Q1.

1. To prove the variance of the $u_1^T x$

$$var(u_1^T x) = \frac{1}{n} u_1^T x c c^T x^T u_1$$

Here we have, c, a centering matrix

$$c = I - \frac{11^T}{n}$$

And a identity matrix I,

1 -> which is an n-dimention vector with has 1 as all of its elements.

the x is whole data we have where,

$$x = [x_1, x_2 \cdots x_n] \in R^{d \times n}$$

with d no. of features

Here u_1 is the first principle component ie, projection of 'n' number of the data samples on the principle component u_1.

Normally the formula for the variance

$$var(x) = E[(x - \mu)^2]$$

μ is the mean as we write is as $\bar{x}$ also

$$= \cancel{E}\ \cancel{(x - \bar{x})}^2$$

$$\cancel{E}\ \cancel{(x - \bar{x})}^2$$

~~var~~

we know that the expected ~~value~~ value, E

$E[x] = \sum_{i=1}^{n} x_i P_i$, we take it to be if x

$\mu = \sum_{i=1}^{n} \frac{x_i}{n}$ -> mean

so, now, $var(x) = \frac{1}{n} \sum (x - \bar{x})^2$

we can write $(x - \bar{x})^2$ as $(x - \bar{x})(x - \bar{x})^T$
as $(x - \bar{x})^2$ will be same as $(x - \bar{x})^T \cdot (x - \bar{x})$
when we compute the values.

Now in the place of x for var(x) we use the
$u_1^T x$.

~~$var(u_1^T x) = \frac{1}{n} u_1^T x \, I$~~

we consider $u_1^T x c$ from the Equation
of the variance

$$u_1^T x c = \begin{bmatrix} u_1^T x_1 \\ u_1^T x_2 \\ u_1^T x_3 \\ \vdots \\ u_1^T x_n \end{bmatrix} \cdot \left[I_{n \times n} - \frac{11^T_{n \times n}}{n} \right]$$

$$= \begin{bmatrix} u_1^T x_1 \, I - \dfrac{u_1^T x_1 \, 1 1^T}{n} \\ u_1^T x_2 \, I - \dfrac{u_1^T x_2 \, 1 1^T}{n} \\ \vdots \\ u_1^T x_n \, I - \dfrac{u_1^T x_n \, 1 1^T}{n} \end{bmatrix}$$

we say I as the identity matrix which has 1 in its diagonal positions and zero's in the remaining,

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}_{3\times3}$$

An 1 & 1^T are a vectors of n dimentions with all ones as its values

$$1 \rightarrow \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \end{bmatrix}_n$$

$$1^T \rightarrow \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_n$$

$$u_1^T X_C = \begin{bmatrix} u_1^T x_1 I - \dfrac{u_1^T x_1 11^T}{n} \\ u_1^T x_2 I - \dfrac{u_1^T x_2 11^T}{n} \\ \vdots \\ u_1^T x_n I - \dfrac{u_1^T x_n 11^T}{n} \end{bmatrix}$$

$$= \begin{bmatrix} u_1^T x_1 \left(I - \dfrac{11^T}{n}\right) \\ u_1^T x_2 \left(I - \dfrac{11^T}{n}\right) \\ \vdots \\ u_1^T x_n \left(I - \dfrac{11^T}{n}\right) \end{bmatrix} = \begin{bmatrix} u_1^T x_1 C \\ u_1^T x_2 C \\ \vdots \\ u_1^T x_n C \end{bmatrix}$$

we know $x_1 x_2 \cdots x_n \rightarrow X$

$\Rightarrow u_1^T X_C$

Now we use this in the Equation in the variance format.

$$\text{var}(x) = \frac{1}{n}(x - \bar{x})(x - \bar{x})^T$$

$$var(u_1^T x) = \frac{1}{n}\left(u_1^T x I - u_1^T x \frac{11^T}{n}\right)\left(\frac{u_1 x I - u_1 x 11^T}{n}\right)^T$$

we also compute for $u_1 x^T c^T$ as similar to

$$u_1^T x c^T \begin{bmatrix} u_1 x_1^T I - u_1 x_1^T \frac{11^T}{n} \\ u_2 x_2^T I^T - u_1 x_2^T \frac{11^T}{n} \\ \vdots \\ u_n x_n^T I^T - u_1 x_n^T \frac{11^T}{n} \end{bmatrix}$$

$$= \begin{bmatrix} u_1 x_1^T \left(I - \frac{11^T}{n}\right)^T \\ u_2 x_2^T \left(I - \frac{11^T}{n}\right)^T \\ \vdots \\ u_1 x_n^T \left(I - \frac{11^T}{n}\right)^T \end{bmatrix} = \begin{bmatrix} u_1 x_1^T c^T \\ u_2 x_2^T c^T \\ \vdots \\ u_1 x_n^T c^T \end{bmatrix}$$

now with
$u_1^T x$ in the expanded definition of
variance

$$var(u_1^T x) = \frac{1}{n}\left(u_1^T x I - u_1^T x \frac{11^T}{n}\right)\left(u_1 x^T I - u_1 x^T \frac{11^T}{n}\right.$$

$$n \cdot \mathrm{var}(u_1^T x)\left(u_1^T x I - u_1^T x \frac{11^T}{n}\right) \cdot$$

$$\left(u_1^T x I - u_1^T x \frac{11^T}{n}\right)^T$$

→ we got this by expanding var(x) with $u_1^T x$ & using the centering matrix.

