NAME: EZE NNAMDI SHEDRACK
ANDREW ID: SEZE
COURSE: MATHEMATICAL FOUNDATIONS OF ML.
ASSIGNMENT 2:

Question 1:

Let $A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 2 & 4 & 1 & 3 \\ 3 & 6 & 2 & 5 \end{bmatrix}$

a. find a basis for the column space of A, (A). What is the rank (A)

<u>Solution</u>

$$A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 2 & 4 & 1 & 3 \\ 3 & 6 & 2 & 5 \end{bmatrix}$$

To find the column basis, we can reduce (A) to its row echelon form (REF) and count the number of pivot columns to know the rank while the pivot columns equals column basis.

$$A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 2 & 4 & 1 & 3 \\ 3 & 6 & 2 & 5 \end{bmatrix}$$

Pivot row 1 column 1 (R_{11}): $R_2 = R_2 - 2R_1$

$$R_2 = R_2 - 2R_1 \longrightarrow \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 3 & 6 & 2 & 5 \end{bmatrix}$$

$$R_3 = R_3 - 3R_1 \longrightarrow \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 2 \end{bmatrix}$$

Let $R_3' = R_3 - 2R_2 \longrightarrow \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Pivot colums = C_1 and C_3 (column 1 and 3)

Since columns 1 and three are the pivot columns, the column basis is the corresponding columns in the original matrix (A)

$$A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 2 & 4 & 1 & 3 \\ 3 & 6 & 2 & 5 \end{bmatrix} \text{ and } REF = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$C_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \; C_3 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \quad \text{Hence, column basis} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$$

Since we have two pivot columns, rank of $A = 2$

b. What is the null space of A (A):

<u>Solution</u>

Null space is equal all cases where $Ax = 0$

from our row echelon form (REF): $\begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

we have x_1, x_2, x_3, and x_4

We can set each row $= 0$: $x_1 + 2x_2 + x_4 = 0$ -- (I)

$x_3 + x_4 = 0$ -- (II)

Since columns 1 and 3 are pivot columns, then x_2 and x_4 from columns 2 and 4 are free variables, hence they can take any form,

Let $x_2 = s$ and $x_4 = t$

In equation (i) write x_1 in terms of s and t

$x_1 + 2x_2 + x_4 = 0$ where $x_2 = s$ and $x_4 = t$

$x_1 = 0 - 2s - t = -2s - t$

In equation (II) $x_3 = 0 - t$

hence: $x_1 = -2s - t$, $x_2 = s$, $x_3 = -t$, $x_4 = t$

finally $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2s-t \\ s \\ -t \\ t \end{bmatrix} \Rightarrow s \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, t \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$

Hence the basis for the null space $= \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$

C. Interpretation: If the matrix A represents a linear transformation, what does the column space represent? What does the null space represent? In the context of a linear system $Ax = b$, what does the null space tell you about the solution set.

<u>solution</u>

The column space in the matrix A represents all possible outputs vectors b that can be produced when the matrix A acts on any input vectors (x) in the domain $\mathbb{R}^4$. Since our rank is 2 then it means the transformation maps the 4-dimensional input space unto a 2-dimensional plane $\mathbb{R}^2$

The null space in the matrix A represents all inputs x that Produces zero outputs since $A \neq 0$ and $Ax = 0$. This also can be called the kernel of matrix A Linear transformation. Since the dimension of nullity is number of non zero rows (2), then we can say the redundancy in the transformation equals to 2.

Finally, in the context of linear system $Ax = b$, the null space tells me that the system $Ax = b$ is considered inconsistent because since the column space is a 2D plane in $\mathbb{R}^3$ then most vectors will not lie in Col(A).

The null space tells us the redundancy, ambiguity or the degree of freedom in this solution.

Question 2: Consider the vectors in $\mathbb{R}^4$: $V_1 = \begin{bmatrix}1\\0\\2\\0\end{bmatrix}$, $V_2 = \begin{bmatrix}2\\1\\0\\-1\end{bmatrix}$, $V_3 = \begin{bmatrix}0\\2\\4\\1\end{bmatrix}$, $V_4 = \begin{bmatrix}1\\1\\2\\0\end{bmatrix}$

Let $S = \{V_1, V_2, V_3, V_4\}$

a. Determine whether the set S is linearly independent. Show all workings and state your final conclusion.

<u>solution</u>

$$S = \begin{bmatrix}1 & 2 & 0 & 1\\0 & 1 & 2 & 1\\2 & 0 & 4 & 2\\0 & -1 & 1 & 0\end{bmatrix}$$

for S to be linearly independent, there must be no dependent rows or columns. That is no row or column can be a linear combination of other rows or columns.

We can reduce to row echelon form (REF) and see if we get a completely zero row or if any column is a free variable.

$$S = \begin{bmatrix}1 & 2 & 0 & 1\\0 & 1 & 2 & 1\\2 & 0 & 4 & 2\\0 & -1 & 1 & 0\end{bmatrix}, \text{ Pivot } R_{11} \rightarrow R_3' = R_3 - 2R_1 \rightarrow \begin{bmatrix}1 & 2 & 0 & 1\\0 & 1 & 2 & 1\\0 & -4 & 4 & 0\\0 & -1 & 1 & 0\end{bmatrix}$$

$$\text{Pivot } R_{22}: R_3' = R_3 + 4R_2 \rightarrow \begin{bmatrix}1 & 2 & 0 & 1\\0 & 1 & 2 & 1\\0 & 0 & 12 & 4\\0 & -1 & 1 & 0\end{bmatrix}$$

$$R_4' = R_4 + R_2 \rightarrow \begin{bmatrix}1 & 2 & 0 & 1\\0 & 1 & 2 & 1\\0 & 0 & 12 & 4\\0 & 0 & 3 & 1\end{bmatrix}$$

$$\text{Pivot } R_{33} \rightarrow R_4'' = R_3 - 4R_4 \rightarrow \begin{bmatrix}1 & 2 & 0 & 1\\0 & 1 & 2 & 1\\0 & 0 & 12 & 4\\0 & 0 & 0 & 0\end{bmatrix}$$

rank(S) = 3.

Since the final row reduces to zero, S is not linearly independent. We can also compute the determinat of S

$|S| = 0$ [Since determinant = 0, it is not linearly independent

Conclusion: S is not linearly independent.

b. Basis and dimension (S): Let V = (S). find a basis B for V and state the dimension of V. Justify why your set B is a basis.

solution

From our row echelon form (REF)

$$S = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 12 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The pivot columns are: Column 1, column 2, and column 3

Hence our basis $B = \{V_1, V_2, V_3\} = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 4 \\ 1 \end{bmatrix}$ in the corresponding matrix (V).

This means the dimension of $V = 3$

B is a basis because basis must span V and be linearly independent. The pivot columns span V and are linearly independent which makes them basis of V.

c. Using the basis B, find the coordinates of the vector $z = \begin{bmatrix} 3 \\ 2 \\ 6 \\ -1 \end{bmatrix}$ with respect to B.

solution

$$\text{Basis } (B) = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 2 & 0 & 4 \\ 0 & -1 & 1 \end{bmatrix} \text{ and } Z = \begin{bmatrix} 3 \\ 2 \\ 6 \\ -1 \end{bmatrix}$$

We want to find the coordinate of z with respect to basis B. which means we want to find scalars c_1, c_2, c_3, such that

$$c_1V_1 + c_2V_2 + c_3V_3 = Z$$

The system becomes:

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 2 & 0 & 4 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 6 \\ -1 \end{bmatrix}$$

Lets solve the augmented matrix $[B \mid z]$

$$\left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 0 & 1 & 2 & 2 \\ 2 & 0 & 4 & 6 \\ 0 & -1 & 1 & -1 \end{array}\right]$$

Pivot $R_{11} \longrightarrow R_3 = R_3 - 2R_1 \longrightarrow \left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 0 & 1 & 2 & 2 \\ 0 & -4 & 4 & 0 \\ 0 & -1 & 1 & -1 \end{array}\right]$

Pivot $R_{22} \longrightarrow R_3' = R_3 - 4R_4 \longrightarrow \left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 4 \\ 0 & -1 & 1 & -1 \end{array}\right]$

$R_4' = R_4 + R_2 \longrightarrow \left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 3 & -1 \end{array}\right]$

There is an inconsistency because the third row says $0 + 0 + 0 = 4$. This means the system has no solution

This means there is no coordinate for z in the span(s) and z does not have coordinates with respect to Basis B.

Question 3: Let $T: \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation defined by
$T(x, y, z) = (2x+y-z,\ x-y+3z,\ 3x+2z)$

a. find the standard matrix representation $[T]$ of the linear transformation T.

solution

To form the standard matrix T, we do the dot product on the basis vectors e_1, e_2, e_3.

$e_1 = (1,0,0)$, $e_2 = (0,1,0)$, $e_3 = (0,0,1)$

$T = T(x,y,z) \cdot e$ and $T(x,y,z) = \begin{bmatrix} 2 & 1 & -1 \\ 1 & -1 & 3 \\ 3 & 0 & 2 \end{bmatrix}$

$T(1,0,0) = 2\times 1 + 1\times 0 + -1\times 0 = 2$

And this is done for all cases:

$$\begin{bmatrix} 2 & 1 & -1 \\ 1 & -1 & 3 \\ 3 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}$$

$R_{11} = 2\times 1 + 1\times 0 + -1\times 0 = 2$

$R_{12} = 2\times 0 + 1\times 1 + -1\times 0 = 1$

$R_{13} = 2\times 0 + 1\times 0 + -1\times 1 = -1$

$R_{21} = 1\times 1 + -1\times 0 + 3\times 0 = 1$ $R_{22} = 1\times 0 + -1\times 1 + 3\times 0 = -1$

$R_{23} = 1\times 0 + -1\times 0 + 3\times 1 = 3$, $R_{31} = 3\times 1 + 0\times 0 + 2\times 0 = 3$

$R_{32} = 3\times 0 + 0\times 1 + 2\times 0 = 0$, $R_{33} = 3\times 0 + 0\times 0 + 2\times 1 = 2$

$$T = \begin{bmatrix} 2 & 1 & -1 \\ 1 & -1 & 3 \\ 3 & 0 & 2 \end{bmatrix}$$ This is the standard matrix T.

