A Matrix Representation for the Combinatorics of Student Club Participation

Challenge Report 2

Math 115A

Arteen Abrishami

Problem Statement

A club is a **non-empty** set of students. No two clubs have the exact same set of students. That is, if $\exists$ two distinct clubs C_1, C_2 then $C_1 \neq C_2$.

Let $\exists n \in \mathbb{N}$ students, where $\mathbb{N} = \{1, 2, 3, ...\}$. We label the students from 1 to n in some arbitrary order. Similarly, let $\exists m \in \mathbb{N}$ clubs, labeled from 1 to m in some arbitrary order. Consider the function:

$$a_{ij} = \begin{cases} 1 & \text{if club } C_i \text{ contains student } S_j \\ 0 & \text{else} \end{cases}$$

We define an **incidence matrix** $A \in M_{m\times n}(\mathbb{R}) = [a_{ij}]$ such that the rows of the matrix correspond to the clubs and the columns to the students.

Later, it will be important to note that addition in $\mathbb{F}_2$ is defined as:

$$+_{\mathbb{F}_2} : \mathbb{F}_2 \times \mathbb{F}_2 \to \mathbb{F}_2$$

$$a +_{\mathbb{F}_2} b := (a +_{\mathbb{R}} b) \bmod 2$$

for $a, b \in \mathbb{F}_2$. For completeness, note that multiplication in $\mathbb{F}_2$ is defined as:

$$\cdot_{\mathbb{F}_2} : \mathbb{F}_2 \times \mathbb{F}_2 \to \mathbb{F}_2$$

$$a \cdot_{\mathbb{F}_2} b := (a \cdot_{\mathbb{R}} b) \bmod 2$$

for $a, b \in \mathbb{F}_2$.

Part 1 Counting Clubs

Let n be the number of students and m be the number of clubs that students are in.

(1) rank(A)

The rank of the incidence matrix $A \in M_{m\times n}(\mathbb{R})$ corresponds to the number of linearly independent columns in A. Considering that a column can be viewed as a student vector S_j, we can say that there are $rank(A)$ number of linearly independent vectors S_j, with each row representing whether they are in a given club C_i or not. Equivalently, we can say that the rank corresponds to the number of students whose club participation can form (as a linear combination) the club participation of the other students (those students corresponding to a linearly dependent vector S_j).

The upper bound on the rank of A is therefore n, as there can be at most n linearly independent columns, or vectors S_j, spanning the column space. In other words, the upper bound on the $rank(A)$ is the number of students whose club participation is being tracked. This occurs when each student S_j cannot be formed as a linear combination of the other students (or columns) in the matrix A.

Another way of looking at this is to say that the upper bound on $rank(A)$ is the dimension of the $Im(A)$, where $dim(Im(A)) \leq n$. On the occasion that $rank(A) = n$, which we call the matrix A having *full rank*, we know that the club participations of every student S_j are completely independent of one another, as linear combinations.

(2) AA^T

Considering the matrix AA^T. We attempt to generate a formula for its entries in terms of the entries of A and A^T.

$$A = \begin{bmatrix} a_{11} & ... & ... & a_{1n} \\ a_{21} & ... & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{m1} & ... & ... & a_{mn} \end{bmatrix}$$

$$A^T = \begin{bmatrix} a_{11} & ... & ... & a_{m1} \\ a_{12} & ... & ... & a_{m2} \\ ... & ... & ... & ... \\ a_{1n} & ... & ... & a_{mn} \end{bmatrix}$$

$$AA^T = \begin{bmatrix} a_{11} & ... & ... & a_{1n} \\ a_{21} & ... & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{m1} & ... & ... & a_{mn} \end{bmatrix} \begin{bmatrix} a_{11} & ... & ... & a_{m1} \\ a_{12} & ... & ... & a_{m2} \\ ... & ... & ... & ... \\ a_{1n} & ... & ... & a_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}a_{11} + a_{12}a_{12} + ... + a_{1n}a_{1n} & ... & a_{11}a_{m1} + a_{12}a_{m2} + ... + a_{1n}a_{mn} \\ a_{21}a_{11} + a_{22}a_{12} + ... + a_{2n}a_{1n} & ... & a_{21}a_{m1} + a_{22}a_{m2} + ... + a_{2n}a_{mn} \\ ... & ... & ... \\ a_{m1}a_{11} + a_{m2}a_{12} + ... + a_{mn}a_{1n} & ... & a_{m1}a_{m1} + a_{m2}a_{m2} + ... + a_{mn}a_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{k=1}^n a_{1k}a_{1k} & ... & \sum_{k=1}^n a_{1k}a_{mk} \\ \sum_{k=1}^n a_{2k}a_{1k} & ... & \sum_{k=1}^n a_{2k}a_{mk} \\ ... & ... & ... \\ \sum_{k=1}^n a_{mk}a_{1k} & ... & \sum_{k=1}^n a_{mk}a_{mk} \end{bmatrix}$$

Therefore, in general, we can say that $AA^T \in M_{m\times m}(\mathbb{R}) = [a^T_{ij}]$ where a^T_{ij} is the function defined by:

$$a_{ij}^T = \sum_{k=1}^{n} a_{ik} a_{jk}$$

in terms of the entries in the incidence matrix A.

The entry a_{ij}^T in $AA^T \in M_{m \times m}(\mathbb{R})$, otherwise known as the **intersection matrix**, describes how many students the clubs C_i and C_j have in common. Note that this matrix is therefore *symmetric*.

