1.6 A Review of Matrix Theory

Notes

In this section admit general matrices: A is $m \times n$ of rank r.

<u>Definition:</u> The <u>rank</u> of A is the number, r, of linearly independent columns in A.

<u>1N</u> Ax is always a combination of the columns of A; it is in the <u>column space</u> of A, denoted $\mathcal{R}(A)$. The system $Ax=b$ has a solution exactly when b is also in the column space of A.

<u>Definition:</u> The nullspace of A, denoted $\mathcal{N}(A)$, is the set of vectors x satisfying $Ax=0$.

For $A \in \mathbb{R}^{m \times n}$:

The column space $\mathcal{R}(A)$ is a subset of $\mathbb{R}^m$ and has dimension r.
The nullspace $\mathcal{N}(A)$ is a subset of $\mathbb{R}^n$ and has dimension $n-r$.

<u>1O</u> If the matrix A has linearly independent columns then:

(1) The nullspace contains only the point $x=0$.
(2) The solution to $Ax=b$ is unique (if it exists).
(3) The rank is $r=n$.

Note that (3) implies we are assuming $n \leq m$. If $n > m$ then $r=n$ is not possible. We could not, for example, find 5 linearly independent vectors of length 4.

Proof: Suppose the columns $a_1, \dots, a_n$ of A are independent.

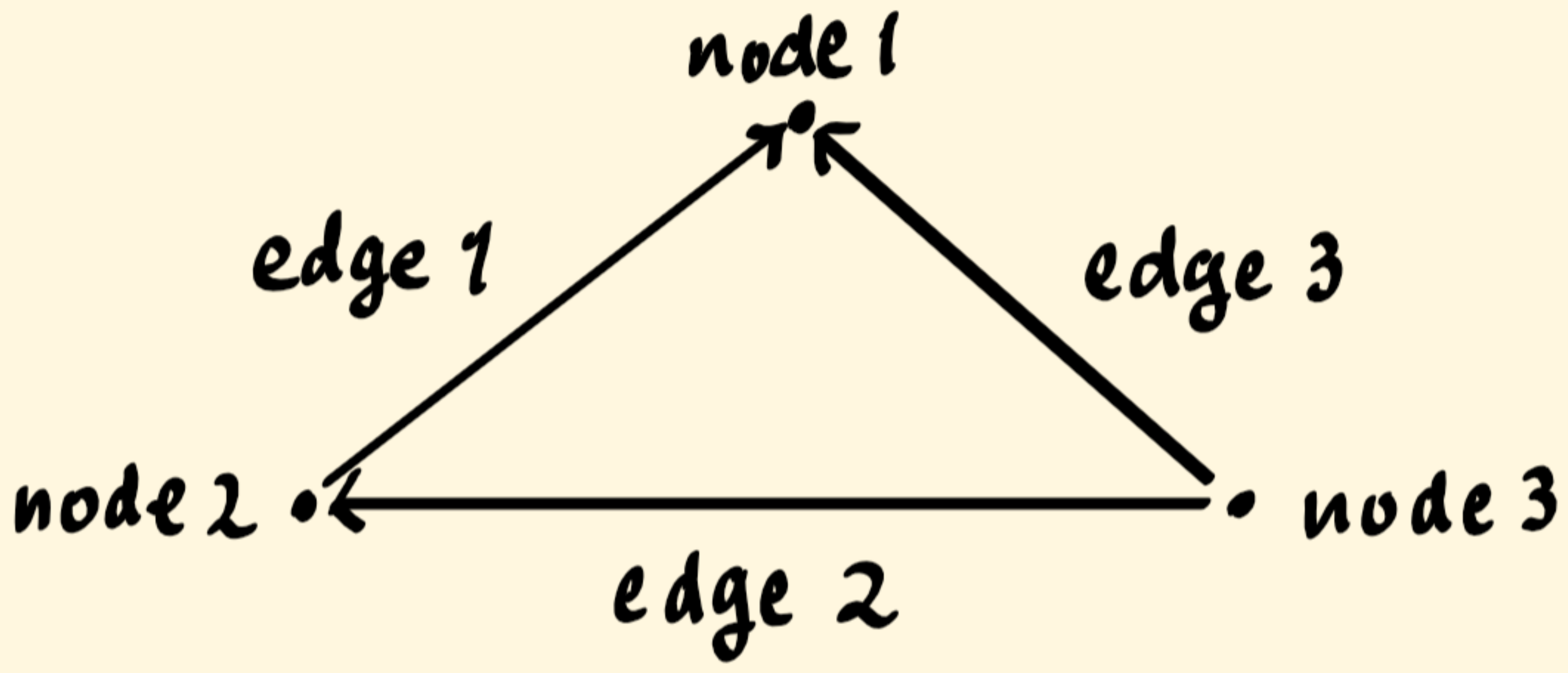
(1) $Ax=0 \rightarrow x_1 a_1 + \dots + x_n a_n = 0 \rightarrow x_i = 0 \;\forall i \rightarrow x=0$.