3b. Determine whether T is one to one (Insective)

solution

For T to be insective, the null space of T must be zero

We can do this by reducing T to row echelon form (REF)

$$T = \begin{bmatrix} 2 & 1 & -1 \\ 1 & -1 & 3 \\ 3 & 0 & 2 \end{bmatrix}$$

Pivot $\xrightarrow{R_1} R_2$; $R_2 = R_1 - 2R_2 \longrightarrow \begin{bmatrix} 2 & 1 & -1 \\ 0 & 3 & -7 \\ 3 & 0 & 2 \end{bmatrix}$

$R_3 = 3R_1 - 2R_3 \longrightarrow \begin{bmatrix} 2 & 1 & -1 \\ 0 & 3 & -7 \\ 0 & 3 & -7 \end{bmatrix}$

Pivot R_{22} ; $R_3'' = R_3 - R_4 \longrightarrow \begin{bmatrix} 2 & 1 & -1 \\ 0 & 3 & -7 \\ 0 & 0 & 0 \end{bmatrix}$

Since the last row reduces to zero, we have a free variable and the null space $= 1$ which is not zero. Hence we can say T is not injective.

C. Determine whether T is surjective (onto);

For T to be onto or surjective, the rank of T must be equal to the number of rows or the dimensions of its co-domain $\mathbb{R}^3$.

From our REF; $\begin{bmatrix} 2 & 1 & -1 \\ 0 & 3 & -7 \\ 0 & 0 & 0 \end{bmatrix}$

The rank equals to 2 which is less than number of rows (3), hence it is not surjective (onto).

3d. Find $T(v)$ where $v = (1, -2, 3)$ and determine if the vector $W = (5, 10, 7)$ is in the range of T.

<u>Solution</u>

$$T = \begin{bmatrix} 2 & 1 & -1 \\ 1 & -1 & 3 \\ 3 & 0 & 2 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} 1 & -2 & 3 \end{bmatrix}^T = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$$

$$T(v) = \begin{bmatrix} 2 & 1 & -1 \\ 1 & -1 & 3 \\ 3 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} = \begin{bmatrix} R_{11} \\ R_{21} \\ R_{31} \end{bmatrix}$$

$R_{11} = 2 \times 1 + 1 \times -2 + -1 \times 3 = 2 - 2 - 3 = -3$

$R_{21} = 1 \times 1 - 1 \times -2 + 3 \times 3 = 1 + 2 + 9 = 12$

$R_{31} = 3 \times 1 + 0 \times -2 + 2 \times 3 = 3 + 0 + 6 = 9$

Hence; $T(v) = (-3, 12, 9)$

Next we check if $W = (5, 10, 7)$ is in the range of T:
To do this we find $[T|W]$ and reduce to REF.

$$T = \begin{bmatrix} 2 & 1 & -1 \\ 1 & -1 & 3 \\ 3 & 0 & 2 \end{bmatrix}, \quad W = \begin{bmatrix} 5 \\ 10 \\ 7 \end{bmatrix}, \quad [T|W] = \left[\begin{array}{ccc|c} 2 & 1 & -1 & 5 \\ 1 & -1 & 3 & 10 \\ 3 & 0 & 2 & 17 \end{array}\right]$$

Pivot R_{11}: $R_2 = R_1 - 2R_2 \longrightarrow \left[\begin{array}{ccc|c} 2 & 1 & -1 & 5 \\ 0 & 3 & -7 & -15 \\ 3 & 0 & 2 & 17 \end{array}\right]$

$R_3 = 3R_1 - 2R_3 \longrightarrow \left[\begin{array}{ccc|c} 2 & 1 & -1 & 5 \\ 0 & 3 & -7 & -15 \\ 0 & 3 & -7 & -19 \end{array}\right]$

$R_3' = R_3 - R_4 \longrightarrow \left[\begin{array}{ccc|c} 2 & 1 & -1 & 5 \\ 0 & 3 & -7 & -15 \\ 0 & 0 & 0 & 4 \end{array}\right]$

The last is inconsistent since $0 + 0 + 0 \neq 4$. Hence we can say W is not in the range of T.

Question 4: consider a linear transformation $\varphi\ \mathbb{R}^3 \to \mathbb{R}^2$ defined by $\varphi(x_1, x_2, x_3) = (2x_1 - x_2 + x_3,\ x_1 + 3x_2 - 2x_3)$

a. Find the matrix representation A_φ of this transformation with respect to the standard bases of $\mathbb{R}^3$ and $\mathbb{R}^2$.

SOLUTION

$$\varphi(x_1, x_2, x_3) = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 3 & -2 \end{bmatrix}$$

standard basis $= (e_1, e_2, e_3)$ and $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

$$A_\varphi = \varphi(x_1, x_2. x_3) \cdot e$$

$$A_\varphi = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 3 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \end{bmatrix}$$

$R_{11} = 2 \times 1 + -1 \times 0 + 1 \times 0 = 2$, $R_{12} = 2 \times 0 + -1 \times 1 + 1 \times 0 = -1$

$R_{13} = 2(0) - 1(0) + 1(1) = 1$, $R_{21} = 1(1) + 3(0) - 2(0) = 1$

$R_{22} = 1(0) + 3(1) - 2(0) = 3$, $R_{23} = 1(0) + 3(0) - 2(1) = -2$

$$A_\varphi = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 3 & -2 \end{bmatrix}$$

b. compute $\varphi(1, -1, 2)^T$ using both the transformation and matrix multiplication.

solution

Using the transformation definition:

$$A_\varphi = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 3 & -2 \end{bmatrix}, \text{ and } \varphi(x_1, x_2, x_3) = \begin{pmatrix} 2x_1 - x_2 + x_3, \text{ and} \\ x_1 + 3x_2 - 2x_3 \end{pmatrix}$$

$$\varphi(1, -1, 2) = (2(1) - (-1) + 2,\ 1 + 3(-1) - 2(2)) = (5, -6)$$

Using the definition $2x_1 - x_2 + x_3 = 2(1) - (-1) + 2 = 5$

$x_1 + 3x_2 + -2x_3 = 1 + 3(-1) - 2(2) = -6$

Using matrix multiplication:

$$A_\varphi = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 3 & -2 \end{bmatrix} \text{ and } \varphi(1, -1, 2)$$

$$= \begin{bmatrix} 2 & -1 & 1 \\ 1 & 3 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} R_{11} \\ R_{21} \end{bmatrix}$$

$R_{11} = 2(1) - 1(-1) + 1(2) = 2 + 1 + 2 = 5$

$R_{21} = 1(1) + 3(-1) - 2(2) = 1 - 3 - 4 = -6$

$$\begin{bmatrix} R_{11} \\ R_{21} \end{bmatrix} = \begin{bmatrix} 5 \\ -6 \end{bmatrix}$$

In both cases, we end up with the same outputs.

4c. In the context of machine learning, if this transformation represents the first layer of a neural network mapping 3 input features to 2 hidden neurons (before activation) what ~~does~~ would the weight matrix be? Explain the connection between linear transformations and neural network layers

<u>Solution</u>

If this transformation represents the first layer of a neural network mapping 3 inputs to 2 hidden neurons the the weight matrix would be given as:

Weight $= A_\varphi$ and the output before activation would be Wx. Weight matrix $(W) = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 3 & -2 \end{bmatrix}$

The connection between linear transformations and neural networks layer is that a linear transformation is the matrix multiplication in neural networks. Hence each row of W contains weight connecting all inputs to one neuron.[1]

Mathematically, we can say:

$$\text{Linear transformation} = \frac{\text{neural layer without bias}}{\text{Activation}}$$

Question 5: Let $B = \{b_1, b_2\}$ be a basis of $\mathbb{R}^2$ where

$$b_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \text{ and } b_2 = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$$

Let $C = \{c_1, c_2, c_3\}$ be a basis of $\mathbb{R}^3$ where

$$c_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad c_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad c_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Consider the linear transformation $T: \mathbb{R}^2$ defined on the basis B by

$$T(b_1) = \begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix}, \quad T(b_2) = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$$

a. Find the transformation matrix A_T of the linear transformation T with respect to bases B and C.

<u>solution</u>

We want to find the transformation matrix A_T such that if $[x]_B$ is the coordinate vector of x in basis B then $A_T[x]_B = [T(x)_C]$.

