Vector spaces

Recall that the initial definition of a vector was a list of numbers. An n component vector $\mathbf{u}$ is:

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

Now, we will also let the lists have an infinite length:

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ \vdots \end{bmatrix}$$

With **vector spaces**, vectors can now be any quantity that **can** be denoted by a list of numbers. The requirements for set V to be a vector space are as follows:

- There must be a system for which each object $\mathbf{v}$ from V is denoted by a unique list $[\mathbf{v}]_{\text{vec}}$ of numbers, also referred to as a **coordinate**. This list of numbers must have the same length (possibly infinite) across all objects from V, and the common length of the list is the number of "dimensions" of V, $\dim(V)$.

- Every object $\mathbf{v}$ from V is represented by exactly one list $[\mathbf{v}]_{\text{vec}}$, and every list $\mathbf{u}$ is associated with exactly one object $[\mathbf{u}]_{\text{obj}}$ from V.

- Given any two objects $\mathbf{u}$ and $\mathbf{v}$ from V, the sum of the objects $\mathbf{u}$ and $\mathbf{v}$ must agree with the vector sum of the lists:
$$\mathbf{u} + \mathbf{v} = [[\mathbf{u}]_{\text{vec}} + [\mathbf{v}]_{\text{vec}}]_{\text{obj}}$$

- Given any object $\mathbf{u}$ from V and any scalar c, the scalar vector product of $\mathbf{u}$ and c must agree with the scalar vector product involving the lists:

$$c\mathbf{u} = [c[\mathbf{u}]_{\text{vec}}]_{\text{obj}}$$

In order to satisfy the above properties, the list of numbers needs to be the coefficients of a linear combination generated from a basis set of V. A set A is a basis set of V if and only if

- The elements of A are linearly independent. No elements from A are linear combinations of other elements from A.

- Every object from V is a linear combination of objects from A, and no linear combinations of objects from A generate objects outside of V. $V = \text{span}(A)$.

With a basis set, representing an arbitrary object from V is done by listing the coefficients of A needed to generate said object.

The set of all polynomials is denoted by $\mathbb{P}$, and is a vector space. An arbitrary polynomial $p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + ...$ can be denoted by listing its coefficients:

$$[p(x)]_{\text{vec}} = [a_0 + a_1 x + a_2 x^2 + a_3 x^3 + ...]_{\text{vec}} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ \vdots \end{bmatrix}$$

The vector space $\mathbb{P}$ has an infinite number of dimensions: $\dim(\mathbb{P}) = +\infty$.
One possible basis set is: $A = \{1, x, x^2, x^3, ...\}$

The set of all polynomials with a degree of up to n is denoted by $\mathbb{P}_n$, and is a vector space. An arbitrary polynomial $p(x) = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n$ can be denoted by listing its coefficients:

$$[p(x)]_{\text{vec}} = [a_0 + a_1 x + a_2 x^2 + ... + a_n x^n]_{\text{vec}} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

The vector space $\mathbb{P}_n$ has $n+1$ dimensions: $\dim(\mathbb{P}_n) = n + 1$.
One possible basis set is: $A = \{1, x, x^2, ..., x^n\}$

The set of all $m \times n$ matrices is denoted by $\mathbb{M}_{m,n}$, and is a vector space. An arbitrary matrix $A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$ can be denoted by listing its entries, in this case column by column. For example if $m = 3$ and $n = 2$,

$$\left[\begin{bmatrix} 5 & 4 \\ -2 & 3 \\ 1 & -1 \end{bmatrix}\right]_{\text{vec}} = \begin{bmatrix} 5 \\ -2 \\ 1 \\ 4 \\ 3 \\ -1 \end{bmatrix} \quad \text{and} \quad \left[\begin{bmatrix} 0 & 1 \\ 8 & -3 \\ 2 & 6 \end{bmatrix}\right]_{\text{vec}} = \begin{bmatrix} 0 \\ 8 \\ 2 \\ 1 \\ -3 \\ 6 \end{bmatrix}$$

