Preferences Resulting From Weighted Sums

Clay Thomas
claytont@princeton.edu

December 4, 2018

1 Motivation

Suppose students are told to rank different schools they would like to get into. The preferences of the students are likely correlated in some way due to an inherent notion of the quality of different schools. One way to define such a correlation is to assume there is some underlying space of quality along different attributes (e.g. STEM education vs liberal arts education) and the students' preference are determined by these attributes. The simplest instance of this is for each student to rank schools according to a weighted sum of the different attributes of the school.

We want to study the inherent complexity of the collection of preferences that result from these procedure, as a function of the number of attributes the schools have. In other words, what sort of correlation arises in the preferences of students in this model?

2 Definitions

Let $\{x_1, \ldots, x_n\} = X \subseteq \mathbb{R}^d_{\geq 0}$ be any set of points with nonnegative coordinates. Given any nonzero $a \in \mathbb{R}^d_{\geq 0}$, define a total order $>_a$ on X as follows: $x_i >_a x_j$ if and only if $\langle a, x_i \rangle > \langle a, x_j \rangle$, or $\langle a, x \rangle = \langle a, y \rangle$ and $i > j$ (note[1]). Let $P(X) = \{>_a | a \in \mathbb{R}^d_{\geq 0} \setminus \{0\}\}$. Thus, X represents the set of schools, the vectors a represent the preference weights of students, and $P(X)$ denotes the set of all possible preferences of schools.

Question. *As a function of d and n, how "rich" can $P(X)$ be (and what is the right notion of "richness")?*

Some observations:

- If $d = 1$, then $|P(X)| = 1$, i.e. students preferences are completely determined by the underlying set X.
- If $d = n$, then every linear preference on X can occur in $P(X)$.

[1] It's subtly important that ties are broken in the same way by every a. Otherwise, if X was n copies of the same point, any preference would be possible.

Proof. Let $X = \{e_i\}$ simply be the standard basis vectors. To induce an ordering $i_1, i_2, \ldots, i_n$, just create a preference vector a which gives weight $1/k$ to coordinate i_k. $\square$

The above hints that there should be some sort of continuum between $d = 1$ and $d = n$ of how complex the set $P(X)$ can be.

3 Lemmas

Here are some simple ways that the structure of X induces structure on $P(X)$:

Definition. *Let $x, y \in X \subseteq \mathbb{R}^d_{\geq 0}$. We say x* dominates *y, denoted $x \gg y$, if $x[k] > y[k]$ for each $k = 1, \ldots, d$.*

Proposition 1. *If $x \gg y$, then $x >_a y$ for any nonzero $a \in \mathbb{R}^d_{\geq 0}$.*

Proof. Simply observe $a_i x_i \geq a_i y_i$ for each $i = 1, \ldots, d$. Because $a \neq 0$, there is also some i where $a_i x_i > a_i y_i$. $\square$

Definition. *For points $x_1, \ldots, x_n \in \mathbb{R}^d$, let $\mathrm{hull}(x_1, \ldots, x_n) = \{u_1 x_1 + \ldots + u_n x_n | 0 \leq u_i \leq 1, \sum_{i=1}^n u_i = 1\}$ denote the convex hull of $x_1, \ldots, x_n$.*

Lemma 2. *Let $z, x_1, \ldots, x_k \in \mathbb{R}^d_{\geq 0}$ and $a \in \mathbb{R}^d_{\geq 0} \setminus \{0\}$. If $z >_a x_i$ for $i = 1, \ldots, k$, then $z >_a w$ for any $w \in \mathrm{hull}(x_1, \ldots, x_k)$.*

Proof. We have $\langle a, z \rangle > \langle a, x_i \rangle$ for each $i = 1, \ldots, k$. If $w = u_1 x_1 + \ldots + u_n x_n$ and $\sum_i u_i = 1$, then $\langle a, w \rangle = u_1 \langle a, x_1 \rangle + \ldots + u_n \langle a, x_n \rangle < u_1 \langle a, z \rangle + \ldots + u_n \langle a, z \rangle = \langle a, z \rangle$ $\square$

Proposition 3. *Let $z, x_1, \ldots, x_k \in \mathbb{R}^d_{\geq 0}$. Suppose that there exists $w \in \mathrm{hull}(x_1, \ldots, x_k)$ such that $w \gg z$. Then no a satisfies $z >_a x_i$ for each $i = 1, \ldots, k$.*

Proof. For contradiction, suppose such an a exists. Then $z >_a w$ as well. However, because $w \gg z$, this is a contradiction. $\square$

Let's look at how preferences are related when different weight vectors a are related.

