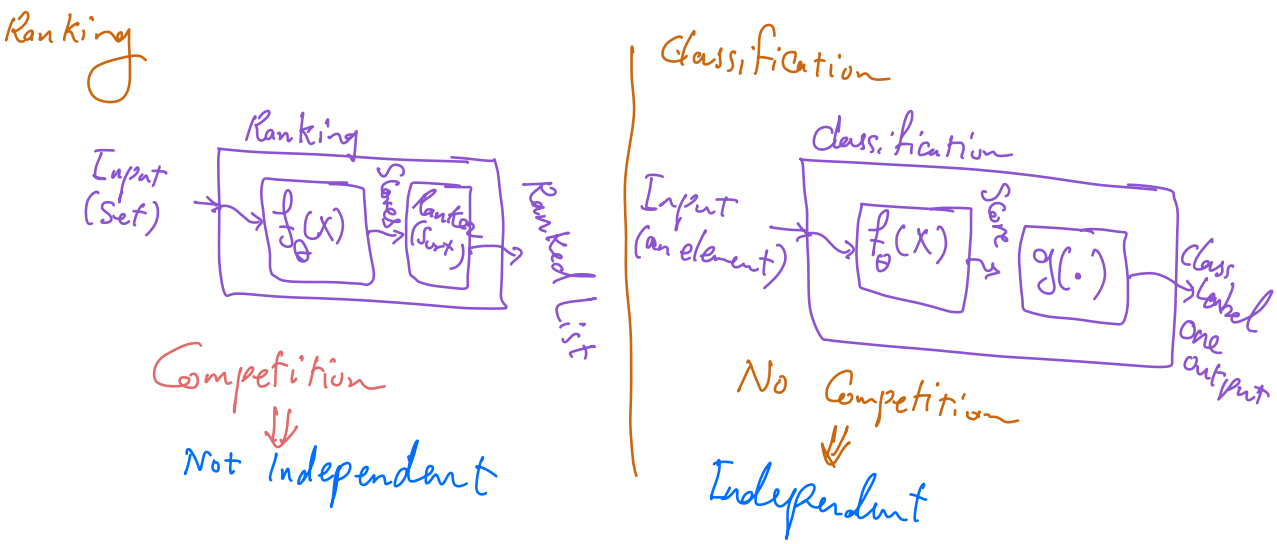
# Ranking

- Fix Data (to be ranked), Fix Ranking Function, Current Performance
  - Rank Billionaires Based on their Success (wealth)
  - Rank Universities Based on Some metrics?
- " " , (Expected) Future Performance
  - Ranking of University applicants based on expected Success?
    ↳ SAT + GPA + ...

- Variable Data, Variable Function, Current Performance
  - Search Results
    ↳ Every user has a pref. function in mind
    - For every query the Relevant Data is diff.

→ Query Answering System






| candidate | $A_1$ | $A_2$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $Y_1$ | $Y_2$ | $Y_3$ |
|---|---|---|---|---|---|---|---|---|---|
| b | male | White | 4 | 5 | 5 | {cs:0.9; art:0.2} | 14 | 9 | 1 |
| c | male | Asian | 5 | 3 | 4 | {math:0.9; cs:0.5} | 12 | 9 | 1 |
| d | female | White | 5 | 4 | 2 | {lit:0.8; math:0.8} | 11 | 4 | 6 |
| e | male | White | 3 | 3 | 4 | {math:0.8; econ:0.4} | 10 | 7 | 6 |
| f | female | Asian | 3 | 2 | 3 | {econ:0.9; math:0.5} | 8 | 5 | 8 |
| k | female | Black | 2 | 2 | 3 | {lit:0.9;art:0.8} | 7 | 1 | 9 |
| l | male | Black | 1 | 1 | 4 | {lit:0.5; math:0.7} | 6 | 6 | 2 |
| o | female | White | 1 | 1 | 2 | {econ:0.9; cs:0.8} | 4 | 7 | 8 |

(a)

| $\tau_1$ |
|---|
| b |
| c |
| d |
| e |
| f |
| k |
| l |
| o |

(b)