(2) $Ax = b = Ay \rightarrow A(x-y) = b-b = 0 \rightarrow x-y=0$ by (1) $\rightarrow x=y$

(3) A has n columns. We assumed these columns are independent. The rank is defined to be the number of independent columns. $\therefore r=n$.

Incidence Matrices

Graph:



Edge-node incidence matrix:

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

rows correspond to edges
columns correspond to nodes
+1 for an edge entering a node
−1 for an edge leaving a node

$Ax = b$:

$$\left.\begin{aligned} x_1 - x_2 &= b_1 \\ x_2 - x_3 &= b_2 \\ x_1 - x_3 &= b_3 \end{aligned}\right\} \rightarrow \begin{gathered}(x_1 - x_2) + (x_2 - x_3) = (x_1 - x_3) \\ b_1 + b_2 = b_3\end{gathered}$$

For $Ax=b$ to have a solution, we must have $y^T b = 0$, $y = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$

This is an example of <u>Kirchoff's voltage law</u>: the sum of the potential drops around a closed loop is zero.

Consider $A^T y = f$. Call A^T the <u>node-edge incidence matrix</u>.

$$A^T y = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} y_1 + y_3 \\ -y_1 + y_2 \\ -y_2 - y_3 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = f$$

Questions about $A^T y = f$:

(1) For which f can the system be solved?

$$(y_1 + y_3) + (-y_1 + y_2) - (y_2 + y_3) = 0 \longrightarrow f_1 + f_2 + f_3 = 0$$

(The y components sum to 0 $\longrightarrow$ The f components sum to 0)

The column space of A^T, $\mathcal{R}(A^T)$ contains vectors f s.t. $x^T f = 0$, $x^T = [1\ 1\ 1]$.

(2) Are the solutions to $A^T y = f$ unique?

Any vector $y_0 = [c, c, -c]^T$ gives $A^T y_0 = 0$ and so for any solution y of $A^T y = f$, $A^T(y + y_0) = f$ as well. The solution of $A^T y$ is not unique.

Note that $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is in $\mathcal{N}(A)$ and is perpendicular to the columns of A^T.

Note that $y = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is in $\mathcal{N}(A^T)$ and is perpendicular to the columns of A.

From each matrix A come <u>four fundamental subspaces</u>, two from $Ax = b$ and two from $A^T y = f$:

$\mathcal{R}(A)$: column space of A, a subspace of $\mathbb{R}^m$

$\mathcal{N}(A)$: nullspace of A, a subspace of $\mathbb{R}^n$

$\mathcal{R}(A^T)$: row space of A, a subspace of $\mathbb{R}^n$

$\mathcal{N}(A^T)$: left nullspace of A, a subspace of $\mathbb{R}^m$

The dimensions and perpendicularity of these four spaces make up:

<u>1P</u> (The Fundamental Theorem of Linear Algebra)

The row space and null space are perpendicular to each other and their dimensions add up to n:

(i) $\mathcal{R}(A^T) \perp \mathcal{N}(A)$

(ii) $\dim \mathcal{R}(A^T) + \dim \mathcal{N}(A) = r + (n-r) = n$

The column space and left nullspace are perpendicular to each other and their dimensions add up to m:

(iii) $\mathcal{R}(A) \perp \mathcal{N}(A^T)$

(iv) $\dim \mathcal{R}(A) + \dim \mathcal{N}(A^T) = r + (m-r) = m$

Proof:

$$\text{(i)} \left.\begin{array}{l} f \in \mathcal{R}(A^T) \rightarrow \exists y \text{ s.t. } A^T y = f \\ x \in \mathcal{N}(A) \rightarrow Ax = 0 \end{array}\right\} \rightarrow x^T f = x^T A^T y = (Ax)^T y = 0^T y = 0$$

$$\text{(iii)} \left.\begin{array}{l} b \in \mathcal{R}(A) \rightarrow \exists x \text{ s.t. } Ax = b \\ y \in \mathcal{N}(A^T) \rightarrow A^T y = 0 \end{array}\right\} \rightarrow y^T b = y^T A x = (A^T y)^T x = 0^T x = 0$$

The theorem claims $\dim \mathcal{R}(A) = \dim \mathcal{R}(A^T) = r$. That is, the number of independent columns of A equals the number of independent rows of A. One proof of this fact uses orthogonality (orthogonal = perpendicular here).

Let $A \in \mathbb{R}^{m \times n}$ with $\dim \mathcal{R}(A^T) = r$. Let $x_1, \ldots, x_r$ be a basis for $\mathcal{R}(A^T)$.

Consider $0 = c_1 A x_1 + \ldots + c_r A x = A(c_1 x_1 + \ldots + c_r x_r) = Av$

(a) $v = c_1 x_1 + \ldots + c_r x_r$ means $v \in \mathcal{R}(A^T)$

(b) $Av = 0$ means v is orthogonal to each row of A and thus orthogonal to any linear combination of the rows of A ($= \mathcal{R}(A^T)$).

$$\left.\begin{array}{l}\text{(a)}\\ \text{(b)}\end{array}\right\} \quad v \in \mathcal{R}(A^T) \;\&\; v \perp \mathcal{R}(A^T) \rightarrow v = 0$$

$\therefore c_1 x_1 + \ldots + c_r x_r = 0 \rightarrow c_1 = \ldots = c_r = 0$ since $x_1, \ldots, x_r$ are independent.

This means $Ax_1, \ldots, Ax_r$ are independent as well. Each Ax_i is a vector in $\mathcal{R}(A)$, so we have found r independent vectors in $\mathcal{R}(A)$.

$$\dim \mathcal{R}(A) \geq r = \dim \mathcal{R}(A^T)$$

Next suppose $\dim \mathcal{R}(A) = s$ and let $x_1, \ldots, x_s$ be a basis for $\mathcal{R}(A)$.

Consider $0 = c_1 A^T x_1 + \ldots + c_s A^T x_s = A^T(c_1 x_1 + \ldots + c_n x_s) = A^T u$. We have

$$\left.\begin{array}{l}\text{(a) } u \in \mathcal{R}(A)\\ \text{(b) } u \perp \mathcal{R}(A)\end{array}\right\} \longrightarrow u = 0 \longrightarrow c_1 = \ldots = c_s = 0 \rightarrow A^T x_1, \ldots, A^T x_s \text{ independent}$$

Since we have s independent vectors in $\mathcal{R}(A^T)$,

$$\dim \mathcal{R}(A) = s \leq \dim \mathcal{R}(A^T)$$

Conclude that since $\dim \mathcal{R}(A) \geq \dim \mathcal{R}(A^T)$ and $\dim \mathcal{R}(A) \leq \dim \mathcal{R}(A^T)$, we have $s = r$ and

$$\dim \mathcal{R}(A) = \dim \mathcal{R}(A^T) = r$$

Finally

(ii) If $\dim \mathcal{R}(A^T) = r$, A^T has r independent columns. Then the nullspace of A has a basis with $n-r$ vectors and $r + (n-r) = n$

(iv) If $\dim \mathcal{R}(A) = r$, A has r independent columns. Then the nullspace of A^T has a basis with $m-r$ vectors and $r + (m-r) = m$.