Given $T(b_1)$ and $T(b_2)$ in the standard basis coordinates in $\mathbb{R}^3$, we need to express them in C basis. Hence we need to express $T(b_1)$ and $T(b_2)$ with respect to C.

$$T(b_1) = [4, 2, 6]^T \text{ and } T(b_2) = [2, 5, 1]^T$$

Let $T(b_1) = c_1 \cdot c_1 + c_2 \cdot c_2 + c_3 \cdot c_3$ and c_1, c_2, and c_3 is given

Then $c_1 \cdot c_1 = c_1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $c_2 \cdot c_2 = c_2 \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$ and $c_3 \cdot c_3 = c_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

$$\Rightarrow c_1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix}$$

This gives the system below

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix}$$

Next we solve the augmented matrix to reduced row echelon:

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 4 \\ 1 & 1 & 0 & 2 \\ 0 & 2 & 1 & 6 \end{array}\right]$$

Pivot R_{11}: $R_2 = R_2 - R_1 \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & 4 \\ 0 & 1 & -1 & -2 \\ 0 & 2 & 1 & 6 \end{array}\right]$

Pivot R_{22}: $R_3 = R_3 - 2R_2 \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & 4 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 3 & 10 \end{array}\right]$

$R_3 \longrightarrow \frac{R_3}{3} \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & 4 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 1 & 10/3 \end{array}\right]$

$R_1 \longrightarrow R_1 - R_3 \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 4-10/3 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 1 & 10/3 \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2/3 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 1 & 10/3 \end{array}\right]$

$R_2 = R_2 + R_3 \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2/3 \\ 0 & 1 & 0 & -2+10/3 \\ 0 & 0 & 1 & 10/3 \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2/3 \\ 0 & 1 & 0 & 4/3 \\ 0 & 0 & 1 & 10/3 \end{array}\right]$

Hence: $[T(b_1)]_C = \begin{bmatrix} 2/3 \\ 4/3 \\ 10/3 \end{bmatrix}$

Now we repeat the process for $T(b_2) = [2, 5, 1]^T$

$C_1 \cdot C_1 + C_2 \cdot C_2 + C_3 \cdot C_3 = [2, 5, 1]^T$

$\Rightarrow C_1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + C_2 \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} + C_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$

$\Rightarrow \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$

The augmented matrix becomes:

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 1 & 1 & 0 & 5 \\ 0 & 2 & 1 & 1 \end{array}\right]$$

Now we reduce the augment to RREF:

Pivot R_{11}: $R_2 = R_2 - R_1 \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & -1 & 3 \\ 0 & 2 & 1 & 1 \end{array}\right]$

Pivot R_{22}: $R_3 = R_3 - 2R_2 \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 3 & -5 \end{array}\right]$

Let $R_3 \longrightarrow \frac{1}{3} R_3 \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 1 & -5/3 \end{array}\right]$

$R_1 = R_1 + -R_3 \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2+5/3 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 1 & -5/3 \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 11/3 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 1 & -5/3 \end{array}\right]$

$R_2 = R_2 + R_3 \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 11/3 \\ 0 & 1 & 0 & 3-\frac{5}{3} \\ 0 & 0 & 1 & -5/3 \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 11/3 \\ 0 & 1 & 0 & 4/3 \\ 0 & 0 & 1 & -5/3 \end{array}\right]$

Hence: $\left[T(b_2)\right]_C = \begin{bmatrix} \frac{11}{3} \\ \frac{4}{3} \\ \frac{-5}{3} \end{bmatrix}$

Therefore the matrix A_T with respect to the bases B and C is:

$$A_T = \begin{bmatrix} \frac{2}{3} & \frac{11}{3} \\ \frac{4}{3} & \frac{4}{3} \\ \frac{10}{3} & \frac{-5}{3} \end{bmatrix}$$

5b. Consider new bases:

$\tilde{B} = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}$ of $\mathbb{R}^2$, $\tilde{C} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \right\}$ of $\mathbb{R}^3$. Under this basis, find the transformation matrix $\tilde{A}_T$

Solution

Recall that $T(b_1) = \begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix}$, $T(b_2) \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$

We are looking for $\tilde{A}_T$ with respect to $\tilde{B}$ and $\tilde{C}$

First we need to find the standard matrix of T, such that $T(x) = [T]_{std}\, x$ for every standard coordinates.

Recall basis B in $\mathbb{R}^2 = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} = P_B$

basis C in $\mathbb{R}^3 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} = P_C$

$$A_T = \begin{bmatrix} 2/3 & 11/3 \\ 4/3 & 4/3 \\ 10/3 & -5/3 \end{bmatrix}$$

Recall that $[T(x)]_C = A_T\, [X]_B$ and $[X]_B = P_B^{-1} X_{std}$

Therefore our standard matrix $T(x)_{std} = P_C \cdot [T(x)]_C$

$T(x)_{std} = P_C \cdot A_T \cdot [X]_B$

Then $T(x)_{std} = P_C \cdot A_T \cdot P_B^{-1} X_{std}$

$$P_C = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}, \quad A_T = \begin{bmatrix} 2/3 & 11/3 \\ 4/3 & 4/3 \\ 10/3 & -5/3 \end{bmatrix}, \quad P_B = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}$$

Next we compute P_B^{-1}: $\frac{1}{|P_B|} \cdot [\text{Adj } P_B]$

$$P_B = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}, \quad \text{adj } P_B = \begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix}$$

$$|P_B| = 6-(-1) = 7 \quad \text{Then } P_B^{-1} = \frac{1}{7}\begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 3/7 & 1/7 \\ -1/7 & 2/7 \end{bmatrix}$$

$$T_{std} = P_C \cdot A_T \cdot P_B^{-1} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2/3 & 11/3 \\ 4/3 & 4/3 \\ 10/3 & -5/3 \end{bmatrix} \begin{bmatrix} 3/7 & 1/7 \\ -1/7 & 2/7 \end{bmatrix}$$

Lets compute $P_C \cdot A_T \rightarrow \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \cdot \frac{1}{3}\begin{bmatrix} 2 & 11 \\ 4 & 4 \\ 10 & -5 \end{bmatrix}$

$$\text{first column} = \frac{2}{3}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \frac{4}{3}\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} + \frac{10}{3}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{2}{3} + \frac{2}{3}(0) + \frac{10}{3} \\ \frac{2}{3} + \frac{4}{3}(1) + \frac{10}{3}(0) \\ \frac{2}{3}(0) + \frac{4}{3}(2) + \frac{10}{3}(1) \end{bmatrix} = \begin{bmatrix} \frac{12}{3} \\ \frac{6}{3} \\ \frac{18}{3} \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix}$$

$$\text{Second column: } \frac{11}{3}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \frac{4}{3}\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} + -\frac{5}{3}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{11}{3} + \frac{4}{3}(0) - \frac{5}{3}(1) \\ \frac{11}{3}(1) + \frac{4}{3}(1) - \frac{5}{3}(0) \\ \frac{11}{3}(0) + \frac{4}{3}(2) - \frac{5}{3}(1) \end{bmatrix} = \begin{bmatrix} \frac{6}{3} \\ \frac{15}{3} \\ \frac{3}{3} \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$$

$$\text{Hence, } P_C A_T = \begin{bmatrix} 4 & 2 \\ 2 & 5 \\ 6 & 1 \end{bmatrix}$$

$$T_{std} = P_C A_T \cdot P_B^{-1} = \begin{bmatrix} 4 & 2 \\ 2 & 5 \\ 6 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3/7 & 1/7 \\ -1/7 & 2/7 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

$$R_{11} = 4(3/7) + 2(-1/7) = \frac{12}{7} - \frac{2}{7} = \frac{10}{7}$$

$$R_{12} = 4(1/7) + 2(2/7) = \frac{4}{7} + \frac{4}{7} = \frac{8}{7}$$

$$R_{21} = 2(3/7) + 5(-1/7) = \frac{6}{7} - \frac{5}{7} = \frac{1}{7}$$

$$R_{22} = 2\left(\frac{1}{7}\right) + 5\left(\frac{2}{7}\right) = \frac{2}{7} + \frac{10}{7} = \frac{12}{7}$$

$$R_{31} = 6\left(\frac{3}{7}\right) + 1\left(\frac{-1}{7}\right) = \frac{18}{7} - \frac{1}{7} = \frac{17}{7}$$

$$R_{32} = 6\left(\frac{1}{7}\right) + 1\left(\frac{2}{7}\right) = \frac{6}{7} + \frac{2}{7} = \frac{8}{7}$$

$$\begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} = \frac{1}{7} \begin{bmatrix} 10 & 8 \\ 1 & 12 \\ 17 & 8 \end{bmatrix} = T_{std}.$$

Changing to new basis $\tilde{B}$ and $\tilde{C}$

$$\tilde{B} = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} = P_{\tilde{B}} \quad \text{and} \quad C = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix} = P_{\tilde{C}}$$

Again we want $\tilde{A}_T$ such that $[T(x)]_{\tilde{C}} = \tilde{A}_T [X]_{\tilde{B}}$ and $[X]_{\tilde{B}} = P_{\tilde{B}}^{-1}$ Then $[T(x)]_{\tilde{C}} = \tilde{A}_T [X]_{\tilde{B}}$ becomes

$T_{std} = \tilde{A}_T \cdot P_{\tilde{B}}^{-1} \cdot P_{\tilde{C}}$ since $T_{std} = [T(x)]_C \cdot P_{\tilde{C}}$

If we make $\tilde{A}_T$ the subject:

$$\tilde{A}_T = P_{\tilde{C}}^{-1} T_{std} P_{\tilde{B}}^{+1} = P_{\tilde{C}}^{-1} T_{std} P_{\tilde{B}}$$

Given: $T_{std} = \frac{1}{7}\begin{bmatrix} 10 & 8 \\ 1 & 12 \\ 17 & 8 \end{bmatrix}$ and $P_{\tilde{B}} = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}$

We need to compute $P_{\tilde{C}}^{-1}$.