$$\mathrm{var}(u_1^T x) = \frac{1}{n}\left(u_1^T x\left(I - \frac{11^T}{n}\right)\right) \cdot \left(u_1^T x I - u_1^T x \frac{11^T}{n}\right)^T$$

$$= \frac{1}{n} u_1^T x c \cdot u_1 x^T \left(I - \frac{11^T}{n}\right)^T$$

$$= \frac{1}{n} u_1^T x c\, u_1 x^T c^T$$

$$\Rightarrow \frac{1}{n} u_1^T x c\, x^T u_1$$

→ $c \cdot c^T$ will be c as it is the centering matrix and the matrix multiply by its transpose with is same as c

so $c \cdot c^T = c$ →

$$\Rightarrow \frac{1}{n} u_1^T x c c^T x^T u_1$$

$$\mathrm{var}(u_1^T x) = \frac{1}{n} u_1^T x c c^T x^T u_1$$

Q.1
1.2. To prove the covariance S. as

$cov = X C C^T X^T$

where X is the data with d-dimention

we usually know that the covariance of x, y

$cov(x,y) = E[(x - \mu_x)(y - \mu_y)]$

if we expand or solve it further,

$cov(x,y) = E[x]E[y] - E[xy]$

we know that E is the expected value so we can say it as mean.

$cov(x,y) = \mu_x \mu_y - E[xy]$

~~...~~ if and usually covariance of a variable for it self is same as the variance we say var = s^2, Now we say the covariance of whole data

$$cov(x,y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

if we see the covariance matrix, all its ~~this~~ diagonal elements will be the square of its value.

we know the C is the centering matrix

$$C = I - \frac{11^T}{n}$$

& the X -> will be the whole data with d number of dimentions.

~~we can say covariance as $S = \frac{1}{n}$~~

C, the centering matrix when transposed and multiplied we get the same centering matrix C.

XC -> will be the centered matrix.

we say $S = \frac{1}{n} XCX^T$

As the value y_n is a constant we can ignore it.

And as we said that $c \cdot c^T$ will be the same value c' as the centering matrix.

finally $S = \frac{1}{n} X C X^T \Rightarrow X C X^T$

$S \Rightarrow X C C^T X^T$

Q2.)

Non-Negitive Matrix Factorization

- The NMF is kind of Algorithm used in multivariant analysis.
- In this we the original matrix V which is also Non-negative is factorized into 2 differ and smaller matrices W & H, where,

$V \rightarrow d \times n$

$W \rightarrow d \times K$ & $H \rightarrow K \times N$

→ we keep on changing the values of W & H for the optimal solution using gradient decent techniques.

→ As we are minimizing the W & h we formulate NMF as minimizing problem.

$$\min_{W,H} : \| X - WH \|_F^2$$

where $X = [x_1 \cdots x_n] \in R^{d \times n}$

$W \in R^{d \times K}, H \in R^{K \times n}$

s.t., $W \geq 0$ & $H \geq 0$

→ Now, if we implement NMF using other algorithms like Projected gradient decent or Integrative co-ordinate decent,

→ First solving using Projected gradient decent

1. Initially we assign some positive values to W & H.

2. we compute the the gradient of the ~~objective~~ objective function for the values of W & H

3. After the 2nd step the W & H will be in the negative gradient direction, so we Project the W & H into the non-negative direction As we donot need the negative values. the following are the Equations for Projecting back and updating the

$$W' = W - \alpha \nabla f(W)$$
$$H' = H - \alpha \nabla f(H)$$

And for projecting it back we have.

$$W' = \underset{v \in C}{\arg\min} \|v - W^*\|$$
$$H' = \underset{v \in C}{\arg\min} \|v - H^*\|$$

4. Now we Repeat steps 2 & 3 until we reach the optimum solution.

→ Solving the Non-Negative matrix factorization using the integrated coordinate decent.

- this is a technique where we keep one of the factors i.e., W or H ~~as~~ as constant and update the others in iterations.

we usually define the Non Negative matrix factorization as a minimizing problem.

$$\min_{W,H} : \|X - WH\|_F^2$$

$$St., W \geq 0, H \geq 0$$

- Initially we assign some values for W & H,
- The following the process of the updating W & H keeping the others constant:
- the updating of the W & H will be computed as.

→ First updating W keeping H constant

$$W^* = \frac{W(XH^T)}{(WHH^T)}$$

→ Now keeping W as constant and update H

~~$H^* =$~~

$$H^* = \frac{H(XW^T)}{(WW^TH)}$$

→ we minimize the problem using ~~minimizat~~ W & H → As we say

matrix → W X H