The vector space $\mathbb{M}_{m,n}$ has mn dimensions: $\dim(\mathbb{M}_{m,n}) = mn$.
One possible basis set is the set of all $m \times n$ matrices where exactly one entry is 1 and the other entries are all 0. For example, a basis set for $\mathbb{M}_{3,2}$ is:

$$A = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

The set of all functions on the interval $[a, b]$ is denoted by $C[a, b]$, and is a vector space. An arbitrary function $f : [a, b] \to \mathbb{R}$ can be denoted by listing the assignment to each of the real numbers from $[a, b]$. It is worth noting that the number of components is not only infinite, but uncountably infinite (super infinite). There is a dimension for each real number from the set $[a, b]$. One possible basis set for $C[a, b]$ is the set of all functions that return 1 for exactly 1 value of $x \in [a, b]$, and 0 for all other values of $x \in [a, b]$:

$$A = \left\{ f(x) = \left\{ \begin{array}{ll} 1 & (x = x_0) \\ 0 & (x \neq x_0) \end{array} \right. \middle| x_0 \in [a, b] \right\}$$

It should also be noted that there are multiple ways to quantify each object from a vector space as a list of numbers by choosing different basis sets for V. For example each polynomial $p(x) = a_2x^2 + a_1x + a_0$ from $\mathbb{P}_2$ can also be quantified by the 3 component list:

$$[a_2x^2 + a_1x + a_0]_{\text{vec}} = \begin{bmatrix} a_2 - a_1 \\ a_1 - a_0 \\ a_0 \end{bmatrix}$$

Equivalently,

$$\left[\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}\right]_{\text{obj}} = (c_1 + c_2 + c_3)x^2 + (c_2 + c_3)x + c_3$$

What is important is that there is an exact 1-to-1 correspondence between polynomials and lists, and that addition and scalar multiplication can be performed directly on the lists: given two polynomials $p(x) = a_2x^2 + a_1x + a_0$ and $q(x) = b_2x^2 + b_1x + b_0$,

$$p(x) + q(x) = (a_2 + b_2)x^2 + (a_1 + b_1)x + (a_0 + b_0)$$

and

$$\begin{bmatrix} a_2 - a_1 \\ a_1 - a_0 \\ a_0 \end{bmatrix} + \begin{bmatrix} b_2 - b_1 \\ b_1 - b_0 \\ b_0 \end{bmatrix} = \begin{bmatrix} (a_2 + b_2) - (a_1 + b_1) \\ (a_1 + b_1) - (a_0 + b_0) \\ a_0 + b_0 \end{bmatrix}$$

so

$$[p(x) + q(x)]_{\text{vec}} = [p(x)]_{\text{vec}} + [q(x)]_{\text{vec}}$$

Also given an arbitrary scalar c,

$$c \cdot p(x) = (c \cdot a_2)x^2 + (c \cdot a_1)x + (c \cdot a_0)$$

and

$$c \begin{bmatrix} a_2 - a_1 \\ a_1 - a_0 \\ a_0 \end{bmatrix} = \begin{bmatrix} (c \cdot a_2) - (c \cdot a_1) \\ (c \cdot a_1) - (c \cdot a_0) \\ (c \cdot a_0) \end{bmatrix}$$

so

$$[c \cdot p(x)]_{\text{vec}} = c \cdot [p(x)]_{\text{vec}}$$

In this case the basis set that generates the polynomial to list correspondence is:

$$A = \left\{ x^2, x^2 + x, x^2 + x + 1 \right\}$$

$$\left[\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}\right]_{\text{obj}} = c_1x^2 + c_2(x^2 + x) + c_3(x^2 + x + 1)$$

If each each polynomial $p(x) = a_2x^2 + a_1x + a_0$ from $\mathbb{P}_2$ were instead quantified by the 3 component list $[a_2x^2 + a_1x + a_0]_{\text{vec}} = \begin{bmatrix} a_2 + 1 \\ a_1 \\ a_0 \end{bmatrix}$, there is still an exact 1-to-1 correspondence between polynomials and lists, but the addition of polynomials no longer corresponds to the addition of lists: given two polynomials $p(x) = a_2x^2 + a_1x + a_0$ and $q(x) = b_2x^2 + b_1x + b_0$,

$$p(x) + q(x) = (a_2 + b_2)x^2 + (a_1 + b_1)x + (a_0 + b_0)$$

but

$$\begin{bmatrix} a_2 + 1 \\ a_1 \\ a_0 \end{bmatrix} + \begin{bmatrix} b_2 + 1 \\ b_1 \\ b_0 \end{bmatrix} = \begin{bmatrix} (a_2 + b_2 + 1) + 1 \\ a_1 + b_1 \\ a_0 + b_0 \end{bmatrix} \neq \begin{bmatrix} (a_2 + b_2) + 1 \\ a_1 + b_1 \\ a_0 + b_0 \end{bmatrix}$$

so

$$[p(x) + q(x)]_{\text{vec}} \neq [p(x)]_{\text{vec}} + [q(x)]_{\text{vec}}$$

This does not invalidate the fact that $\mathbb{P}_2$ is a vector space. The coordinate system is simply invalid. No basis can generate the given polynomial to list correspondence.

