (x^T u_i)(u_i^T x) \\ &= \sum_{i=1}^d \lambda_i (x^T u_i)^2. \end{aligned}$$

$$\sum_{i=1}^d \lambda_i (x^T u_i)^2 \geq 0. \quad x^T u_i \geq 0$$

$$\lambda_i \geq 0 \quad \forall i, i = 1, 2, \dots, d$$

② $\lambda_i \geq 0 \Rightarrow A$ is PSD.

Since $\lambda_i \geq 0$.

$$\text{then } x^T A x = \sum_{i=1}^d \lambda_i (x^T u_i)^2 \geq 0.$$

A is PSD.

(b) ① A is PD $\Rightarrow \lambda_i > 0$.

$$A = \sum_{i=1}^d \lambda_i u_i u_i^T, \quad A \text{ is PD } \Rightarrow x^T A x > 0$$

$$x^T A x = \sum_{i=1}^d \lambda_i (x^T u_i)(u_i^T x) = \sum_{i=1}^d \lambda_i (x^T u_i)^2 > 0$$

$$\text{Since } x^T u_i \geq 0, \quad \lambda_i > 0.$$

② $\lambda_i > 0 \Rightarrow A$ is PD

Since $\lambda_i > 0$,

$$x^T A x = \sum_{i=1}^d \lambda_i (x^T u_i)^2 > 0$$

A is PD.

III:) False.

$$C = \begin{bmatrix} \vdots & c_1^T & \vdots \\ \vdots & c_2^T & \vdots \\ \vdots & c_3^T & \vdots \end{bmatrix} \quad C^T = \begin{bmatrix} c_1 & c_2 & c_3 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

$$CC^T = \begin{bmatrix} c_1^T c_1 & \vdots & \vdots \\ \vdots & c_2^T c_2 & \vdots \\ \vdots & \vdots & c_3^T c_3 \end{bmatrix}$$

$$c_i^T c_i = c_i^2 \geq 0.$$

But in $A = \begin{bmatrix} -8 & -1 & -6 \\ -3 & -5 & -7 \\ -4 & -1 & -2 \end{bmatrix}$

which diagonal cannot be all non-negative.

So, A can't be written as $A = CC^T$.

(iii) False.

If A could be written as $A = CC^T$

Then $(\text{eig}(C)) = \text{eig}(CC^T) = \text{eig}(C^T C) = \text{eig}(A)$.

if A have
eigen decomposition

$$\Rightarrow A = U \Lambda U^T$$

$$= (U \sqrt{\Lambda}) (\sqrt{\Lambda} U^T)$$

$$= (U \sqrt{\Lambda}) (U^* \sqrt{\Lambda})^T$$

$$= (U \sqrt{\Lambda}) (U \sqrt{\Lambda})^T$$

Since Λ is diagonal.

$$= CC^T$$

But $A = \begin{bmatrix} -8 & -1 & -6 \\ -3 & -5 & -7 \\ -4 & -9 & -2 \end{bmatrix}$ apparently not a symmetric

If A could be written as CC^T ,

then A could be written as $(U \sqrt{\Lambda}) (U \sqrt{\Lambda})^T$ i.e., $U \Lambda U^T$

which means A needs to be symmetric.

So, the assumption is not right.

A can not be written as $A = CC^T$.

$$4. (a) \quad f(x) = a^T x + b$$

$$f(y) = a^T y + b$$

$$f(tx + (1-t)y) = a^T(tx + (1-t)y) + b$$

$$= t a^T x + (1-t) a^T y + b$$

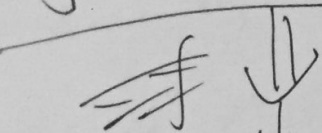
$$= t a^T x + (1-t) a^T y + (t + 1 - t)b$$

$$= t(a^T x + b) + (1-t)(a^T y + b)$$

$$= t f(x) + (1-t) f(y)$$

$$\leq t f(x) + (1-t) f(y) \Rightarrow \text{Convex}$$

$$-f(tx + (1-t)y) = -(t f(x) + (1-t) f(y))$$



Same as, this is convex

$\therefore f(tx + (1-t)y)$ is concave.

Since $f(tx + (1-t)y) = t f(x) + (1-t) f(y)$

Not strictly inequality.

$f(x)$ is not strictly convex.