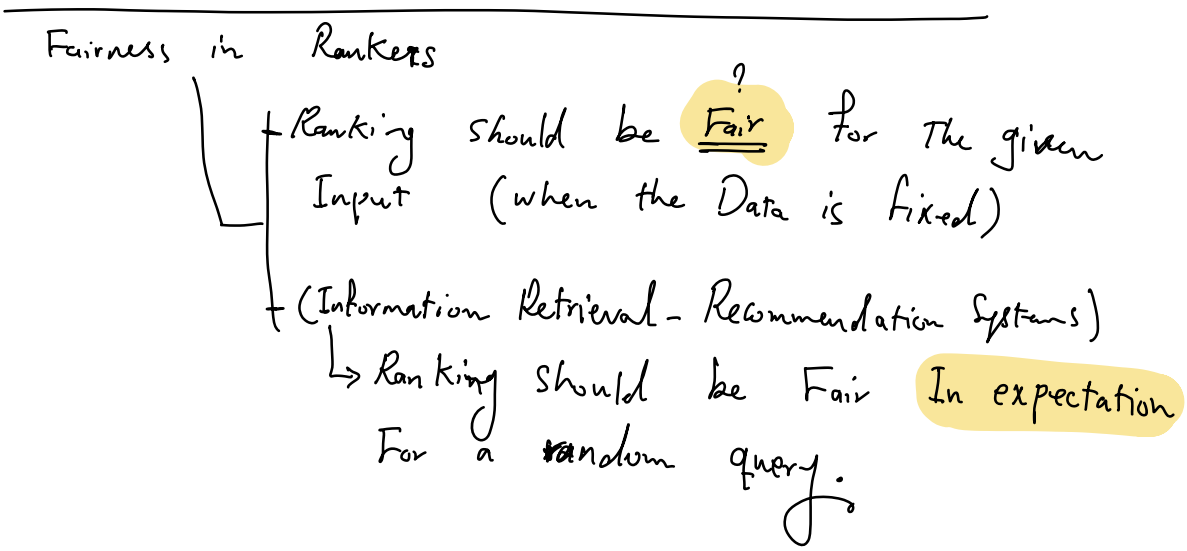
| $\tau_2$ |
|---|
| c |
| b |
| e |
| f |
| d |
| o |
| l |
| k |

(c)

| $\tau_3$ |
|---|
| k |
| l |
| b |
| d |
| e |
| f |
| c |
| o |

(d)

Fig. 1. (a) Dataset $C$ of college applicants, with demographic attributes $A_1$ (sex) and $A_2$ (race), numerical attributes $X_1$ (high school GPA), $X_2$ (verbal SAT), and $X_3$ (math SAT), and attribute $X_4$ (choice), that is, a vector extracted from the applicants' essays; (b) is a ranking $\tau_1$ on $Y_1$, computed as the sum of $X_1$, $X_2$, and $X_3$; (c) is a ranking on $Y_2$, predicted based on historical performance of STEM (cs, econ, math) majors; (d) is a ranking on $Y_3$, predicted based on historical performance of humanities (art, lit) majors. In all cases, the top-4 candidates will be interviewed in score order, and potentially admitted.



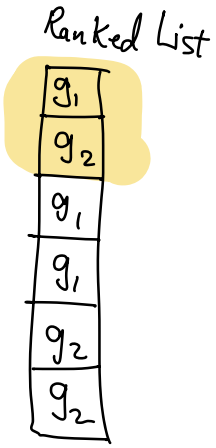
# Fairness Definitions For Ranking

- Parity at Top-k

Ranking is Fair if

$\forall g_i$

$$\frac{|g_i \cap \text{Top}_k|}{k} \simeq \frac{|g_i|}{n}$$




- Parity at moving top_k:

Instead of a single value k, looks at a set of possible values for k

$$k = \{10, 20, \dots\}$$

$$k = \{2, 4, \dots\}$$

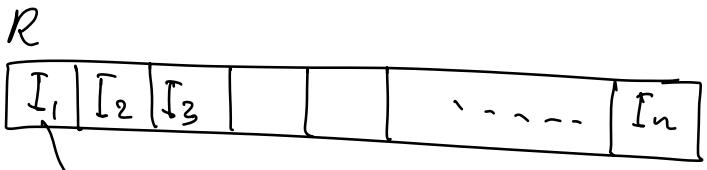
Normalization Term

$$rKD = \frac{1}{Z} \sum_{k=\{10,20,..\}} \frac{1}{\log(k+1)} \left( \frac{\rho_1}{\rho_2} - \frac{|g_1|}{|g_2|} \right)$$