As a corrollary, the diagonal entries $\{a_{11}^T,\ a_{22}^T,\ ...\}$, where $i = j$, simply denote the number of students in the the corresponding club $\{C_1,\ C_2,\ ...\}$.

Part 2 Two Rules on Clubs

To recap, a club is a **non-empty** subset of students where $C_i \neq C_j$ when $i \neq j$. That is, no two distinct clubs have the exact same set of students.

We now impose two rules:

(a) Every club must have an *odd* number of members

(b) Every club must have an *even* number of members in common

Considering these rules, we find an upper bound on the number of possible clubs. Once again, let n be the number of students and m be the number of clubs that the students are in.

(3) Generalizing AA^T over $\mathbb{F}_2$

As a first step, we describe the entries in the matrix $AA^T \in M_{m \times m}(\mathbb{F}_2)$. Note that we are now considering AA^T as a matrix over the field $\mathbb{F}_2$, as opposed to $\mathbb{R}$. Recall the definitions of addition and multiplication in $\mathbb{F}_2$ from the abstract.

Consider then the entries of $AA^T \in M_{m\times m}(\mathbb{F}_2)$:

$$AA^T = \begin{bmatrix} \sum_{k=1}^n a_{1k}a_{1k} & ... & \sum_{k=1}^n a_{1k}a_{mk} \\ \sum_{k=1}^n a_{2k}a_{1k} & ... & \sum_{k=1}^n a_{2k}a_{mk} \\ ... & ... & ... \\ \sum_{k=1}^n a_{mk}a_{1k} & ... & \sum_{k=1}^n a_{mk}a_{mk} \end{bmatrix}$$

With the definition of addition and multiplication over $\mathbb{F}_2$, we find that a_{ij}^T is 1 if the clubs C_i, C_j share an odd number of members in common and 0 if the clubs C_i, C_j share an even number of members in common.

(4) Restricting AA^T to our Problem

With our restrictions that we defined earlier, denoted (a) and (b), we know that when $i = j$ then $A_{ij}^T = 1$ since every club has an odd number of members (therefore, every club has an odd number of members in common with itself). We also know that when $i \neq j$, then $A_{ij}^T = 0$, since every club may only have an even number of members in common. Therefore, our matrix $AA^T \in M_{m\times m}(\mathbb{F}_2)$ is exactly:

$$AA^T = \begin{bmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ 0 & 0 & ... & 0 \\ 0 & 0 & ... & 1 \end{bmatrix}$$

In other words, $AA^T = I_m$ where $I_m \in M_{m\times m}(\mathbb{F}_2)$ is the identity matrix with m rows and m columns over $\mathbb{F}_2$.

The rank of AA^T is therefore exactly m, as the identity matrix I_m has exactly m linearly independent columns. This is the number of student clubs.

(5) Proof on Limit on Student Clubs with Restrictions

Consider two arbitrary matrices A, B over a field $\mathbb{F}$, where we can form AB. Note that this means that A and B must be the appropriate dimensions. We want to show that $rank(AB) \leq min\{rank(A), rank(B)\}$.

To do this, we must show that $rank(AB) \leq rank(A)$ and $rank(AB) \leq rank(B)$.

Show: $rank(AB) \leq rank(A)$

Proof:

Let $v \in Im(AB)$.

$\exists u$ s.t. $v = ABu$ by definition.

Let $w = Bu$. So $v = Aw$.

Therefore, $v \in Im(A)$.

Conclude $Im(AB) \subseteq Im(A)$.

Therefore, $dim(Im(AB)) \leq dim(Im(A))$.

Therefore, $rank(AB) \leq rank(A)$. $\square$

Show: $rank(AB) \leq rank(B)$

Proof:

Let $\exists v \in ker(B)$.

Then $Bv = 0$.

Therefore, $ABv = A(Bv) = A(0) = 0$.

Conclude $v \in ker(AB)$.

So $ker(B) \subseteq ker(AB)$.

Therefore, $dim(ker(B)) \leq dim(ker(AB))$.

Equivalently, $nullity(B) \leq nullity(AB)$.

Let n be the dimension of the input space.

By rank-nullity, $rank(AB) + nullity(AB) = n$.

By previous proof, $nullity(B) \leq nullity(AB)$.

Therefore, $rank(AB) + nullity(B) \leq n$.

Equivalently, $rank(AB) \leq n - nullity(B)$.

Recall AB and B share the same input space.

By rank-nullity, $rank(B) + nullity(B) = n$

Equivalently, $rank(B) = n - nullity(B)$.

Therefore, we can say $rank(AB) \leq rank(B)$. $\square$

We conclude that $rank(AB) \leq min\{rank(A), rank(B)\}$, as desired.

Finally, we deduce an upper bound for the number of clubs that n students can participate in. We know that $AA^T = I_m \in M_{m\times m}(\mathbb{F}_2)$, where m is the number of clubs. From the above result, we know that

$$rank(AA^T) \leq min\{rank(A), rank(A^T)\}$$

Previously, we showed that $rank(AA^T) = m$. We also deduced that $rank(A) \leq n$, which is true for any field $\mathbb{F}$. From this, we can say that

$$rank(AA^T) \leq rank(A)$$

$$m \leq n$$

so that the number of clubs must always be less than or equal to the number of students. With that, we conclude our discussion of a matrix representation for the combinatorics of student club participation. $\square$