In general terms, for there to exist an 1-to-1 correspondence between objects and lists, as well as for addition and scalar multiplication to properly correspond to addition and scalar multiplication for lists, there must be defined a basis set A for vector space V. Every object from V is a unique linear combination of the objects from A, and the corresponding list consists of the coefficients of the objects from A. Moreover, every linear combination of the objects from A generates an object from V. The 1-to-1 correspondence between objects and linear combinations of the objects from A establishes that addition and scalar multiplication of the lists is consistent with addition and scalar multiplication involving the objects from V.

Returning to the example where every polynomial $p(x) = a_2 x^2 + a_1 x + a_0$ from $\mathbb{P}_2$ is quantified by the 3 component list $[a_2 x^2 + a_1 x + a_0]_{\text{vec}} = \begin{bmatrix} a_2 - a_1 \\ a_1 - a_0 \\ a_0 \end{bmatrix}$, the list $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ corresponds to the polynomial $\left[\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right]_{\text{obj}} = x^2$, the list $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ corresponds to the polynomial $\left[\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right]_{\text{obj}} = x^2 + x$, and the list $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ corresponds to the polynomial $\left[\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right]_{\text{obj}} = x^2 + x + 1$. The set $[A]_{\text{vec}} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^3$. The set $A = \{x^2, x^2 + x, x^2 + x + 1\}$ is a basis for $\mathbb{P}_2$. Every polynomial $p(x)$ from $\mathbb{P}_2$ has a unique list of coefficients $[p(x)]_{\text{vec}} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$ where $p(x) = c_1 x^2 + c_2(x^2 + x) + c_3(x^2 + x + 1)$.

Linear subspaces

Let V denote an arbitrary vector space. Any **subset** S of V is a vector space itself if and only if S is a **linear subspace** of V, which is to say that S is closed under addition and scalar multiplication.

If and only if S is a linear subspace of V, then there exists a basis set $A = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$, such that for every vector $\mathbf{v}$ from S, there exists a unique linear combination of the vectors from A that is equal to $\mathbf{v}$:

$$\mathbf{v} = c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2 + ... + c_k \mathbf{u}_k$$

Closure under addition and scalar multiplication also ensures that every choice of coefficients c_1, c_2, ..., c_k results in a vector from S.

The list of coefficients $(c_1, c_2, ..., c_k)$ then create a coordinate system wherein each vector $\mathbf{v}$ from S can be denoted by the list $\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}$ independent of the actual components that comprise $\mathbf{v}$ in the context of the "parent" vector space V. This makes S itself a vector space as per the above understanding.

Lastly, it should be noted that basis sets, and hence the dimension, may be infinite, but the math surrounding basis sets is unchanged. When a vector has an infinite number of components, the components may be indexed by quantities other than integers confined to the finite range $1, 2, ..., n$.

Examples:

- Consider the linear subspace S of $\mathbb{P}_2$ defined by $S = \text{span}\{4x^2 - 3, 2x + 5, 8x^2 + 2x - 1\}$. An arbitrary polynomial $p(x) \in S$ has the form $p(x) = a_2 x^2 + a_1 x + a_0$, but representing $p(x)$ via the coefficient list $\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$ is invalid since not every 3 element list will be associated with a polynomial from S. For example,

  the polynomial $p(x) = -8x^2 + 6x + 1$ has the coefficient list $\begin{bmatrix} 1 \\ 6 \\ -8 \end{bmatrix}$, but is not a polynomial from S.

  The "coordinate system" for S will differ from the coordinate system for $\mathbb{P}_2$ despite the fact that S is a subspace of $\mathbb{P}_2$. To properly associate each object from S with a list, a basis for S must be determined. This basis set will create a local coordinate system for S where every coordinate is valid and every polynomial from S has a a unique coordinate. For now, an arbitrary polynomial $p(x) = a_2 x^2 + a_1 x + a_0$ will have the standard representation $[a_2 x^2 + a_1 x + a_0]_{\text{vec}} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$ as a basis set for S is computed.

