Team Members:

Priyanka Upadhyay, Email Id : s8prupad@stud.uni-saarland.de
Gopal Bhattrai, Email Id : gobh00001@stud.uni-saarland.de

2.1

(a) Covariance Matrix = X^TX = Square Matrix, Symmetric matrix

Say X^TX is a 2x2 matrix ⇒

$$X^TX = \begin{bmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{21} & \sigma_2^2\end{bmatrix}$$

σ_1^2, σ_2^2 represents the Variance of the Variable 1 and 2 respectively

σ_{12} and σ_{21} represents the Correlation between Variable 1 and 2. The Correlation ranges $[-1,1]$, The more close it is to $+1$ more positively Correlated. The more to 0, No Correlation & More close to -1 negatively Correlated.

$[\sigma_{12} = \sigma_{21}]$

(b) X^TX

- Symmetric ⇒ $[A^T = A]$

$(X^TX)^T \longrightarrow X^T(X^T)^T = X^TX$ hence Symmetric

$\because (AB)^T = B^TA^T$

- Square Matrix ⇒ Say X is a $n\times m$ matrix then X^T will be $m\times n$.

Hence $X_{n\times m} \cdot X^T_{m\times n} = (XX^T)_{n\times n}$ Square matrix

- Positive Semidefinite ⇒ The Energy i.e. $x^TCx \geq 0$ for any x

$\therefore v^T(x^Tx)v \geq 0$ here $\boxed{C = x^Tx}$

$(v^Tx^T)xv \geq 0$

$(\cancel{xv})^T$

$(xv)^T xv \geq 0 \quad \therefore \left[\|A\|_2^2 = A^T A \right]$

$\therefore \|xv\|_2^2 \geq 0$ which is always true since Square cant be a negative value

*Has Positive Eigen Values and Orthogonal Eigen Vectors ⇒

- Since the matrix is Positive definite, the matrix will have Positive Eigenvalues

Property → A real Symmetric matrix will always have real Eigen Values and Orthogonal Eigen Vectors.

Eigen Values ...

2.2 $X = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 0 & -1 & -2 & -3 & -4 \end{bmatrix}$

$X - \mu = \begin{bmatrix} 0-2 & 1-2 & 2-2 & 3-2 & 4-2 \\ 0-2 & -1-2 & -2+2 & -3+2 & -4+2 \end{bmatrix} = \underset{X}{\begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ -2 & -3 & 0 & -1 & -2 \end{bmatrix}}$

$XX^T = \underset{2\times5}{\begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ -2 & -3 & 0 & -1 & -2 \end{bmatrix}} \underset{5\times2}{\begin{bmatrix} -2 & -2 \\ -1 & -3 \\ 0 & 0 \\ 1 & -1 \\ 2 & -2 \end{bmatrix}} = \underset{2\times2}{\begin{bmatrix} 10 & -10 \\ -10 & 10 \end{bmatrix}}$

$\therefore a_{11} = -2\times-2 + (-1\times-1) + 0 + 1\times1 + 2 * +2$

$|A - \lambda I| = 0$

$\begin{vmatrix} 10-\lambda & -10 \\ -10 & 10-\lambda \end{vmatrix} \Rightarrow (10-\lambda)^2 - 100 = 0$

$(10-\lambda)^2 = 100$

$10 - \lambda = \pm 10$

$\lambda = 10 \pm 10 \Rightarrow$

$\boxed{\lambda = 0, 20}$

$\begin{bmatrix} 10 & -10 \\ -10 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

Eigen vector $\Rightarrow \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ ✓ $\longrightarrow V = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$

As they Satisfy the Equation

Correspond to 20 — Correspon to 0.

The Eigen Values tells the amount of Variance Captured by its Eigen Vector.

Hence Max Variance ie $PC1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $PC2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

(Corresponding to 20)

So the Reduced Dimensionality $\Rightarrow$

$$= \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}_{2\times2} \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ -2 & -3 & 0 & -1 & -2 \end{bmatrix}_{2\times3} = \begin{bmatrix} 4 & 2 & 0 & -2 & -4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$D \cdot X$$

(b) There can be different uses of PCA $\Rightarrow$

- For Visualization, plotting higher dimensional data using smaller dimensions.
- To reduce the dimensionality due to which Overfitting of model may decrease.
- It also removes linearly Correlated features i.e. Redundant features.
- Might Improve Perfomance.

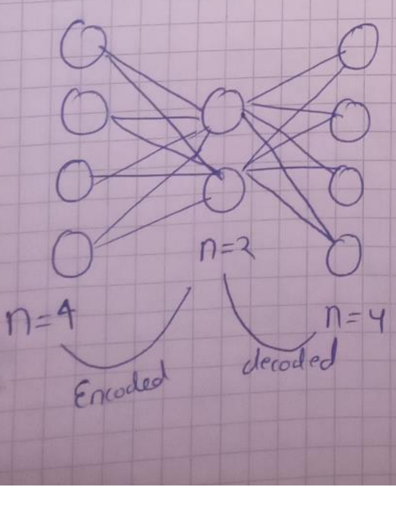
Disadvantage $\Rightarrow$

- PCA might loose information which might lead to poor Perfomance than before.
- Once Compressed, we cant get back exactly the dataset before
- Data Standardization is must before PCA.

• After doing PCA, Our dataset turns into Principal Components that are nothing but linear comb^n of your Original features so less Interpretable.

ii) Well, PCA sort of use the Covariance Matrix, i.e. it only takes linear Correlation into account. Its perfomance will not be good on Non linear dataset. But In practise, we apply PCA and then we see how much % of variance was encoded by Principal Components. If % is high means dataset was sort of linear and PCA encoded it but if very low, means data is sort of Non linear So applying PCA doesnt make sense there.

(C) Autoencoder ⇒ Is a NN approach of Compressing a high dimensional data to a lower dimension → Encoder and again Converting an Encoded lower dimension to original dimension, using NN technique.



• If we choose right Parameters of Neural Network, we can perform this Non linear transformation to smaller dimension and again back to Original dimension using appropiate weight matrix