Another way to see the orthogonality relationships $\mathcal{R}(A^T) \perp \mathcal{N}(A)$ and $\mathcal{R}(A) \perp \mathcal{N}(A^T)$ is:

$$x \in \mathcal{N}(A) \longleftrightarrow Ax = 0$$

$$\longleftrightarrow \begin{bmatrix} \text{row } 1 \\ \vdots \\ \text{row } m \end{bmatrix} x = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

$\longleftrightarrow$ x orthogonal to each row

$\longleftrightarrow$ x orthogonal to any linear combination of rows

$\longleftrightarrow$ x orthogonal to each $f \in \mathcal{R}(A^T)$

$$y \in \mathcal{N}(A^T) \longleftrightarrow A^T y = 0$$

$$\longleftrightarrow \begin{bmatrix} \text{column } 1 \\ \vdots \\ \text{column } n \end{bmatrix} y = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

$\longleftrightarrow$ y orthogonal to each column

$\longleftrightarrow$ y orthogonal to any linear combination of columns

$\longleftrightarrow$ y orthogonal to each $b \in \mathcal{R}(A)$

Factorizations Based on A^TA

$A \in \mathbb{R}^{m \times n} \longrightarrow A^TA \in \mathbb{R}^{n \times n}$

A^TA is at least symmetric positive semidefinite:

$$x^TA^TAx = \|Ax\|^2 \geq 0 \quad \text{with equality iff } Ax = 0$$

Whether A^TA is positive definite then depends on whether there $x \neq 0$ s.t. $Ax = 0$. Nontrivial solutions exist iff the columns of A are dependent. Theorem 1R (1) follows from this.

1R

(1) If the columns of A are independent then A^TA is s.p.d.

(2) If the columns of A are dependent then A^TA is positive semi-definite but not invertible and $\lambda = 0$ is an eigenvalue of A^TA.

(3) $\mathcal{N}(A) = \mathcal{N}(A^TA)$, $\mathcal{R}(A^T) \underset{\uparrow}{=} \mathcal{R}((A^TA)^T)$, and $\operatorname{rank}(A) = \operatorname{rank}(A^TA)$

(A, A^TA have same rowspace)

Proof:

(2) If the columns of A are dependent $\exists\, x \neq 0$ s.t. $Ax = 0$ and for such x we have also $A^TAx = A^T0 = 0$. This implies A^TA is singular. Also $\lambda = 0$ is an eigenvalue for which any $x \neq 0$ s.t. $Ax = 0$ is an eigenvector:

$A^TAx = \vec{0} = 0x$ (used $\vec{0}$ to distinguish between scalar/vector zero)

(3) $\mathcal{N}(A) = \mathcal{N}(A^TA)$: $x \in \mathcal{N}(A) \longleftrightarrow Ax = 0 \longleftrightarrow A^TAx = 0 \longleftrightarrow x \in \mathcal{N}(A)$

Of these statements, $A^TAx = 0 \rightarrow Ax = 0$ is likely the least obvious. We use an argument that has been seen a few times in this section. Suppose x satisfies $A^TAx = 0$ and consider $Ax = b$. We have $b \in \mathcal{R}(A)$ and also $A^Tb = 0$. $A^Tb = 0$ implies b is orthogonal to every vector in $\mathcal{R}(A)$. In particular $0 = b^Tb$, which holds iff $b = 0$. Conclude that if $A^TAx = 0$ then $Ax = 0$.

$\operatorname{rank} A = \operatorname{rank} A^TA$: Suppose $\operatorname{rank} A = r$. We have seen that this means $\operatorname{rank} A^T = r$ as well. Recall $\operatorname{rank} A^T = \dim \mathcal{R}(A^T)$ by definition. From 1P, $r = n - \dim \mathcal{N}(A)$. Using $\mathcal{N}(A) = \mathcal{N}(A^TA)$, 1P, $A^TA = (A^TA)^T$, and the definition of $\operatorname{rank} A^TA$,

$$n = \dim \mathcal{R}((A^TA)^T) + \dim \mathcal{N}(A^TA) = \operatorname{rank} A^TA + \dim \mathcal{N}(A)$$

Conclude $\operatorname{rank} A^TA = n - \dim \mathcal{N}(A) = r = \operatorname{rank} A$.