$P_{\tilde{C}} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix}$ Then we can compute $P_{\tilde{C}}$ by augmenting the identity matrix I.

$$P_{\tilde{C}}^{-1} = \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 & 0 & 1 \end{array}\right]$$

Pivot $R_{11} \rightarrow R_3 = R_3 - R_1 \rightarrow \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 1 \end{array}\right]$

Pivot R_{22}: $R_3' = R_2 + R_3 \rightarrow \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 3 & -1 & 1 & 1 \end{array}\right]$

$R_3 = \frac{1}{3}R_3 \rightarrow \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1/3 & 1/3 & 1/3 \end{array}\right]$

$R_2 = R_2 - R_3 \rightarrow \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1/3 & 2/3 & -1/3 \\ 0 & 0 & 1 & -1/3 & 1/3 & 1/3 \end{array}\right]$

$R_1 = R_1 - R_2 \rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1-\frac{1}{3} & 0-\frac{2}{3} & 0+\frac{1}{3} \\ 0 & 1 & 0 & 1/3 & 2/3 & -1/3 \\ 0 & 0 & 1 & -1/3 & 1/3 & 1/3 \end{array}\right]$

$$P_{\tilde{C}}^{-1} = \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \\ 1/3 & 2/3 & -1/3 \\ -1/3 & 1/3 & 1/3 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 & -2 & 1 \\ 1 & 2 & -1 \\ -1 & 1 & 1 \end{bmatrix}$$

$$\tilde{A}_T = P_{\tilde{C}}^{-1} \cdot T_{std}\, P_{\tilde{B}} = \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \cdot \begin{bmatrix} \frac{10}{7} & \frac{8}{7} \\ \frac{1}{7} & \frac{12}{7} \\ \frac{17}{7} & \frac{8}{7} \end{bmatrix} \cdot \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}$$

Lets compute $T_{std} \cdot P_{\tilde{B}} = \frac{1}{7} \begin{bmatrix} 10 & 8 \\ 1 & 12 \\ 17 & 8 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \\ R_{31} & R_{32} \end{bmatrix}$

$$R_{11} = \frac{10}{7}(1) + \frac{8}{7}(2) = \frac{10}{7} + \frac{16}{7} = \frac{26}{7}$$

$$R_{12} = \frac{10}{7}(3) + \frac{8}{7}(1) = \frac{30}{7} + \frac{8}{7} = \frac{38}{7}$$

$$R_{21} = \frac{1}{7}(1) + \frac{12}{7}(2) = \frac{1}{7} + \frac{24}{7} = \frac{25}{7}$$

$$R_{22} = \frac{1}{7}(3) + \frac{12}{7}(1) = \frac{3}{7} + \frac{12}{7} = \frac{15}{7}$$

$$R_{31} = \frac{17}{7}(1) + \frac{8}{7}(2) = \frac{17}{7} + \frac{16}{7} = \frac{33}{7}$$

$$R_{32} = \frac{17}{7}(3) + \frac{8}{7}(1) = \frac{51}{7} + \frac{8}{7} = \frac{59}{7}$$

$$T_{std} \cdot P_{\tilde{B}} = \frac{1}{7} \begin{bmatrix} 26 & 38 \\ 25 & 15 \\ 33 & 59 \end{bmatrix}$$

Next we compute $P_{\tilde{C}}^{-1} \cdot T_{std}\, P_{\tilde{B}}$

$$\tilde{A}_T = \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \cdot \begin{bmatrix} 26/7 & 38/7 \\ 25/7 & 15/7 \\ 33/7 & 59/7 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \\ R_{31} & R_{32} \end{bmatrix}$$

$$R_{11} = \frac{2}{3}\left(\frac{26}{7}\right) - \frac{2}{3}\left(\frac{25}{7}\right) + \frac{1}{3}\left(\frac{33}{7}\right) = \frac{52}{21} - \frac{50}{21} + \frac{33}{21} = \frac{35}{21}$$

$$R_{12} = \frac{2}{3}\left(\frac{38}{7}\right) - \frac{2}{3}\left(\frac{15}{7}\right) + \frac{1}{3}\left(\frac{59}{7}\right) = \frac{76}{21} - \frac{30}{21} + \frac{59}{21} = \frac{105}{21}$$

$$R_{21} = \frac{1}{3}\left(\frac{26}{7}\right) + \frac{2}{3}\left(\frac{25}{7}\right) - \frac{1}{3}\left(\frac{33}{7}\right) = \frac{26}{21} + \frac{50}{21} - \frac{33}{21} = \frac{43}{21}$$

$$R_{22} = \frac{1}{3}\left(\frac{38}{7}\right) + \frac{2}{3}\left(\frac{15}{7}\right) - \frac{1}{3}\left(\frac{59}{7}\right) = \frac{38}{21} + \frac{30}{21} - \frac{59}{7} = \frac{9}{21}$$

$$R_{31} = -\frac{1}{3}\left(\frac{26}{7}\right) + \frac{1}{3}\left(\frac{25}{7}\right) + \frac{1}{3}\left(\frac{33}{7}\right) = \frac{33}{21} + \frac{25}{21} - \frac{26}{21} = \frac{32}{21}$$

$$R_{32} = -\frac{1}{3}\left(\frac{38}{7}\right) + \frac{1}{3}\left(\frac{15}{7}\right) + \frac{1}{3}\left(\frac{59}{7}\right) = \frac{-38}{21} + \frac{15}{21} + \frac{59}{21} = \frac{36}{21}$$

$$\tilde{A}_T = \frac{1}{21}\begin{bmatrix} 35 & 105 \\ 43 & 9 \\ 32 & 36 \end{bmatrix} = \begin{bmatrix} \frac{35}{21} & \frac{105}{21} \\ \frac{43}{21} & \frac{9}{21} \\ \frac{32}{21} & \frac{36}{21} \end{bmatrix} = \begin{bmatrix} \frac{5}{3} & 5 \\ \frac{43}{21} & 3/7 \\ 32/21 & 12/7 \end{bmatrix}$$

5c. Find the image and kernel of T

Recall: $T_{std} = \frac{1}{7}\begin{bmatrix} 10 & 8 \\ 1 & 12 \\ 17 & 8 \end{bmatrix}$ We can find the kernel by solving $T_{std} = 0$

$$\frac{1}{7}\begin{bmatrix} 10 & 8 \\ 1 & 12 \\ 17 & 8 \end{bmatrix} = 0 \quad \text{multiply both side by } 7$$

$$= \begin{bmatrix} 10 & 8 \\ 1 & 12 \\ 17 & 8 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$\Rightarrow 10x_1 + 8x_2 = 0$ --- (i)

$x_1 + 12x_2 = 0$ --- (ii)

from (ii) $x_1 = 0 - 12x_2 = -12x_2$ --- (iii)

substitute $x_1 = -12x_2$ into (1)

$10x_1 + 8x_2 = 0$

$10(-12x_2) + 8x_2 = 0 \Rightarrow -120x_2 + 8x_2 = 0$

divide through by 4: $-30x_2 + 2x_2 = 0$

Hence $-28x_2 = 0$ then $x_2 = 0/-28 = 0$

$x_1 = -12x_2 = -12(0) = 0$

therefore kernel $= 0$

The image of the columns $(\text{Im}(T)) = \text{span}\{T(b_1), T(b_2)\}$ in $\mathbb{R}^3$ coordinates, Therefore:

$$\text{Im}(T) = \text{span}\left\{\begin{bmatrix}4\\2\\6\end{bmatrix}\begin{bmatrix}2\\5\\1\end{bmatrix}\right\}$$

$$\text{kernel}(T) = \{0\}$$

5d. Explain why the property of a linear transformation being one to one is important in machine learning, Particularly for neural network layers and feature extraction. Discuss what information is preserved or lost when a transformation is not one-to-one.

<u>Solution</u>

The property of a linear transformation being one to one in machine learning is important because if a neural network is not one to one, it loses information as different ~~output~~ ~~produces~~ inputs produces the same outputs. For feature extraction, this non insective mapping would cause a collapse of distinct data points to the same features thereby making them indinstinguishable later. This is bad if important distinctions are lost. [2] [3] [4]

What information is preserved or lost ?

<u>solution</u>

* What is lost !

1. Uniqueness of representation is lost since multiple inputs maps to the same outputs
2. Invertibility is lost since you cannot recover input from output uniquely, which means the layer discards some outputs/input information
3. Distinct data sample can become identical

* What is preserved !

1. Coset structure is preserved since all inputs of the form $x+v$ where $v \in \ker(T)$ maps to the same outputs
2. Linear dependencies is preserved since inputs are related by something in the kernel and outputs are the same.
3. Noise reduction is preserved by collapsing unimportant variations. [4]

Question 6: Consider a linear layer (no bias) acting on a 2D input by the matrix

$$W = \begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix}$$

a. show that $T: \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ defined by $T(x) = Wx$ is linear i.e show that $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$

<u>Solution</u>

Recall that for a transformation to be linear, it must satisfy two Conditions:

All vectors $x, y \in \mathbb{R}^2$, and

Any scalar $\alpha, \beta \in \mathbb{R}$.

1. Additive rule: $T(x+y) = T(x) + T(y)$
2. Homogeneity: $T(\alpha x) = \alpha T(x)$

To show that $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$ Given

$T(x) = Wx$ and $W = \begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix}$ hence $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$$T(x) = Wx = \begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 - x_2 \\ x_1 + 3x_2 \end{pmatrix}$$

Hence: $T(x_1, x_2) = (2x_1 - x_2, x_1 + 3x_2)$

This also means $T(y) = \in \mathbb{R}$ and $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ as $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

Let α and β be scalars,

$$\alpha x + \beta y = \begin{pmatrix} \alpha x_1 + \beta y_1 \\ \alpha x_2 + \beta y_2 \end{pmatrix}$$

If we apply T to $\alpha x + \beta y$ where $\alpha x_1 + \beta y_1 = x_1$, $\alpha x_2 + \beta y_2 = x_2$

$$T(\alpha x + \beta y) = \begin{pmatrix} 2(\alpha x_1 + \beta y_1) - (\alpha x_2 + \beta y_2) \\ 1(\alpha x_1 + \beta y_1) + 3(\alpha x_2 + \beta y_1 \end{pmatrix}$$

Now we have two coordinates:

$$2(\alpha x_1 + \beta y_1) - (\alpha x_2 + \beta y_2) = 2\alpha x_1 + 2\beta y_1 - \alpha x_2 - \beta y_2$$
$$= 2\alpha x_1 - \alpha x_2 + 2\beta y_1 - \beta y_2$$
$$= \alpha(2x_1 - x_2) + \beta(2y_1 - y_2)$$

Similarly: $1(\alpha x_1 + \beta y_1) + 3(\alpha x_2 + \beta y_2)$
$$= \alpha x_1 + 3\alpha x_2 + \beta y_1 + \beta y_2$$
$$= \alpha(x_1 + 3x_2) + \beta(y_1 + y_2)$$

Hence:
$$T(\alpha x + \beta y) = \begin{pmatrix} \alpha(2x_1 - x_2) + \beta(2y_1 - y_2) \\ \alpha(x_1 + 3x_2) + \beta(y_1 + y_2) \end{pmatrix}$$

Likewise:
$$\alpha T(x) + \beta T(y)$$

Recall $T(x) = \begin{pmatrix} 2x_1 - x_2 \\ x_1 + 3x_2 \end{pmatrix}$ also $T(y) = \begin{pmatrix} 2y_1 - y_2 \\ y_1 + 3y_2 \end{pmatrix}$

$$\alpha T(x) = \begin{pmatrix} \alpha(2x_1 - x_2) \\ \alpha(x_1 + 3x_2) \end{pmatrix},$$

$$\beta T(y) = \begin{pmatrix} \beta(2y_1 - y_2) \\ \beta(y_1 + 3y_2) \end{pmatrix}$$

Adding them we get: $\alpha T(x) + \beta T(y) =$
$$\begin{pmatrix} \alpha(2x_1 - x_2) + \beta(2y_1 - y_2) \\ \alpha(x_1 + 3x_2) + \beta(y_1 + 3y_2) \end{pmatrix}$$

Hence we see $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$

Therefore T is linear.