The importance of Top-k @ k

$$\rho_i = \frac{|g_i \cap \text{Top\_k}|}{k}$$

Fairness in Exposure




(Aka Exposure) at

e.g.,

$$I_i = \frac{1}{\log(i+1)}$$

Exposure Parity : Equal Exposure for various groups

$$\frac{1}{|g_f|}\sum I_i \, \mathbb{1}(R_i \in g_j) = \frac{1}{|g_k|}\sum I_i \, \mathbb{1}(R_i \in g_k)$$

- In two-Sided Markets (Linked In),
  Fairness is two-Sided
  - The users who Initiate the queries
  - Those " are returned as the result.

---

- Utility Parity (Information Retrieval Systems)

- given a query $q$, every element $e$ has a relevance score $r(e, q)$

the utility of a ranking is defined as

$$U(R) = \sum_{i=1}^{n} I_i \, r(R_i, q)$$

$R_i$: the element at position $i$

$\Rightarrow$ Utility Parity

$$\frac{1}{|g_i|} \sum I_j \, r(R_j, q) \, \mathbb{1}(R_j \in g_i) =$$

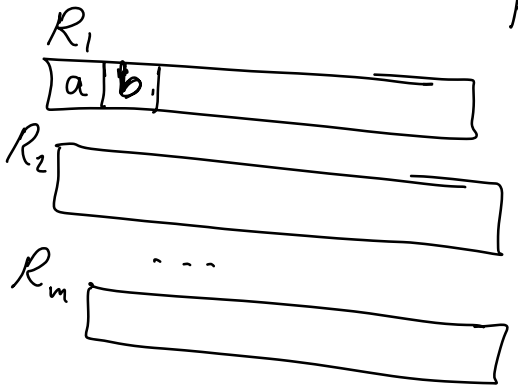
$$\frac{1}{|g_k|} \sum I_j \, r(R_j, q) \, \mathbb{1}(R_j \in g_k)$$