Definition. *For any vectors $a_1, \ldots, a_k \in \mathbb{R}^d_{\geq 0}$, define $\mathrm{cone}(a_1, \ldots, a_k) = \{u_1 a_1 + \ldots + u_k a_k | u_i \geq 0 \forall j\}$. That is, the convex cone of vectors defined by $a_1, \ldots, a_k$.*

Proposition 4. *Suppose that for preference weights $a_1, \ldots, a_k$, we have $x >_{a_1} y, \ldots, x >_{a_k} y$. Then for any nonzero $b \in \mathrm{cone}(a_1, \ldots, a_k)$, $x >_b y$ as well.*

Proof. Let $b = u_1 a_1 + \ldots + u_k a_k$. We get $\langle b, x \rangle = u_1 \langle a_1, x \rangle + \ldots + u_k \langle a_k, x \rangle > u_1 \langle a_1, y \rangle + \ldots + u_k \langle a_k, y \rangle = \langle b, y \rangle$, as desired. $\square$

4 Impossibility Results for $d = 2$

Proposition 5. *If $d = 2$, then the preferences*

$$
\begin{aligned}
x &>_a y >_a z \\
y &>_b z >_b x \\
z &>_c x >_c y
\end{aligned}
$$

cannot occur for $x, y, z \in X \subseteq \mathbb{R}^2_{\geq 0}$.

Proof. Suppose the above preferences exist. Because $d = 2$, one of a, b, c must be contained in the convex cone of the other two, i.e. $d \in \{a, b, c\}$ with $d \in \text{cone}(e, f)$ where $\{e, f\} = \{a, b, c\} \setminus \{d\}$. (This is visually very obvious: given three different vectors in two dimensions, one must lie between the other two. You can prove it by considering the angle of the vectors from the x axis). Any pair of the above preferences has exactly one pair among x, y, z which they rank the same, so $u >_e v$ and $u >_f v$ for some u, v. However, this pair is reversed in the remaining preference, i.e. $u <_d v$. $\square$

More generally, we have the following:

Proposition 6. *Let $d = 2$ and $|X| = 3$. Then it is not possible for every point to be ranked highest by some preference, and ranked lowest by some other preference.*

Proof. Let $\{x, y, z\}$ be any set of distinct points in $\mathbb{R}^2_{\geq 0}$. Because $d = 2$, there must exist some $p \in \{x, y, z\}$ such that $p \gg w$ or $p \ll w$ for some $w \in \text{hull}(r, s)$, where $\{r, s\} = \{x, y, z\} \setminus \{p\}$ (again, a proof by picture is best here, but it can be formally proven). But then, no e satisfies $p >_e r$ and $p >_e s$, OR no e satisfies $p <_e r$ and $p <_e s$. $\square$

Corollary 7. *If $d = 2$, then the preferences*

$$
\begin{aligned}
x &>_a y >_a z \\
x &>_b z >_b y \\
y &>_c z >_c x \\
z &>_d y >_d x
\end{aligned}
$$

cannot occur for $x, y, z \in X \subseteq \mathbb{R}^2_{\geq 0}$.

It turns out that the above two patterns are, up to relabeling, the *only* impossible subsets of preferences with $|X| = 3$. We believe these patterns extend to higher dimensions, but the proofs above do not generalize because of the geometry involved.

5 Experimental Counts

The following gives upper bounds on the maximal number of preferences $|P(X)|$ as a function of n and d. I've only included entries if those points seem relatively close to the true maximum (i.e. when things seem pretty stable to checking more preference vectors or doing more trials).

n	d=2	d=3	d=4
3	4	6	6
4	7	18	24
5	10	41	85
6	15	87	
7	20	121	
8	24		
9	28		
10	39		
11	43		
12	48		