6b. Let $B = \{b_1, b_2\}$ be the basis $b_1 = (1,1)^T$, $b_2 = (1,-1)^T$
Compute the matrix representation $[T]_B$ of T which is with respect to the basis B.

<u>Solution</u>

$$W = \begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix} \text{ and } b_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, b_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$B = \{b_1, b_2\} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\} = P$$

If we write the vector V in B coordinates:

$$[V]_B = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \text{ then } c_1 b_1 + c_2 b_2 = V$$

Recall $[T]_B = P^{-1} W P$ and $P = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

Next we compute P^{-1}: $P^{-1} = \frac{1}{|P|} \operatorname{Adj} P$

$$|P| = -1-1 = -2 \text{ and } \operatorname{Adj} P = \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$P^{-1} = \frac{-1}{2} \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$W = \begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix}$$ so now we can compute $[T]_B$

$$[T]_B = P^{-1} W P = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$WP = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

$R_{11} = 2(1) - 1(1) = 1$

$R_{12} = 2(1) - 1(-1) = 3$

$R_{21} = 1(1) + 3(1) = 4$

$R_{22} = 1(1) + 3(-1) = -2$

$$\Rightarrow \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 4 & -2 \end{bmatrix}$$

$[T]_B = P^{-1}WP$ becomes $\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 4 & -2 \end{bmatrix}$

$$= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 4 & -2 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

$$R_{11} = \frac{1}{2}(1) + \frac{1}{2}(4) = \frac{1}{2} + \frac{2}{1} = \frac{5}{2}$$

$$R_{12} = \frac{1}{2}(3) + \frac{1}{2}(-2) = \frac{3}{2} - \frac{1}{1} = \frac{1}{2}$$

$$R_{21} = \frac{1}{2}(1) - \frac{1}{2}(4) = \frac{1}{2} - \frac{2}{1} = \frac{-3}{2}$$

$$R_{22} = \frac{1}{2}(3) - \frac{1}{2}(-2) = \frac{3}{2} + \frac{1}{1} = \frac{5}{2}$$

$$\begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 5 & 1 \\ -3 & 5 \end{bmatrix}$$

Hence $[T]_B = \begin{bmatrix} 5/2 & 1/2 \\ -3/2 & 5/2 \end{bmatrix}$

6c) Interprete in machine learning terms what changing the basis means for a linear layer and how this relates to feature representation.

<u>Solution</u>

Changing the basis is like describing the same data using a different set of measuring metric. This corresponds to transforming input features into a new coordinate system before applying the weights. This can make the features more meaningful like in PCA thereby making learning more efficient

[5][6]

Question 7:

a. Prove that the transformation $T: \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x,y) = (x+2y, 3x-y)$ is linear by verifying both the required properties.

<u>Solution</u>

The two properties are given as

1. Homogeneity: $T(\alpha x) = \alpha T(x)$

2: Additivity: $T(x+y) = T(x) + T(y)$

First Lets test homogeneity: $T(\alpha x) = \alpha T(x)$

We can rewrite this as $T(\alpha u) = \alpha T(u)$

Let $u_1 = (x_1, y_1)$ and $v = (x_2, y_2)$ and α be a scalar

$$T(\alpha u) = T(\alpha x_1, \alpha y_1) = (\alpha x_1 + 2\alpha y_1, 3\alpha x_1 - \alpha y_1)$$

$$= \alpha(x_1 + 2y_1, 3x_1 - y_1) = \alpha T(u)$$

Hence $T(\alpha u) = \alpha T(u)$

For additivity: $T(u+v) = T(u) + T(v)$

where $u+v = (x_1 + x_2, y_1 + y_2)$

$$T(u+v) = T(x_1 + x_2, y_1 + y_2) \Rightarrow$$

$$= ((x_1 + x_2) + 2(y_1 + y_2), 3(x_1 + x_2) - (y_1 + y_2))$$

$$T(u+v) = ((x_1 + x_2) + 2(y_1 + y_2), 3(x_1 + x_2) - 1(y_1 + y_2))$$

$$T(u) + T(v) \Rightarrow T(u) = (x_1 + 2y_1, 3x_1 - y_1)$$

$$T(v) = (x_2 + 2y_2, 3x_2 - y_2)$$

$$T(u) + T(v) = (x_1 + 2y_1 + x_2 + 2y_2, 3x_1 - y_1 + 3x_2 - y_2)$$

$$= ((x_1 + x_2 + 2y_1 + 2y_2, 3x_1 + 3x_2 - (y_1 + y_2))$$

$$= (x_1 + x_2 + 2(y_1 + y_2), 3(x_1 + x_2) - 1(y_1 + y_2))$$

Since $T(u+v) = T(u) + T(v)$

Then $T(x, y)$ is linear, and T is a linear transformation

7b. Find the kernel (null space and image of the transformation represented by the matrix

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \\ -1 & -2 & 1 \end{bmatrix}$$

What is the dimension of each.

<u>Solution</u>

The kernel $\ker(A)$ is given by $Ax = 0$ where

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

which means $Ax = 0 = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \\ -1 & -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

We can reduce A to Row echelon form and see if We have any free variables or dependent rows/columns.

$$\begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \\ -1 & -2 & 1 \end{bmatrix}$$

Pivot R_{11}: $R_2 = R_2 - 2R_1$

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

Again $R_3 = R_1 + R_3 \longrightarrow \begin{bmatrix} 1 & 2 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

The only pivot column is column 1 so rank $= 1$

From the REF we can compute $Ax = 0$ and since column 2 and 3 are free variables, set $x_2 = s$ and $x_3 = t$

$x_1 + 2x_2 - x_3 = 0$ becomes

$x_1 + 2s - t = 0$ Therefore $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} t-2s \\ s \\ t \end{bmatrix}$

$x_1 = t - 2s$

Hence: $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = s\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, t\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

Which means kernel (null space) $= \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ span

and the dimension $= 2$

Image (Range) can be gotten from the pivot column which corresponds to:

$$\text{Im}(A) = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \text{ and the dimension } = 1$$

7c. Explain why dimensionality reduction like PCA relies fundamentally on linear transformations, and what limitations this imposes on the types of patterns PCA can capture. compare this with non-Linear dimensionality reduction methods.

<u>Solution</u>

Principal component analysis (PCA) is a linear dimensionality reduction technique that transforms data into a new vector coordinate system such that the first Principal components captures the direction of maximum variance and each subsequent component captures the next highest variance, orthogonal to the previous ones. [7] [8]

PCA relies on linear combinations of the original features while using orthogonal projection to project onto new axis called principal components, which means PCA captures only

Linear relationships in the data. Now because PCA is Linear, some of the other limitations includes:

a. It cannot capture non linear patterns such as curved manifolds, spirals, clusters in non linear shapes.

b. If the data is on a non linear manifold, PCA will flatten it incorrectly, losing important structure. [9]

Non Linear Dimensionality Reduction Methods

Non linear methods are used to overcome PCA's limitations, such method includes:

1. t-SNE (t-distributed stochastic Neighbouring Embedding which preserves local structures and ensures similar points stay close. This method is great for visualizing clusters in 2D or 3D but not great for generalization tasks.

2. Isomap which preserves geodesic distances on a manifold and is good for unfolding non linear structures

3. Auto encoders (Neural networks) which is used to learn non linear mappings using deep learning, and can compress data with complex patterns. [9] [10]

Question 8: Feature Engineering and dimensionality Reduction

A data scientist is working with a dataset of house prices using three features: size (s in sq. ft), Number of bedrooms (b) and Age (a) in years. The raw features are

$$S = [1000, 1500, 1200, 2000, 1800]^T$$

$$b = [2, 3, 2, 4, 3]^T, \quad a = [5, 10, 2, 20, 15]^T;$$

The corresponding prices are: $y = [200, 300, 240, 400, 350]^T$ in thousands.

a. Create a design matrix $X \in \mathbb{R}^{5\times 4}$ that includes a bias term (column of ones) and the three features: $X = [1, s, b, a]$

Solution

Given:

$$S = [1000, 1500, 1200, 2000, 1800]^T$$

$$b = [2, 3, 2, 4, 3]^T, \quad a = [5, 10, 2, 20, 15]^T$$

bias term $= [1, 1, 1, 1, 1]^T$ then the matrix X:

$$X = \begin{bmatrix} 1 & 1000 & 2 & 5 \\ 1 & 1500 & 3 & 10 \\ 1 & 1200 & 2 & 2 \\ 1 & 2000 & 4 & 20 \\ 1 & 1800 & 3 & 15 \end{bmatrix}$$

* Determine whether the columns are linearly dependent, if they are find a non trivial combination linearly that gives the zero vector. What is the practical implication of linear dependence for machine learning.