- Learn to Rank

Max Utility

S.t. Utility Parity

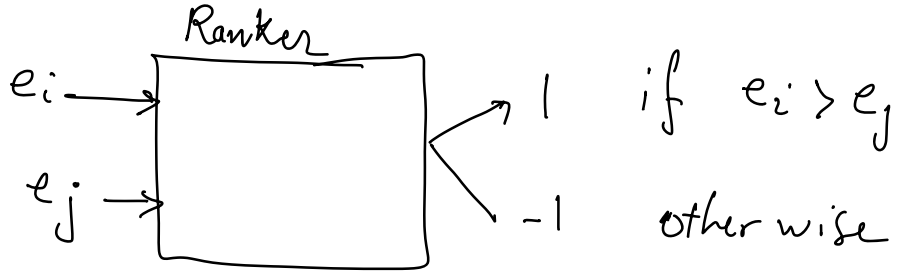
- Training Data : A set of Rankings



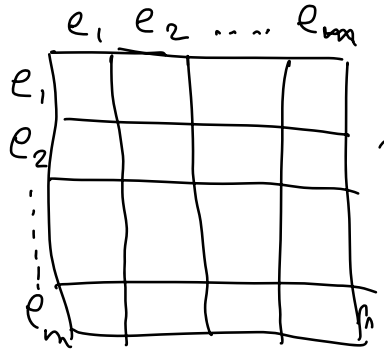

$r(a) > r(b)$

$e_i = \langle x_1, \cdots, x_d \rangle \; \langle q_i \rangle$

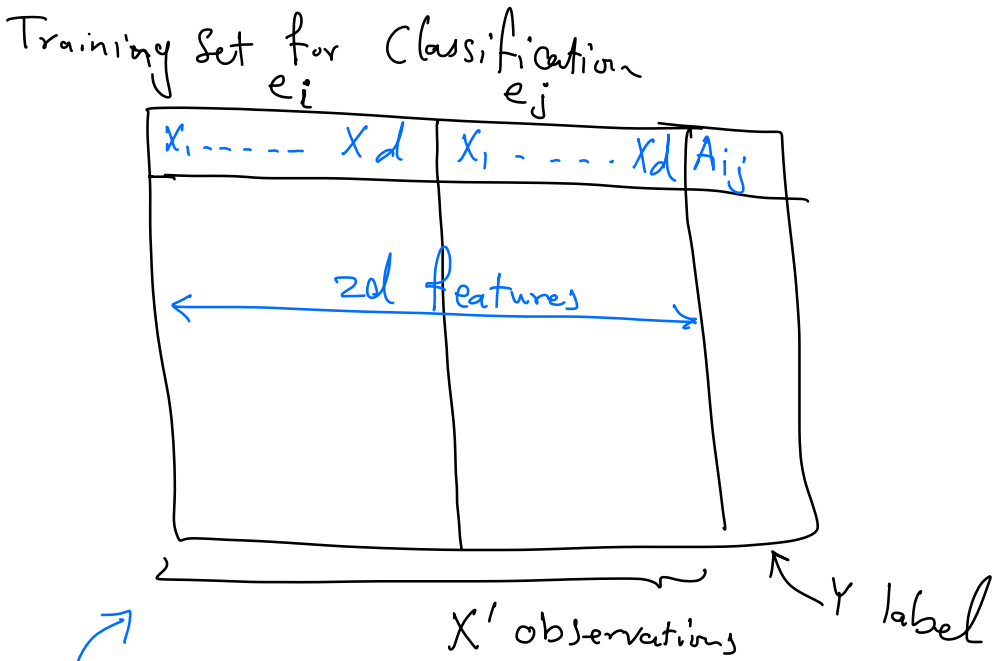
$\{e_1 > e_2 \;,\; e_3 > e_5 \;, \cdots\}$




Transformed Ranking to a Classification Instance




$$A[i,j] = \begin{cases} 1 & \text{if } e_i > e_j \\ -1 & \text{otherwise} \end{cases}$$




A ranker that is expected to be Fair

---

Suppose we have a fixed data to rank, we have an Initial ranking Func. & we want to minimally change the Function to generate a Fair output

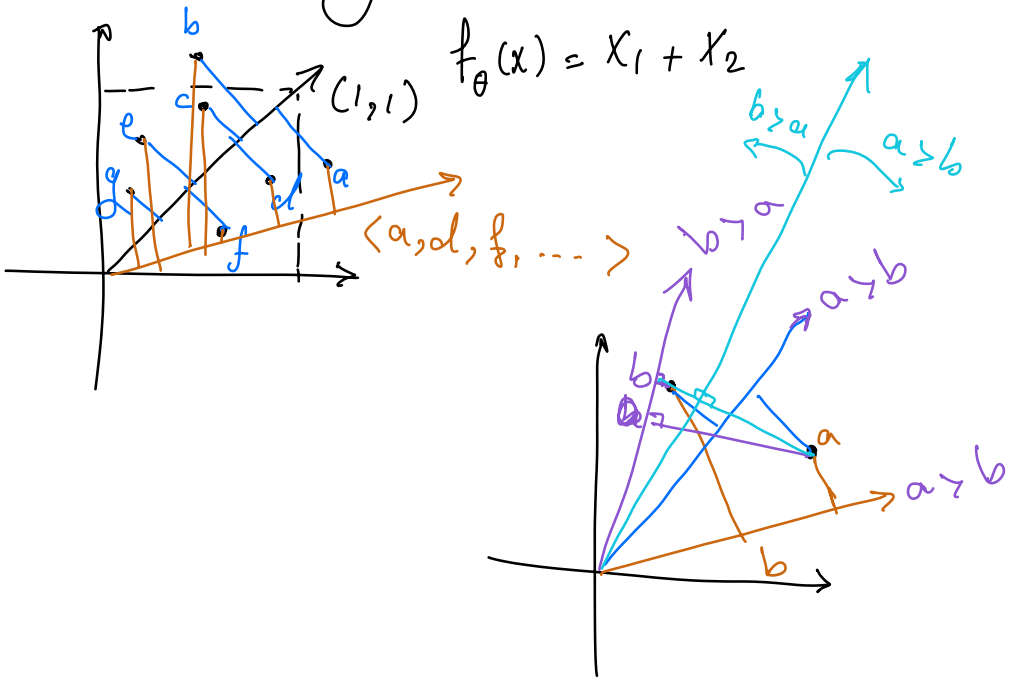
Students

| GPA | GRE | Gender | |
|---|---|---|---|
| .9 | .61 | M | 1.51 |
| .7 | .78 | F | 1.48 |
| .82 | .32 | M | 1.14 |
| ... | | ... | .. |
| | | | |