  The set of lists associated with $S = \text{span}\{4x^2 - 3, 2x + 5, 8x^2 + 2x - 1\}$ is:

  $$[S]_{\text{vec}} = \text{span} \left\{ \begin{bmatrix} -3 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 8 \end{bmatrix} \right\}$$

  Filtering out linearly dependent redundant vectors leaves behind the basis set:

  $$[A]_{\text{vec}} = \left\{ \begin{bmatrix} -3 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \\ 0 \end{bmatrix} \right\}$$

  so $A = \{4x^2 - 3, 2x + 5\}$ is a basis set for S. Every polynomial $p(x)$ from S is associated with a unique list $\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ where $p(x) = c_1(4x^2 - 3) + c_2(2x + 5)$.

- Now a coordinate system is not what is desired, but instead a determination of whether S is a valid vector space. If S is a subset of a known vector space V, then all that needs to be checked is whether S is a linear subspace of V. S is a linear subspace of V if and only if S is closed under addition and closed under scalar multiplication.

  * The span of any subset A of vector space V is a linear subspace of V, and is hence a vector space itself. $\text{span}(A)$ will always be a vector space.

  * Consider the set S of all polynomials $a_2 x^2 + a_1 x + a_0$ from $\mathbb{P}_2$ where $a_0 \geq 0$.
    Is S closed under addition? $a_0 \geq 0$ and $b_0 \geq 0$ imply that $a_0 + b_0 \geq 0$ so the addition of two polynomials from S remains in S. S is closed under addition.
    Is S closed under scalar multiplication? $p(x) = 1$ belongs to S, but $(-1)p(x) = -1$ does not belong to S. S is not closed under scalar multiplication.
    S is **not** a linear subspace of $\mathbb{P}_2$, and is hence not a vector space.

  * Consider the set S of all polynomials $a_2 x^2 + a_1 x + a_0$ from $\mathbb{P}_2$ where a_0, a_1, and a_2 are all integers.
    Is S closed under addition? Adding integers results in other integers so the addition of two polynomials from S remains in S. S is closed under addition.
    Is S closed under scalar multiplication? $p(x) = 1$ belongs to S, but $(1/2)p(x) = 1/2$ does not belong to S. S is not closed under scalar multiplication.
    S is **not** a linear subspace of $\mathbb{P}_2$, and is hence not a vector space.

Row vectors and row spaces

Recall that the transpose of a matrix swaps rows and columns:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}^T = \begin{bmatrix} a_{1,1} & a_{2,1} & \cdots & a_{m,1} \\ a_{1,2} & a_{2,2} & \cdots & a_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,n} & a_{2,n} & \cdots & a_{m,n} \end{bmatrix}$$

A $1 \times n$ matrix A is a matrix that describes a linear mapping that accepts an n component vector $\mathbf{x}$ as input and returns a 1 component vector (essentially a scalar) as output. Such a matrix extracts a single number from the input vector:

$$A\mathbf{x} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = a_1 x_1 + a_2 x_2 + ... + a_n x_n$$

Matrices that consist of a single row are referred to as **row vectors**. If the column vector of coefficients for a row vector is $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, then the row vector itself is

$$\mathbf{a}^T = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

Note that the row vector column vector product $\mathbf{a}^T \mathbf{x}$ is equal to the dot product $\mathbf{a} \bullet \mathbf{x}$.

For this discussion, and to aid in the understanding of row vectors, row vectors will be said to denote "linear features" of the input vector.

The matrix vector product $A\mathbf{x}$ is often interpreted as computing a linear combination of the columns of A with the entries of $\mathbf{x}$ serving as the coefficients. Here, the matrix vector product $A\mathbf{x}$ where A is an $m \times n$ matrix will be interpreted as the computing m linear features of the input vector $\mathbf{x}$, where each row is a linear feature. If the m linear features are $\mathbf{a}_1^T$, $\mathbf{a}_2^T$, ..., $\mathbf{a}_m^T$, then:

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_m^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{a}_1^T \mathbf{x} \\ \mathbf{a}_2^T \mathbf{x} \\ \vdots \\ \mathbf{a}_m^T \mathbf{x} \end{bmatrix}$$

so the resultant vector is the vector that consists of the m features of $\mathbf{x}$.