$\mathcal{R}(A^T) = \mathcal{R}((A^TA)^T)$: $f \in \mathcal{R}(A^T) \leftrightarrow f \perp \mathcal{N}(A) \leftrightarrow f \perp \mathcal{N}(A^TA) \leftrightarrow f \in \mathcal{R}((A^TA)^T)$

Here we use the claim that $f \in \mathcal{R}(B^T)$ iff $f \perp \mathcal{N}(B)$ for a matrix B:

$f \in \mathcal{R}(B^T) \rightarrow \exists y$ s.t. $B^Ty = f \rightarrow x^Tf = x^TB^Ty = 0^Ty = 0 \;\forall x \in \mathcal{N}(B) \rightarrow f \perp \mathcal{N}(B)$

To prove $f \perp \mathcal{N}(B)$ implies $f \in \mathcal{R}(B^T)$ use orthogonal projection. For any subspace V of $\mathbb{R}^n$, any vector $f \in \mathbb{R}^n$ can be written as $f = v+u$, where $v \in V$ and $u \in V^\perp$, where $V^\perp$ is the orthogonal complement of V (the set of all vectors orthogonal to every vector in V). Since $\mathcal{N}(A)$ is a subspace of $\mathbb{R}^n$ and $\mathcal{N}(A)^\perp = \mathcal{R}(A^T)$ we have:

$$f = v + u \;, \quad v \in \mathcal{N}(A) \quad u \in \mathcal{R}(A^T)$$

Since $f \perp \mathcal{N}(B)$, $x^Tf = 0 \;\forall x \in \mathcal{N}(A)$. In particular $v^Tf = 0$, so

$$0 = v^Tf = v^Tv + v^Tu = v^Tv + 0 = v^Tv \rightarrow v = 0 \rightarrow f = u \in \mathcal{R}(A^T)$$

Conclude that $f \perp \mathcal{N}(B)$ implies $f \in \mathcal{R}(A^T)$.

Remark: An important observation follows from 1R:

$$Ax = 0 \text{ iff } A^TAx = 0 \text{ iff } x^TA^TAx = 0$$

(Recall $0 = x^TA^TAx = ||Ax||^2 \rightarrow Ax = 0$ so one could look at these statements as a 'logical loop').

Exercises

1.6.15 The fundamental theorem says either b is in the column space or b is not orthogonal to the left nullspace. Either:

(1) $Ax = b$ for some x

or

(2) $A^T y = 0$, $y^T b \neq 0$ for some y

Show directly that (1) and (2) cannot both be true.

Suppose $\exists\, x, y$ s.t. both (1) and (2) hold.

$$0 \neq y^T b = y^T A x = (A^T y)^T b = 0^T b = 0$$

This produces the contradiction $0 \neq 0$. $\therefore$ (1) and (2) cannot both be true.

1.6.24 Show that $A^T A$ can never have a negative eigenvalue.

$$A^T A x = \lambda x \longrightarrow x^T A^T A x = \lambda x^T x \longrightarrow \|Ax\|^2 = \lambda \|x\|^2 \longrightarrow \lambda = \|Ax\|^2 / \|x\|^2$$

$\|Ax\|^2, \|x\|^2 > 0$ since $x \neq 0$ (by definition of an eigenvector). $\therefore \lambda > 0$

1.6.21 Let $v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $w = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ so that $v^T w = 0$.

(a) What is $A = vw^T$?

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

(b) What are the eigenvalues of A?

$$0 = (1-\lambda)(-1-\lambda) + 1 = -1 - \lambda + \lambda + \lambda^2 + 1 = \lambda^2 \rightarrow \lambda = 0$$

(c) What are its eigenvectors?

$$\left.\begin{aligned} &\left[\begin{array}{cc|c} 1 & -1 & 0 \\ 1 & -1 & 0 \end{array}\right] \rightarrow x_1 = x_2 \quad \text{Take } x = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &\left[\begin{array}{cc|c} 1 & -1 & 1 \\ 1 & -1 & 1 \end{array}\right] \rightarrow y_1 = 1 + y_2 \quad \text{Take } y = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \end{aligned}\right\} \quad M = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$$

(d) What is the Jordan form of A?

$$J = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Since A has one eigenvalue $\lambda = 0$ with multiplicity 2, J consists of one 2×2 block.