(b) f is strictly convex.

$$\text{then } f(tx + (1-t)y) < hf(x_1) + (1-h)f(x_2), t \in (0,1)$$

Suppose x_1 is global minimizer, and x_2 is also a minimizer.

$$\text{where } f(x_1) \leq f(x_2)$$

$$\text{Since } t \in (0,1) \Rightarrow tf(x_1) + (1-t)f(x_2) \leq tf(x_2) + (1-t)f(x_2)$$

$$tf(x_1) + (1-t)f(x_2) \leq (1-t)f(x_2)$$

$$f(tx_1 + (1-t)x_2) < tf(x_1) + (1-t)f(x_2) \leq f(x_2)$$

$$f(tx_1 + (1-t)x_2) < f(x_2)$$

We assume x_2 is one of the global minimizer,

i.e. $f(x) \geq f(x_2)$, which exactly the opposite of inequality shows ~~be~~ above.

So, ~~there is~~ f at most has one global minimizer.

d ① f convex $\rightarrow \nabla^2 f(x) \succeq 0$.

$$f(x+\lambda d) = f(x) + \nabla f(x)^T(\lambda d) + \frac{1}{2}(\lambda d)^T H(x)(\lambda d) + o(\|\lambda d\|^2)$$

$$\because f(x+\lambda d) \geq f(x) + \langle \nabla f(x), \lambda d \rangle$$

$$\therefore \lambda^2 \left(\frac{1}{2} d^T H(x) d + \frac{o(\|\lambda d\|^2)}{\lambda^2} \right) \geq 0.$$

$$\lambda^2 \geq 0 \text{ and } \frac{o(\|\lambda d\|^2)}{\lambda^2} \geq 0.$$

$$\therefore d^T H(x) d \geq 0. \quad \text{i.e. } \nabla^2 f(x) \succeq 0$$

② $\nabla^2 f(x) \succeq 0 \rightarrow f$ convex

$$f(y) = f(x) + \nabla f(x)^T(y-x) + \frac{1}{2}(y-x)^T H(z)(y-x)$$

\downarrow
 $t(y-x) + y$, where $t \in (0,1)$

$$\therefore H(z) \succeq 0.$$

$$\therefore f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle$$

$$\therefore f \text{ is convex}$$

$$10) \quad f(x) = \frac{1}{2} x^T A x + b^T x + c.$$

$$\nabla f(x) = \frac{1}{2} A^T x + \frac{1}{2} A x + b$$

$$= \frac{1}{2} (A^T + A) x + b$$

$$= A x + b.$$

$$\nabla^2 f(x) = A.$$

if A is a PSD matrix, then f is convex.

if A is a PD matrix, then f is strictly convex.

1c) h denotes to an arbitrary vector. $\|h\|=1$.

use suggestion $f(x^*+th) = f(x^*) + \langle \nabla f(x^*), th \rangle + \frac{1}{2} \langle th, \nabla^2 f(x^*) th \rangle + O(t^3)$

$$\begin{cases} X = X^* + th \\ Y = X^* \end{cases} \quad X - Y = th.$$

$$f(x^*+th) = f(x^*) + (th)^T \nabla f(x^*) + \frac{1}{2} (\nabla^2 f(x^*) th)^T th + O(t^3)$$

$$f(x^*+th) - f(x^*) = th^T \nabla f(x^*) + \frac{1}{2} t^2 h^T \nabla^2 f(x^*) h + O(t^3)$$

Since $f(x^*)$ is local minimizer.