If each of the m features $\mathbf{a}_1^T \mathbf{x}$, $\mathbf{a}_2^T \mathbf{x}$, ..., $\mathbf{a}_m^T \mathbf{x}$ of $\mathbf{x}$ are known, then any linear combination of these features is known as well. Given arbitrary coefficients c_1, c_2, ..., c_m, then the linear feature $c_1 \mathbf{a}_1^T + c_2 \mathbf{a}_2^T + ... + c_m \mathbf{a}_m^T$ can be easily computed:

$$(c_1 \mathbf{a}_1^T + c_2 \mathbf{a}_2^T + ... + c_m \mathbf{a}_m^T)\mathbf{x} = c_1(\mathbf{a}_1^T \mathbf{x}) + c_2(\mathbf{a}_2^T \mathbf{x}) + ... + c_m(\mathbf{a}_m^T \mathbf{x})$$

The set of all linear combinations of the rows of matrix A is referred to as the **row space** of matrix A. Every linear feature from the row space can be computed for an arbitrary vector $\mathbf{x}$ when only the m initial features $\mathbf{a}_1^T \mathbf{x}$, $\mathbf{a}_2^T \mathbf{x}$, ..., $\mathbf{a}_m^T \mathbf{x}$ are known.

If a feature $\mathbf{b}^T$ is outside of the row space of A, then $\mathbf{b}^T\mathbf{x}$ cannot be determined if only $\mathbf{a}_1^T\mathbf{x}$, $\mathbf{a}_2^T\mathbf{x}$, ..., $\mathbf{a}_m^T\mathbf{x}$ are known. Feature $\mathbf{b}^T$ is "independent" of the features $\mathbf{a}_1^T$, $\mathbf{a}_2^T$, ..., $\mathbf{a}_m^T$, and this independence is linear independence. Feature $\mathbf{b}^T$ contributes information that cannot be computed from the features $\mathbf{a}_1^T$, $\mathbf{a}_2^T$, ..., $\mathbf{a}_m^T$. The "dimension" of the row space, referred to as the **row rank** of a matrix, is the maximum number of independent features that can be derived from the row space of A.

The "rank" of a matrix A, denoted by $\mathrm{rank}(A)$, was defined as the number of dimensions in the range of the linear mapping L_A denoted by A, or equivalently the number of dimensions of the column space of A. Since the range of L_A has $\mathrm{rank}(A)$ dimensions, each possible set of values that the m features can achieve can actually be fully and uniquely described by $\mathrm{rank}(A)$ numbers, and hence $\mathrm{rank}(A)$ is the maximum number of independent features that the row space will allow. The dimension of the row space is $\mathrm{rank}(A)$, and therefore the both the column rank and row rank have the same value of $\mathrm{rank}(A)$. This also means that:

$$\mathrm{rank}(A^T) = \mathrm{rank}(A)$$

Let matrix B be an $m \times m$ invertible matrix. Let the matrix product $A' = BA$ have the rows: $BA = \begin{bmatrix} (\mathbf{a}_1')^T \\ (\mathbf{a}_2')^T \\ \vdots \\ (\mathbf{a}_m')^T \end{bmatrix}$. The m features that correspond to the rows of A' can be computed from the m features that correspond to the rows of A using matrix B. Also since B is invertible, the m features that correspond to the rows of A can be computed from the m features that correspond to the rows of A' using matrix B^{-1}

$$\begin{bmatrix} (\mathbf{a}_1')^T\mathbf{x} \\ (\mathbf{a}_2')^T\mathbf{x} \\ \vdots \\ (\mathbf{a}_m')^T\mathbf{x} \end{bmatrix} = B \begin{bmatrix} \mathbf{a}_1^T\mathbf{x} \\ \mathbf{a}_2^T\mathbf{x} \\ \vdots \\ \mathbf{a}_m^T\mathbf{x} \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} \mathbf{a}_1^T\mathbf{x} \\ \mathbf{a}_2^T\mathbf{x} \\ \vdots \\ \mathbf{a}_m^T\mathbf{x} \end{bmatrix} = B^{-1} \begin{bmatrix} (\mathbf{a}_1')^T\mathbf{x} \\ (\mathbf{a}_2')^T\mathbf{x} \\ \vdots \\ (\mathbf{a}_m')^T\mathbf{x} \end{bmatrix}$$

This means that both A and A' convey the same "information". In total multiplying A by invertible matrix B to get the matrix $A' = BA$ does the following:

- Linear independence relationships among the columns of A hold for the columns of A' and vice versa. This fact is used when computing a basis for the column space of A to eliminate redundant columns.