Solution

To determine where the columns of X are linearly dependent, we can reduce X to its row echelon form (REF)

$$X = \begin{bmatrix} 1 & 1000 & 2 & 5 \\ 1 & 1500 & 3 & 10 \\ 1 & 1200 & 2 & 2 \\ 1 & 2000 & 4 & 20 \\ 1 & 1800 & 3 & 15 \end{bmatrix}$$

Pivot R_{11}: $R_2 = R_2 - R_1 \longrightarrow \begin{bmatrix} 1 & 1000 & 2 & 5 \\ 0 & 500 & 1 & 5 \\ 1 & 1200 & 2 & 2 \\ 1 & 2000 & 4 & 20 \\ 1 & 1800 & 3 & 15 \end{bmatrix}$

$R_3 = R_3 - R_1 \longrightarrow \begin{bmatrix} 1 & 1000 & 2 & 5 \\ 0 & 500 & 1 & 5 \\ 0 & 200 & 0 & -3 \\ 1 & 2000 & 4 & 20 \\ 1 & 1800 & 3 & 15 \end{bmatrix}$

$R_4 = R_4 - R_1 \longrightarrow \begin{bmatrix} 1 & 1000 & 2 & 5 \\ 0 & 500 & 1 & 5 \\ 0 & 200 & 0 & -3 \\ 0 & 1000 & 2 & 15 \\ 1 & 1800 & 3 & 15 \end{bmatrix}$

$R_5 = R_5 - R_1 \longrightarrow \begin{bmatrix} 1 & 1000 & 2 & 5 \\ 0 & 500 & 1 & 5 \\ 0 & 200 & 0 & -3 \\ 0 & 1000 & 2 & 15 \\ 0 & 800 & 1 & 10 \end{bmatrix}$

Pivot R_{22} $\rightarrow$ $R_3 = 2R_2 - 5R_3$ $\rightarrow$ $\begin{bmatrix} 1 & 1000 & 2 & 5 \\ 0 & 500 & 1 & 5 \\ 0 & 0 & 2 & 25 \\ 0 & 1000 & 2 & 15 \\ 0 & 800 & 1 & 10 \end{bmatrix}$

$R_4 = R_4 - 2R_2$ $\rightarrow$ $\begin{bmatrix} 1 & 1000 & 2 & 5 \\ 0 & 500 & 1 & 5 \\ 0 & 0 & 2 & 25 \\ 0 & 0 & 0 & 5 \\ 0 & 800 & 1 & 10 \end{bmatrix}$

$R_5 = 5R_5 - 8R_2$ $\rightarrow$ $\begin{bmatrix} 1 & 1000 & 2 & 5 \\ 0 & 500 & 1 & 5 \\ 0 & 0 & 2 & 25 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & -3 & 10 \end{bmatrix}$

Pivot R_{33}: $R_5' = 3R_3 + 2R_5$ $\rightarrow$ $\begin{bmatrix} 1 & 1000 & 2 & 5 \\ 0 & 500 & 1 & 5 \\ 0 & 0 & 2 & 25 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 95 \end{bmatrix}$

Pivot R_{44}: $R_5'' = 19R_4 - R_5$ $\rightarrow$ $\begin{bmatrix} 1 & 1000 & 2 & 5 \\ 0 & 500 & 1 & 5 \\ 0 & 0 & 2 & 25 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Hence Rank = 4

Since Rank = 4 and columns = 4, they are linearly independent.

What is the practical implication of linear dependence for machine learning models?

Solution

We know that a set of vectors is linearly independent if no vector in the set can be written as a linear combination of the others. In the case we just solved, there are four columns and we found them to be linearly independent. This implies that in machine learning, linear independence means there can be no redundancies in features, that is each feature contributes unique information and no one feature is a duplicate or a predictable combination of others, and this helps the model learn distinct patterns from each feature. [9] [11]

However, for linear dependence, the implication would be that the coefficients in linear regression would not be uniquely determined which leads to infinitely many solutions Also, model interpretation would be problematic, yielding high variance, unstable estimates, multicollinearity issues. This can be remedied by removing redundan features or regularizing. [11]

8b. Apply the gram-schmidt process to orthogonalize the features vector (excluding the bias term). Show all steps, and Provide the resulting orthogonal basis.

Solution

Recall that $s = \begin{bmatrix} 1000 \\ 1500 \\ 1200 \\ 2000 \\ 1800 \end{bmatrix}$, $b = \begin{bmatrix} 2 \\ 3 \\ 2 \\ 4 \\ 3 \end{bmatrix}$, $a = \begin{bmatrix} 5 \\ 10 \\ 2 \\ 20 \\ 15 \end{bmatrix}$

We want to orthogonalize the features in order of s, b, a.

Let $V_1 = s$

$V_2 = b$ and $V_1 = U_1$ hence $U_1 = \begin{bmatrix} 1000 \\ 1500 \\ 1200 \\ 2000 \\ 1800 \end{bmatrix}$

$V_3 = a$

$U_2 = V_2 - \text{Proj}_{U_1}(V_2)$ where $V_2 = b = \begin{bmatrix} 2 \\ 3 \\ 2 \\ 4 \\ 3 \end{bmatrix}$

$\text{Proj}_{U_1} V_2 = \frac{U_1 V_2}{U_1 . U_1} . U_1 \Rightarrow U_1 \cdot U_1 = 1000^2 + 1500^2 + 1200^2 + 2000^2 + 1800^2$

$U_1 . U_1 = 11,930,000$

$U_1 \cdot V_2 = 2(1000) + 3(1500) + 2(1200) + 4(2000) + 3(1800) = 22,300$

$\text{Proj}_{U_1} V_2 = \frac{22300}{11930000} . (1000, 1500, 1200, 2000, 1800)$

$= (1.869, 2.804, 2.243, 3.738, 3.364)$

$U_2 = V_2 - \text{Proj}_{U_1} V_2 = (2, 3, 2, 4, 3) - (\text{Proj}_{U_1} V_2)$

$U_2 = \begin{bmatrix} 2 \\ 3 \\ 2 \\ 4 \\ 3 \end{bmatrix} - \begin{bmatrix} 1.869 \\ 2.804 \\ 2.243 \\ 3.738 \\ 3.364 \end{bmatrix} \approx \begin{bmatrix} 0.131 \\ 0.196 \\ -0.243 \\ 0.262 \\ -0.364 \end{bmatrix}$

In the same way we can compute

$U_3 = V_3 - \text{Proj}_{U_1} V_3 - \text{Proj}_{U_2} V_3$

$V_3 = a = \begin{bmatrix} 5 \\ 10 \\ 2 \\ 20 \\ 15 \end{bmatrix}, \quad U_1 = \begin{bmatrix} 1000 \\ 1500 \\ 1200 \\ 2000 \\ 1800 \end{bmatrix}, \quad U_2 = \begin{bmatrix} 0.131 \\ 0.196 \\ -0.243 \\ 0.262 \\ -0.364 \end{bmatrix}$

$$\text{Proj}_{u_1} V_3 = V_3 - \frac{V_3 u_1}{u_1 u_1} \cdot u_1$$

$$V_3 \cdot u_1 = (5, 10, 2, 20, 15) \cdot (1000, 1500, 1200, 2000, 1800)^T$$

$$= 5(1000) + 10(1500) + 2(1200) + 20(2000) + 15(1800)$$

$$V_3 \cdot u_1 = 89400$$

Recall that $u_1 \cdot u_1 = 11{,}930{,}000$

$$\text{Proj}_{u_1} V_3 = \frac{V_3 \cdot u_1}{u_1 \cdot u_1} \cdot u_1 = \frac{89400}{11930000} (1000, 1500, 1200, 2000, 1800)$$

$$= 0.007494 (1000, 1500, 1200, 2000, 1800)$$

$$\text{Proj}_{u_1} V_3 = 7.494, 11.241, 8.992, 14.987, 13.489$$

Repeat the steps for $\text{Proj}_{u_2} V_3$

$$\text{Proj}_{u_2} V_3 = \frac{V_3 \cdot u_2}{u_2 \cdot u_2} \cdot u_2 \quad \text{where } V_3 = \begin{bmatrix} 5 \\ 10 \\ 2 \\ 20 \\ 15 \end{bmatrix}, \; u_2 \approx \begin{bmatrix} 0.131 \\ 0.196 \\ -0.243 \\ 0.262 \\ -0.364 \end{bmatrix}$$

$$V_3 \cdot u_2 = (5, 10, 2, 20, 15)(0.131, 0.196, -0.243, 0.262, -0.364)^T$$

$$= 5(0.131) + 10(0.196) + 2(-0.243) + 20(0.262) + 15(-0.364)$$

$$V_3 \cdot u_2 = 1.909$$

$$u_2 \cdot u_2 = 0.131^2 + 0.196^2 + (-0.243)^2 + 0.262^2 + (-0.364)^2$$

$$= 0.3158$$

$$\text{Proj}_{u_2} V_3 = \frac{u_2 V_3}{u_2 u_2} \cdot u_2 = \frac{1.909}{0.3158} \cdot (0.131, 0.196, -0.243, 0.262, -0.364)$$

$$= 6.045 (0.131, 0.196, -0.243, 0.262, -0.364)$$

$$\text{Proj}_{u_2} V_3 = 0.792, 1.185, -1.469, 1.584, -2.200$$

$$u_3 = v_3 - \text{Proj}_{u_1} v_3 - \text{Proj}_{u_2} v_3$$

$$v_3 = \begin{bmatrix} 5 \\ 10 \\ 2 \\ 20 \\ 15 \end{bmatrix}, \quad \text{Proj}_{u_1} v_3 \approx \begin{bmatrix} 7.494 \\ 11.241 \\ 8.992 \\ 14.987 \\ 13.489 \end{bmatrix}, \quad \text{Proj}_{u_2} v_3 \approx \begin{bmatrix} 0.792 \\ 1.185 \\ -1.469 \\ 1.584 \\ -2.200 \end{bmatrix}$$

$$u_3 = \begin{bmatrix} 5 - 7.494 - 0.792 \\ 10 - 11.241 - 1.185 \\ 2 - 8.992 + 1.469 \\ 20 - 14.987 - 1.584 \\ 15 - 13.489 + 2.200 \end{bmatrix} \approx \begin{bmatrix} -3.286 \\ -2.426 \\ -5.523 \\ 3.429 \\ 3.711 \end{bmatrix}$$

$$u_3 \approx \begin{bmatrix} -3.286 \\ -2.426 \\ -5.523 \\ +3.429 \\ +3.711 \end{bmatrix}$$