$$f(x^*+th) - f(x^*) \geq 0.$$

$$\therefore \frac{1}{2} t^2 h^T \nabla^2 f(x^*) h + O(t^3) \geq 0$$

$$h^T \nabla^2 f(x^*) h \geq 0.$$

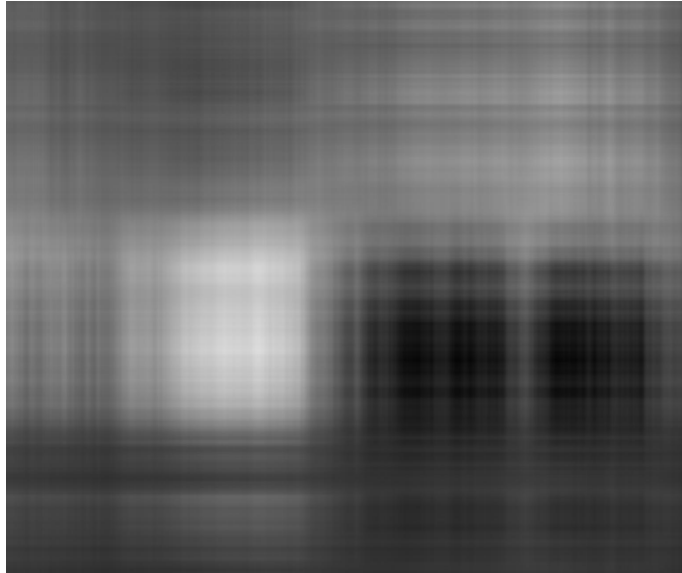
$$\nabla^2 f(x^*) \geq 0$$

$\nabla^2 f(x^*)$ is \neq positive semidefinite.

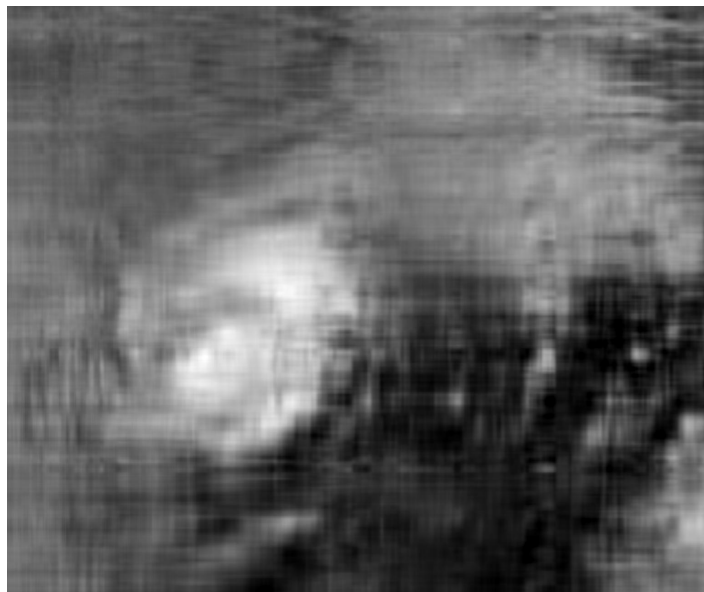
5. Programming

(a) After performing SVD on this matrix and zero out all but top $k(2,10,40)$ singular values to form an approximation. Result images are shown as below:

- $k = 2$



- $k = 10$



- $K = 40$



result = [0.28258451691929531, 0.15928758406875393, 0.084079335040156966]

when $k = 2$, result = 0.28258451691929531

when $k = 10$, result = 0.15928758406875393

when $k = 40$, result = 0.084079335040156966

(b) How many numbers do you need to describe the approximation

Result = [5690, 28450, 113800]

when $k = 2$, result = 5690

when $k = 10$, result = 28450

when $k = 40$, result = 113800

My code is shown as below:

```
import numpy as np
from PIL import Image
```

```
##### Part A #####
# Load Image
image = Image.open('harvey-saturday-goes7am.jpg')
grey = image.convert("L")

# Image to array
X = np.asarray(grey)
```

```

# k value
k_list = [2,10,40]
result_list = []
partb_list = []

for k in k_list:
    # SVD
    U,s,Vt = np.linalg.svd(X, full_matrices = False)
    s[k:] = 0
    S = np.diag(s)
    X_app = np.dot(np.dot(U,S), Vt)

    # Show and save approximate_image
    img = Image.fromarray(X_app)
    if(img.mode != 'RGB'):
        img = img.convert('RGB')

    img.save(str(k)+'.jpg')

    # calculate  $\|X-X_{app}\|_f / \|X\|_f$ 
    temp = X - X_app
    result = np.linalg.norm(temp,'fro') / np.linalg.norm(X,'fro')
    result_list.append(result)

##### Part B #####
# How many numbers do you need to describe the approximation
left = U[:, :k]
left_row = left.shape[0]
left_column = left.shape[1]
num_left = left_row * left_column

right = Vt[k, :]
right_row = right.shape[0]
right_column = right.shape[1]
num_right = right_row * right_column

partb = num_left + num_right + k
partb_list.append(partb)

```