- The null spaces of A and A' are equal. This fact is used to compute the null space of A.

- A and A' convey the same "information". This fact will be used to find a set of independent linear features that spans the row space of A, in other words a basis for the row space of A.

To compute a basis for the row space of A, matrix A will be converted via elementary row operations to a row reduced form A'. The features that form the nonzero rows of A' are linearly independent, and convey the same information as the rows of A. The nonzero rows of A' will be clearly linearly independent since exactly one row is nonzero in each of the pivot columns. All redundant rows are the zero rows at the bottom of A'.

Examples:

- The matrix

$$A = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

  computes the following linear traits of the vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$:

- The average of the first 2 entries, $y_1 = \frac{x_1 + x_2}{2}$.
- The average of all 3 entries, $y_2 = \frac{x_1 + x_2 + x_3}{3}$.

$$A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} (x_1 + x_2)/2 \\ (x_1 + x_2 + x_3)/3 \end{bmatrix}$$

If these traits are all that is known about the vector $\mathbf{x}$, then any other trait that is a linear combination of these traits/rows are also known. If c_1 and c_2 are arbitrary coefficients, then it is possible to compute:

$$c_1 y_1 + c_2 y_2 = c_1 \frac{x_1 + x_2}{2} + c_2 \frac{x_1 + x_2 + x_3}{3} = c_1 \begin{bmatrix} 1/2 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + c_2 \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= (c_1 \begin{bmatrix} 1/2 & 1/2 & 0 \end{bmatrix} + c_2 \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}) \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Let y_3 denote the value of the feature $y_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3$. Since $\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} = -2 \begin{bmatrix} 1/2 & 1/2 & 0 \end{bmatrix} + 3 \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$, it is possible to compute y_3 from y_1 and y_2 alone: $y_3 = -2y_1 + 3y_2$. Since y_3 is simply x_3, we are actually able to recover the value of x_3 knowing only y_1 and y_2.

Let y_4 denote the value of the feature $y_4 = \begin{bmatrix} 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 - x_2 + x_3$. Since the row vector $\begin{bmatrix} 1 & -1 & 1 \end{bmatrix}$ is not a linear combination of $\begin{bmatrix} 1/2 & 1/2 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$, it is impossible to compute $y_4 = x_1 - x_2 + x_3$ from $y_1 = \frac{x_1 + x_2}{2}$ and $y_2 = \frac{x_1 + x_2 + x_3}{3}$ alone.

1) Let

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

The row reduced form of A is:

$$A' = \begin{bmatrix} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 4 \end{bmatrix}$$

Each of the nonzero rows compute the features $y_1 = x_1 - 3x_4$, $y_2 = x_2 + 2x_4$, and $y_3 = x_3 + 4x_4$. Every possible linear feature that can be derived from the output values of A alone is a unique linear combination of the row vectors $\{\begin{bmatrix} 1 & 0 & 0 & -3 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & 4 \end{bmatrix}\}$. This set is a basis set for the row space of A, and it is clear that the number of linearly independent features, which is the row rank, is equal to the column rank which is 3.

2) Let

$$A = \begin{bmatrix} 1 & 1 & 2 & -3 \\ -1 & 0 & 3 & -10 \\ 2 & 1 & -3 & 13 \\ -4 & -1 & 11 & -39 \end{bmatrix}$$

The row reduced form of A is:

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

Each of the nonzero rows compute the features $y_1 = x_1 + x_4$, $y_2 = x_2 + 2x_4$, and $y_3 = x_3 - 3x_4$. Every possible linear feature that can be derived from the output values of A alone is a unique linear combination of the row vectors $\{\begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & -3 \end{bmatrix}\}$. This set is a basis set for the row space of A, and it is clear that the number of linearly independent features, which is the row rank, is equal to the column rank which is 3.