Finally, the orthogonal basis u_1, u_2, and u_3 are given:

$$u_1 = \begin{bmatrix} 1000 \\ 1500 \\ 1200 \\ 2000 \\ 1800 \end{bmatrix}, \quad u_2 \approx \begin{bmatrix} 0.131 \\ 0.196 \\ -0.243 \\ 0.262 \\ -0.364 \end{bmatrix}, \quad u_3 \approx \begin{bmatrix} -3.286 \\ -2.426 \\ -5.523 \\ 3.429 \\ 3.711 \end{bmatrix}$$

8c) Using your orthogonal basis, compute the projection of the price vector y onto the column space spanned by the features. Interprete this projection in the context of Linear regression

Solution

$$u_1 = \begin{bmatrix} 1000 \\ 1500 \\ 1200 \\ 2000 \\ 1800 \end{bmatrix} \quad u_2 \approx \begin{bmatrix} 0.131 \\ 0.196 \\ -0.243 \\ 0.262 \\ -0.364 \end{bmatrix}, \quad u_3 \approx \begin{bmatrix} -3.286 \\ -2.426 \\ -5.523 \\ 3.429 \\ 3.711 \end{bmatrix}$$

and $y = \begin{bmatrix} 200 \\ 300 \\ 240 \\ 400 \\ 350 \end{bmatrix}$ In thousands

$$\text{Proj}_X y = \frac{y.u_1}{u_1.u_1}.u_1 + \frac{y.u_2}{u_2.u_2}.u_2 + \frac{y.u_3}{u_3.u_3}.u_3$$

$$y.u_1 = [200, 300, 240, 400, 350][1000, 1500, 1200, 2000, 1800]$$

$$= 200(1000) + 300(1500) + 240(1200) + 400(2000) + 350(1800)$$

$$= 2,368,000$$

$$u_1.u_1 = 1000^2 + 1500^2 + 1200^2 + 2000^2 + 1800^2$$

$$= 11,930,000$$

Let $c_1 = \frac{y.u_1}{u_1.u_1} = \frac{2,368,000}{11,930,000} = 0.1985$

$$C_2 = \frac{y_1 \cdot u_2}{u_2 \cdot u_2}$$

$$y_1 \cdot u_2 = [200, 300, 240, 400, 350]\,[0.131, 0.196, -0.243, 0.262, -0.364]$$

$$= 200(0.131) + 300(0.196) + 240(-0.243) + 400(0.262) + 350(-0.364)$$

$$y_1 \cdot u_2 = 26.40 + 58.8 - 58.32 + 104.8 - 127.4$$

$$y_1 \cdot u_2 \approx 4.28$$

$$u_2 \cdot u_2 = 0.131^2 + 0.196^2 + (-0.243)^2 + 0.262^2 + (-0.364)^2$$

$$\approx 0.31603$$

$$C_2 = \frac{4.28}{0.31603} \approx 13.54$$

Likewise Let $\frac{y \cdot u_3}{u_3 \cdot u_3} = C_3$

$$y \cdot u_3 = [200, 300, 240, 400, 350]\,[-3.286, -2.426, -5.523, 3.429, 3.711]$$

$$y \cdot u_3 = 200(-3.286) + 300(-2.426) + 240(-5.523) + 400(3.429) + 350(3.711)$$

$$y \cdot u_3 \approx -40.07$$

And $u_3 \cdot u_3 = (-3.286)^2 + (-2.426)^2 + (-5.523)^2 + (3.429)^2 + (3.711)^2$

$$= 10.798 + 5.885 + 30.504 + 11.758 + 13.772$$

$$u_3 \cdot u_3 \approx 72.716$$

$$C_3 = \frac{y \cdot u_3}{u_3 \cdot u_3} = \frac{-40.07}{72.716} = -0.55105$$

$$C_3 \approx -0.5511$$

$$\text{Proj}_x y = c_1 u_1 + C_2 u_2 + C_3 u_3$$

$$C_1 u_1 = 0.1985 \cdot (1000, 1500, 1200, 2000, 1800)$$

$C_1U_1 = (198.50, 297.75, 238.20, 397.00, 357.30)^T$

Also :

$$C_2U_2 = 13.54(0.131, 0.196, -0.243, 0.262, -0.364)$$
$$= (1.787, 2.654, -3.290, 3.547, -4.929)^T$$

And :

$$C_3U_3 = -0.5511(-3.286, -2.426, -5.523, 3.429, 3.711)$$
$$C_3U_3 = (1.811, 1.337, 3.044, -1.890, -2.045)^T$$

$$\text{Pros}_X\, y = C_1U_1 + C_2U_2 + C_3U_3$$

$$= \begin{bmatrix} 198.50 \\ 297.75 \\ 238.20 \\ 397.00 \\ 357.30 \end{bmatrix} + \begin{bmatrix} 1.787 \\ 2.654 \\ -3.290 \\ 3.547 \\ -4.929 \end{bmatrix} + \begin{bmatrix} 1.811 \\ 1.337 \\ 3.044 \\ -1.890 \\ -2.045 \end{bmatrix}$$

$$\text{Pros}_X\, y = \begin{bmatrix} 198.50 + 1.787 + 1.811 \\ 297.75 + 2.654 + 1.337 \\ 238.20 - 3.290 + 3.044 \\ 397 + 3.547 - 1.890 \\ 357.30 + -4.929 - 2.045 \end{bmatrix} = \begin{bmatrix} 202.098 \\ 301.741 \\ 237.954 \\ 398.657 \\ 350.326 \end{bmatrix}$$

Hence $\text{Pros}_X\, y \approx \begin{bmatrix} 202.098 \\ 301.741 \\ 237.954 \\ 398.657 \\ 350.326 \end{bmatrix}$ in thousands

Interprete this projection in the context of linear regression

Solution

The $proj_x y$ represents the best linear approximation of the target vector y using the given features. In linear regression, this is the predicted price vector $\hat{y}$ obtained by projecting y onto the column space of the design matrix.

$P = [202.098, 301.741, 237.954, 398.657, 350.326]$ is the vector of predicted house prices in thousands from the linear model using size, bedrooms, and age.

8 D) Compute the cosine similarity between the original feature vectors s and b. What does this tell you about the relationship between house size and number of bedrooms in this dataset?

Solution

Recall that the cosine similarity $= \dfrac{s \cdot b}{\|s\| \cdot \|b\|}$

Where $s = [1000, 1500, 1200, 2000, 1800]^T$

$b = [2, 3, 2, 4, 3]^T$

$$s \cdot b = 1000(2) + 1500(3) + 1200(2) + 2000(4) + 1800(3)$$
$$= 2000 + 4500 + 2400 + 8000 + 5400$$
$$= 22300$$

$$\|s\| = \sqrt{1000^2 + 1500^2 + 1200^2 + 2000^2 + 1800^2} = \sqrt{11930000}$$
$$= 3453.983$$

$$\|b\| = \sqrt{2^2 + 3^2 + 2^2 + 4^2 + 3^2} = \sqrt{42} = 6.4807$$

$$\text{cosine similarity} = \frac{s \cdot b}{\|s\| \, \|b\|} = \frac{22300}{\sqrt{11930000} \cdot \sqrt{42}}$$

$$\approx \frac{22300}{22384.37} \approx 0.9962$$

Cosine similarity = 0.9962 approximately.

Interpretation: since the cosine similarity is really close to 1, this means s and b are almost perfectly correlated and this implies that in this dataset, as the number of bedrooms increases, the house size increases almost Proportionally.