$f_\theta(X) : (\theta_1 = 1, \theta_2 = 1)$

$f_\theta(X) = \sum \theta_i X_i = \theta^T X$

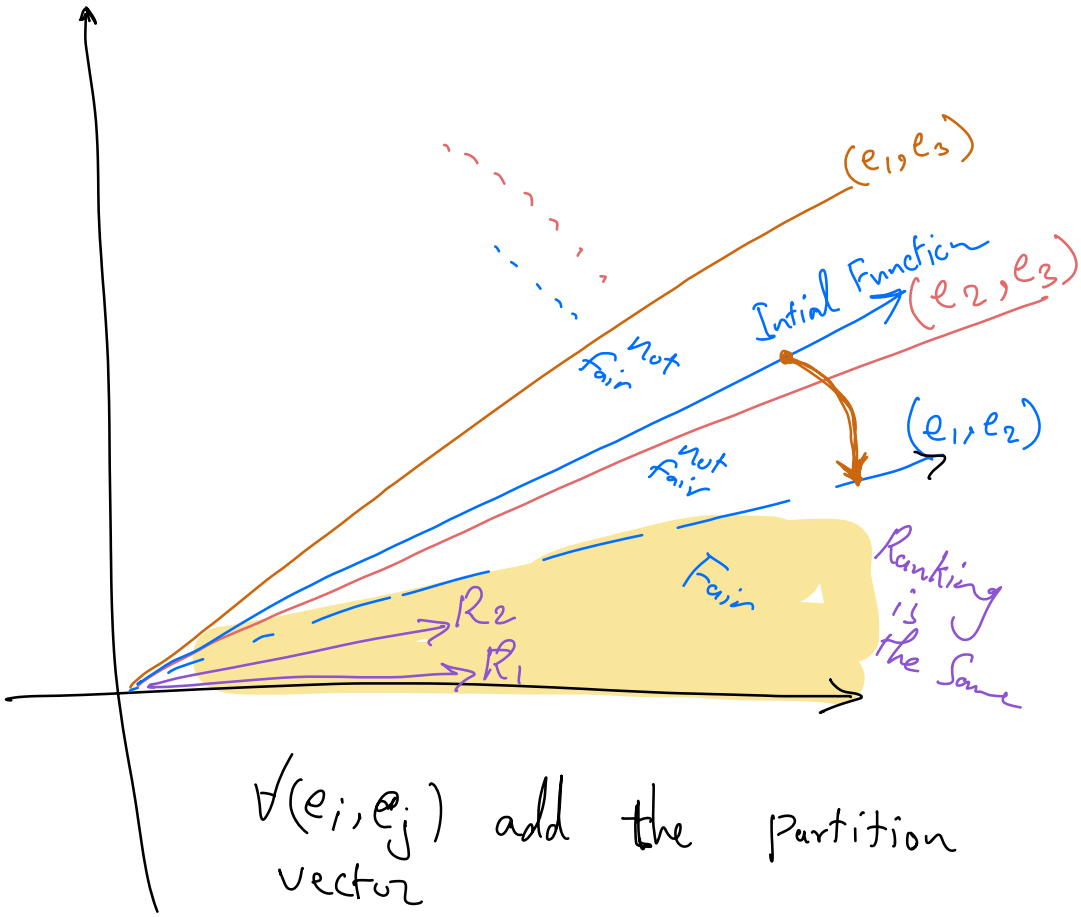
If $\langle 0.9, 1.1 \rangle$ is as good as $\langle 1, 1 \rangle$, $\langle 1.1, 0.9 \rangle$ Can you pick a one that generates a fair ranking




- For every pair of elements, there is a unique function that partitions the ranking functions in two groups $\langle L, R \rangle$

a>b ← L

R → b>a




$\forall (e_i, e_j)$ add the partition vector