3) Let

$$A = \begin{bmatrix} 1 & 3 & 4 & -11 \\ -1 & -1 & 0 & 1 \\ 2 & 4 & 4 & -12 \end{bmatrix}$$

It was previously determined that the row reduced form of A is:

$$A' = \begin{bmatrix} 1 & 0 & -2 & 4 \\ 0 & 1 & 2 & -5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Each of the nonzero rows compute the features $y_1 = x_1 - 2x_3 + 4x_4$, and $y_2 = x_2 + 2x_3 - 5x_4$. Every possible linear feature that can be derived from the output values of A alone is a unique linear combination of the row vectors $\{\begin{bmatrix} 1 & 0 & -2 & 4 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 2 & -5 \end{bmatrix}\}$. This set is a basis set for the row space of A, and it is clear that the number of linearly independent features, which is the row rank, is equal to the column rank which is 2.

4) Let

$$A = \begin{bmatrix} -3 & 9 & -6 & -9 \\ -4 & 12 & -3 & -7 \\ 2 & -6 & 0 & 2 \\ -1 & 3 & 1 & 0 \end{bmatrix}$$

The row reduced form of A is:

$$A' = \begin{bmatrix} 1 & -3 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Each of the nonzero rows compute the features $y_1 = x_1 - 3x_2 + x_4$, and $y_2 = x_3 + x_4$. Every possible linear feature that can be derived from the output values of A alone is a unique linear combination of the row vectors $\{\begin{bmatrix} 1 & -3 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix}\}$. This set is a basis set for the row space of A, and it is clear that the number of linearly independent features, which is the row rank, is equal to the column rank which is 2.

5) Let

$$A = \begin{bmatrix} 1 & 2 & -3 & -1 & -5 & 13 & 4 & -17 \\ 1 & 1 & -1 & 0 & 1 & 9 & 2 & -8 \\ 0 & 1 & -2 & 1 & 4 & 10 & 0 & 3 \\ 1 & 1 & -1 & 1 & 6 & 12 & 2 & -8 \\ 2 & -1 & 4 & 0 & 5 & -3 & 1 & -7 \end{bmatrix}$$

The row reduced form of A is:

$$A' = \begin{bmatrix} 1 & 0 & 1 & 0 & 2 & 2 & 0 & 1 \\ 0 & 1 & -2 & 0 & -1 & 7 & 0 & 3 \\ 0 & 0 & 0 & 1 & 5 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Each of the nonzero rows compute the features $y_1 = x_1 + x_3 + 2x_5 + 2x_6 + x_8$, $y_2 = x_2 - 2x_3 - x_5 + 7x_6 + 3x_8$, $y_3 = x_4 + 5x_5 + 3x_6$, and $y_4 = x_7 - 6x_8$. Every possible linear feature that can be derived from the output values of A alone is a unique linear combination of the row vectors $\{\begin{bmatrix} 1 & 0 & 1 & 0 & 2 & 2 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & -2 & 0 & -1 & 7 & 0 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 1 & 5 & 3 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & -6 \end{bmatrix}\}$. This set is a basis set for the row space of A, and it is clear that the number of linearly independent features, which is the row rank, is equal to the column rank which is 4.

6) Let

$$A = \begin{bmatrix} -1 & -2 & 1 \\ 3 & 6 & 1 \\ 4 & 9 & -6 \\ -1 & 0 & 2 \end{bmatrix}$$

The row reduced form of A is:

$$A' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Each of the nonzero rows compute the features $y_1 = x_1$, $y_2 = x_2$, and $y_3 = x_3$. Every possible linear feature that can be derived from the output values of A alone is a unique linear combination of the row vectors $\{\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}\}$. This set is a basis set for the row space of A, and it is clear that the number of linearly independent features, which is the row rank, is equal to the column rank which is 3.

7) Let

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

It was previously determined that the row reduced form of A is:

$$A' = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Each of the nonzero rows compute the features $y_1 = x_2$. Every possible linear feature that can be derived from the output values of A alone is a unique linear combination of the row vectors $\{\begin{bmatrix} 0 & 1 \end{bmatrix}\}$. This set is a basis set for the row space of A, and it is clear that the number of linearly independent features, which is the row rank, is equal to the column rank which is 1.

8) Let

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The row reduced from of A is:

$$A' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

There are no nonzero rows. Only the zero row/feature, which always returns 0 and conveys no information, can be computed. Every possible linear feature that can be derived from the output values of A alone is a unique linear combination of the row vectors $\{\}$. Only the zero row/feature is a linear combination of the empty set $\{\}$. Zero row/features always return 0 and convey no information. This set is a basis set for the row space of A, and it is clear that the number of linearly independent features, which is the row rank, is equal to the column rank which is 0.