8e) using the given starter code, follow the Instruction and implement your solution using only NumPy.

```python
import numpy as np

# ============================================================================
# Question 8: Feature Engineering and Dimensionality Reduction
# ============================================================================

# Given data
s = np.array([1000, 1500, 1200, 2000, 1800])  # Size in sq. ft.
b = np.array([2, 3, 2, 4, 3])                  # Number of bedrooms
a = np.array([5, 10, 2, 20, 15])               # Age in years
y = np.array([200, 300, 240, 400, 350])        # Prices in thousands

print("=" * 70)
print("GIVEN DATA")
print("=" * 70)
print(f"Size (s): {s}")
print(f"Bedrooms (b): {b}")
print(f"Age (a): {a}")
print(f"Prices (y): {y}")
print()

# ============================================================================
# Part (a): Design Matrix and Linear Dependence
# ============================================================================
print("=" * 70)
print("PART (a): Design Matrix and Linear Dependence")
print("=" * 70)

# TODO: Create the design matrix X with bias term (column of ones) and features
np.set_printoptions(suppress=True, precision=2)

X = np.column_stack([np.ones(5), s, b, a])

print("Design Matrix X:")
print(X.astype(int))
print(f"Shape of X: {X.shape}")
print()

# TODO: Check for linear dependence
rank_X = np.linalg.matrix_rank(X)
num_cols = X.shape[1]

print(f"Rank of X: {rank_X}")
print(f"Number of columns: {num_cols}")
print(f"Are columns linearly independent? {rank_X == num_cols}")
print()

# TODO: Compute determinant of X^T X and its determinant
XTX = X.T @ X
det_XTX = np.linalg.det(XTX)

print(f"Determinant of X^T X: {det_XTX}")
print()

# TODO: Find a non-trivial linear combination (if columns are dependent)
print("Finding linear dependence relationship...")
if rank_X < num_cols:
    U, S, Vt = np.linalg.svd(X)
    null_space_vector = Vt[-1, :]
    print(f"Non-trivial linear combination coefficients: {null_space_vector}")
else:
    print("Columns are linearly independent - no non-trivial linear combination exists")

print()
print("TODO: Write your interpretation of what linear dependence means for ML models")
print("Practical Implication:")
print("""
Linear dependence means that the coefficient of one feature can be expressed as a
combination of others, preventing unique solutions to a linear regression which then
leads to infinitely many solutions. Also, models may become unstable and sensitive to
small changes in data which can lead to overfitting. This is known as multicollinearity
in regression analysis and this can be remedied by removing redundant features or using
regularization techniques, or by applying PCA to reduce dimensionality.
""")
print()

# ============================================================================
# Part (b): Gram-Schmidt Orthogonalization
# ============================================================================
print("=" * 70)
print("PART (b): Gram-Schmidt Orthogonalization")
print("=" * 70)

# Extract feature vectors (excluding bias term)
v1 = s  # Size
v2 = b  # Bedrooms
v3 = a  # Age
```

```python
print("Original feature vectors:")
print(f"v1 (size): {v1}")
print(f"v2 (bedrooms): {v2}")
print(f"v3 (age): {v3}")
print()

# TODO: Apply Gram-Schmidt process
u1 = v1.copy()

proj_u1_v2 = (np.dot(v2, u1) / np.dot(u1, u1)) * u1
u2 = v2 - proj_u1_v2

proj_u1_v3 = (np.dot(v3, u1) / np.dot(u1, u1)) * u1
proj_u2_v3 = (np.dot(v3, u2) / np.dot(u2, u2)) * u2
u3 = v3 - proj_u1_v3 - proj_u2_v3

# TODO: Normalize to get orthonormal basis
q1 = u1 / np.linalg.norm(u1)
q2 = u2 / np.linalg.norm(u2)
q3 = u3 / np.linalg.norm(u3)

print("Orthogonal basis vectors:")
print(f"u1: {u1}")
print(f"u2: {u2}")
print(f"u3: {u3}")
print()

# TODO: Verify orthogonality by checking dot products
print("Verification of orthogonality:")
print(f"u1 · u2: {np.dot(u1, u2):.3f}")
print(f"u1 · u3: {np.dot(u1, u3):.3f}")
print(f"u2 · u3: {np.dot(u2, u3):.3f}")

print()

# ============================================================================
# Part (c): Projection of Price Vector
# ============================================================================
print("=" * 70)
print("PART (c): Projection of Price Vector")
print("=" * 70)

# TODO: Compute projection of y onto the column space of features
y_proj = (np.dot(y, q1) * q1 +
          np.dot(y, q2) * q2 +
          np.dot(y, q3) * q3)

print(f"Projection of y onto column space: {y_proj}")
print()

# TODO: Compute projection matrix and apply to y
X_feat = np.column_stack([s, b, a])
P = X_feat @ np.linalg.inv(X_feat.T @ X_feat) @ X_feat.T
y_proj_matrix = P @ y

print("Interpretation in context of linear regression:")
print("TODO: Explain what this projection represents")
print("""
This projection represents the predicted prices from a linear model
using Size, Bedrooms, and Age as features to predict Prices.
It's the closest point to the prices in the column space of features.
This is essentially the output of a linear regression model fitted to the data.
""")
print()

# ============================================================================
# Part (d): Cosine Similarity
# ============================================================================
print("=" * 70)
print("PART (d): Cosine Similarity")
print("=" * 70)

# TODO: Compute cosine similarity between s and b
dot_product_sb = np.dot(s, b)
norm_s = np.linalg.norm(s)
norm_b = np.linalg.norm(b)
cos_similarity = dot_product_sb / (norm_s * norm_b)

print(f"Cosine similarity between size and bedrooms: {cos_similarity:.4f}")

# TODO: Compute the angle in degrees
angle_rad = np.arccos(np.clip(cos_similarity, -1.0, 1.0))
angle_deg = np.degrees(angle_rad)

print(f"Angle between vectors: {angle_deg:.4f}°")
print()

print("Interpretation:")
print("TODO: Explain what this cosine similarity tells us about the relationship")
print("between house size and number of bedrooms")
```

```python
print(f"""
Cosine similarity of {cos_similarity:.4f} indicates very strong positive correlation
between Size and bedrooms, this means they are almost perfectly aligned, suggesting
multicollinearity. This means as the size of the house increases, the number of bedrooms
also tends to increases proportionally and vice versa.
""")
print()

# ======================================================================
# Additional Analysis (Optional)
# ======================================================================
print("=" * 70)
print("ADDITIONAL ANALYSIS")
print("=" * 70)

print("Correlation matrix of features:")
features = np.column_stack([s, b, a])
correlation_matrix = np.corrcoef(features.T)
print(correlation_matrix)

print()
print("=" * 70)
print("END OF ANALYSIS")
print("=" * 70)
```

```
======================================================================
GIVEN DATA
======================================================================
Size (s): [1000 1500 1200 2000 1800]
Bedrooms (b): [2 3 2 4 3]
Age (a): [ 5 10  2 20 15]
Prices (y): [200 300 240 400 350]

======================================================================
PART (a): Design Matrix and Linear Dependence
======================================================================
Design Matrix X:
[[   1 1000    2    5]
 [   1 1500    3   10]
 [   1 1200    2    2]
 [   1 2000    4   20]
 [   1 1800    3   15]]
Shape of X: (5, 4)

Rank of X: 4
Number of columns: 4
Are columns linearly independent? True

Determinant of X^T X: 14860000.000000093

Finding linear dependence relationship...
Columns are linearly independent - no non-trivial linear combination exists

TODO: Write your interpretation of what linear dependence means for ML models
Practical Implication:

Linear dependence means that the coefficient of one feature can be expressed as a
combination of others, preventing unique solutions to a linear regression which then
leads to infinitely many solutions. Also, models may become unstable and sensitive to
small changes in data which can lead to overfitting. This is known as multicollinearity
in regression analysis and this can be remedied by removing redundant features or using
regularization techniques, or by applying PCA to reduce dimensionality.


======================================================================
PART (b): Gram-Schmidt Orthogonalization
======================================================================
Original feature vectors:
v1 (size): [1000 1500 1200 2000 1800]
v2 (bedrooms): [2 3 2 4 3]
v3 (age): [ 5 10  2 20 15]

Orthogonal basis vectors:
u1: [1000 1500 1200 2000 1800]
u2: [ 0.13  0.2  -0.24  0.26 -0.36]
u3: [-3.28 -2.41 -5.54  3.45  3.69]

Verification of orthogonality:
u1 · u2: 0.000
u1 · u3: 0.000
u2 · u3: -0.000

======================================================================
PART (c): Projection of Price Vector
======================================================================
Projection of y onto column space: [201.66 301.22 238.2  398.25 351.2 ]

Interpretation in context of linear regression:
TODO: Explain what this projection represents

This projection represents the predicted prices from a linear model
using Size, Bedrooms, and Age as features to predict Prices.
It's the closest point to the prices in the column space of features.
This is essentially the output of a linear regression model fitted to the data.


======================================================================
PART (d): Cosine Similarity
======================================================================
Cosine similarity between size and bedrooms: 0.9962
Angle between vectors: 4.9762°

Interpretation:
TODO: Explain what this cosine similarity tells us about the relationship
between house size and number of bedrooms
"
Cosine similarity of 0.9962 indicates very strong positive correlation
between Size and bedrooms, this means they are almost perfectly aligned, suggesting
multicollinearity. This means as the size of the house increases, the number of bedrooms
also tends to increases proportionally and vice versa.


======================================================================
ADDITIONAL ANALYSIS
======================================================================
```

```
Correlation matrix of features:
[[1.   0.94 0.95]
 [0.94 1.   0.96]
 [0.95 0.96 1.  ]]

======================================================================
END OF ANALYSIS
======================================================================
```

REFERENCES

[1] Y. El Khalfaoui, C. Elmejgari, B. Alibouch, Y. Nadir, and A. F. El Ouafdi, "Adaptive Spline Transformations vs Traditional Linear Layers: A Comparative Analysis of CNN Architectures for WCE Images Classification," Oct. 2025. doi: 10.1007/978-3-032-01536-5_90.

[2] F. Peng, J. Gao, X. Lu, K. Liu, Y. Huo, and S. Wang, "Towards Deeper GCNs: Alleviating Over-Smoothing via Iterative Training and Fine-Tuning," Oct. 2025. doi: 10.1007/978-3-032-06066-2_13.

[3] R. Hasani *et al.*, "Closed-form continuous-time neural networks," *Nat. Mach. Intell.*, vol. 4, no. 11, pp. 992–1003, Nov. 2022, doi: 10.1038/s42256-022-00556-7.

[4] Y. Meng *et al.*, "Preserving and combining knowledge in robotic lifelong reinforcement learning," *Nat. Mach. Intell.*, vol. 7, no. 2, pp. 256–269, Feb. 2025, doi: 10.1038/s42256-025-00983-2.

[5] "Linear discriminant analysis," *Nat. Rev. Methods Primer*, vol. 4, no. 1, p. 71, Sept. 2024, doi: 10.1038/s43586-024-00357-9.

[6] "AI for Science 2025", Accessed: Oct. 12, 2025. [Online]. Available: https://www.nature.com/articles/d42473-025-00161-3

[7] "What is Feature Engineering?," GeeksforGeeks. Accessed: Oct. 12, 2025. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/what-is-feature-engineering/

[8] "Feature Engineering: Scaling, Normalization and Standardization," GeeksforGeeks. Accessed: Oct. 12, 2025. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/Feature-Engineering-Scaling-Normalization-and-Standardization/

[9] *Nonlinear Dimensionality Reduction*. Accessed: Oct. 12, 2025. [Online]. Available: https://link.springer.com/book/10.1007/978-0-387-39351-3

[10] "Linear vs. Non-linear Classification: Analyzing Differences Using the Kernel Trick - GeeksforGeeks." Accessed: Oct. 12, 2025. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/linear-vs-non-linear-classification-analyzing-differences-using-the-kernel-trick/

[11] "How can we Handle Multicollinearity in Linear Regression?," GeeksforGeeks. Accessed: Oct. 12, 2025. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/how-can-we-handle-multicollinearity-in-linear-regression/