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

1.6.19 True or False?

1. There is **no** matrix A whose row space contains $[1\ 2\ 1\ 1]^T$ and whose nullspace contains $[1\ {-2}\ 1\ 1]^T$.

   True

   Suppose $f = [1\ 2\ 1\ 1]^T \in \mathcal{R}(A^T)$ and $x = [1\ {-2}\ 1\ 1]^T \in \mathcal{N}(A)$
   By the fundamental theorem of linear algebra, $\mathcal{R}(A^T) \perp \mathcal{N}(A)$.
   We must have $x^T f = 0$. But $x^T f = -1$. From this contradiction conclude there can be no A s.t. $f \in \mathcal{R}(A^T)$ and $x \in \mathcal{N}(A)$.

2. Exactly one vector is in both the row space and the row space.

   True

   Suppose $x \in \mathcal{R}(A^T)$ and $x \in \mathcal{N}(A)$. Since $x \in \mathcal{N}(A)$, $Ax = 0$. Then x is orthogonal to every row of A and therefore orthogonal to every linear combination of rows of A. This means x is orthogonal to every vector in $\mathcal{R}(A^T)$. In particular $x \perp x$. The only vector orthogonal to itself in any vector space is the zero vector. There is exactly one zero vector in $\mathbb{R}^n$ (or any vector space for that matter).

3. If $\operatorname{rank} A = \operatorname{rank} B = 3$, then $\operatorname{rank}(A+B) \leq 6$.

   True

   In general, $\operatorname{rank}(A+B) \leq \operatorname{rank} A + \operatorname{rank} B$. The rank of $A+B$ is the dimension of the column space of $A+B$. If any vector y is in the column space of $A+B$, then y can be written as a linear combination of the vectors $a_1+b_1, a_2+b_2, \dots, a_n+b_n$. But this means y can be written as a linear combination of the vectors $a_1, \dots, a_n, b_1, \dots, b_n$. This means the column space of $A+B$ is a subset of the space spanned by $a_1, \dots, a_n, b_1, \dots, b_n$. Then $\operatorname{rank}(A+B)$, which is the dimension of the space spanned by the columns of $A+B$, is less than the dimension of the space spanned by $a_1, \dots, a_n, b_1, \dots, b_n$ (since the first is a subset of the second). We know there are $\operatorname{rank} A + \operatorname{rank} B$ independent vectors out of $a_1, \dots, a_n, b_1, \dots, b_n$ (= dimension of the span of these vectors). So $\operatorname{rank}(A+B) \geq \operatorname{rank} A + \operatorname{rank} B$. Conclude

$$\operatorname{rank}(A+B) \leq \operatorname{rank} A + \operatorname{rank} B = 6.$$

4. The rank of the matrix with every $a_{ij} = 1$ is 1.

<u>True</u>

In this case every column is the same – only 1 independent column.

5. The rank of the $n \times n$ matrix A with $a_{ij} = i + j$ is $r = n$.

<u>False</u>

Consider the counterexample given by the case $n = 3$:

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

Suppose $Ax = 0$: $\left.\begin{aligned} 2x_1 + 3x_2 + 4x_3 &= 0 \\ 3x_1 + 4x_2 + 5x_3 &= 0 \\ 4x_1 + 5x_2 + 6x_3 &= 0 \end{aligned}\right\}$ If $x_1 + x_2 + x_3 = 0$ and $2x_1 + 3x_2 + 4x_3 = 0$ then all 3 equations are satisfied.

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 2 & 3 & 4 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \end{array}\right] \rightarrow x_1 = x_3 \text{ and } x_2 = -2x_3$$

Take $x = [1 \ -2 \ 1]^T$. Then $Ax = 0$ for this nonzero x.
Conclude that $r = \text{rank}\, A < 3 = n$.