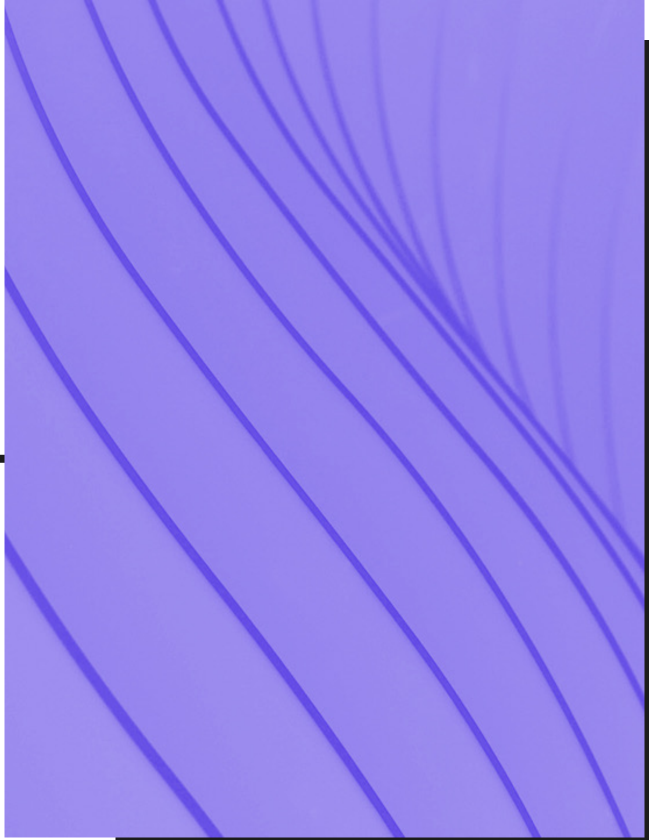
Holistic AI × The Alan Turing Institute

Bias in Clustering Systems Part III

Sachin Beepath
Assurance Officer, Holistic AI

Contents

Measuring Bias in Clustering



1. Formally introduce several group and individual fairness metrics.
2. Demonstrate how to measure fairness and bias in clustering systems.

Balance

For m protected groups and k clusters, we can define two ratios:

- r, the proportion of samples in the data set that belong to the protected group b
- r_a, the proportion of samples in cluster a belonging to group b

The balance is then defined over all clusters and protected groups as:

$$\min_{a \in [k], b \in [m]} \min\{R_{a,b}, \frac{1}{R_{a,b}}\}, R_{a,b} = \frac{r}{r_a},$$

- Balance can take values between 0 and 1.
- The closer the balance is to 1, the more fair the clustering is.

Balance (Example)

– Consider $m = 2$ protected groups and $k = 3$ clusters

– For group A we have 10 data points with cluster assignments: [1,2,1,3,2,1,1,1,2,3]

– For group B we have 8 data points with with cluster assignments: [2,3,1,2,1,2,3,3]

Balance (Example)

The proportions of group A are:

Entire dataset: $\frac{10}{18} = 0.56$

Cluster 1: $\frac{5}{7} = 0.71$

Cluster 2: $\frac{3}{6} = 0.5$

Cluster 3: $\frac{2}{5} = 0.4$

The proportions of group B are:

Entire dataset: $\frac{8}{18} = 0.44$

Cluster 1: $\frac{2}{7} = 0.29$

Cluster 2: $\frac{3}{6} = 0.5$

Cluster 3: $\frac{3}{5} = 0.6$

Balance (Example)

The ratios group group A are:

- $R_{A,1} = \frac{0.56}{0.71} = 0.79$
- $R_{A,2} = \frac{0.5}{0.56} = 0.89$
- $R_{A,3} = \frac{0.4}{0.56} = 0.71$

The ratios of group B are:

- $R_{B,1} = \frac{0.29}{0.44} = 0.66$
- $R_{B,2} = \frac{0.44}{0.5} = 0.88$
- $R_{B,3} = \frac{0.44}{0.6} = 0.73$

The balance of the clustering then 0.66

Cluster Distribution KL Divergence

- For $m = 2$ protected groups and k clusters, we can calculate the distribution of clusters amongst each group and calculate the KL divergence between the two distributions:

$$KL(P||Q) = \sum_k P(k) \log(\frac{P(k)}{Q(k)})$$

- Where $P(k)$ represents the cluster distribution for the first group and $Q(k)$ represents the cluster distribution for the second group.
- The more similar the distributions are, the fairer the system is.
- An ideal value is 0.

Cluster Distribution KL Divergence

- Consider $m = 2$ protected groups and $k = 3$ clusters.
- For group A we have 10 data points with cluster assignments: [1,2,1,3,2,1,1,1,2,3]
- For group B we have 8 data points with with cluster assignments: [2,3,1,2,1,2,3,3]
- $P(k) = [0.71, 0.5, 0.4]$
- $Q(k) = [0.29, 0.5, 0.6]$
- $KL(P||Q) = 0.71 \log \frac{0.71}{0.29} + 0.5 \log \frac{0.5}{0.5} + 0.6 \log \frac{0.6}{0.4} = 0.88$

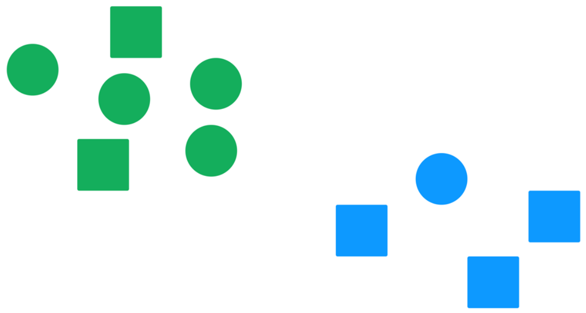
Part 3 Question 1

What is the cluster balance of the system?

Total: 5 Squares, 5 Circles

Blue: 3 Squares, 1 Circle

Green: 2 Squares, 4 Circles



Social Fairness

– The clustering cost O for a set cluster centres $U = \{U_1, \ldots, U_k\}$ and the input dataset X is defined as:

$$O(U, X) = \sum_{x \in X} \min_{u \in U} \|x - u\|^2$$

For m protected groups, let X_a be the samples of X that belong to protected group a, the social fairness cost is then:

$$\max_{a \in [m]} \frac{O(U, X_a)}{|X_a|}$$

Social Fairness

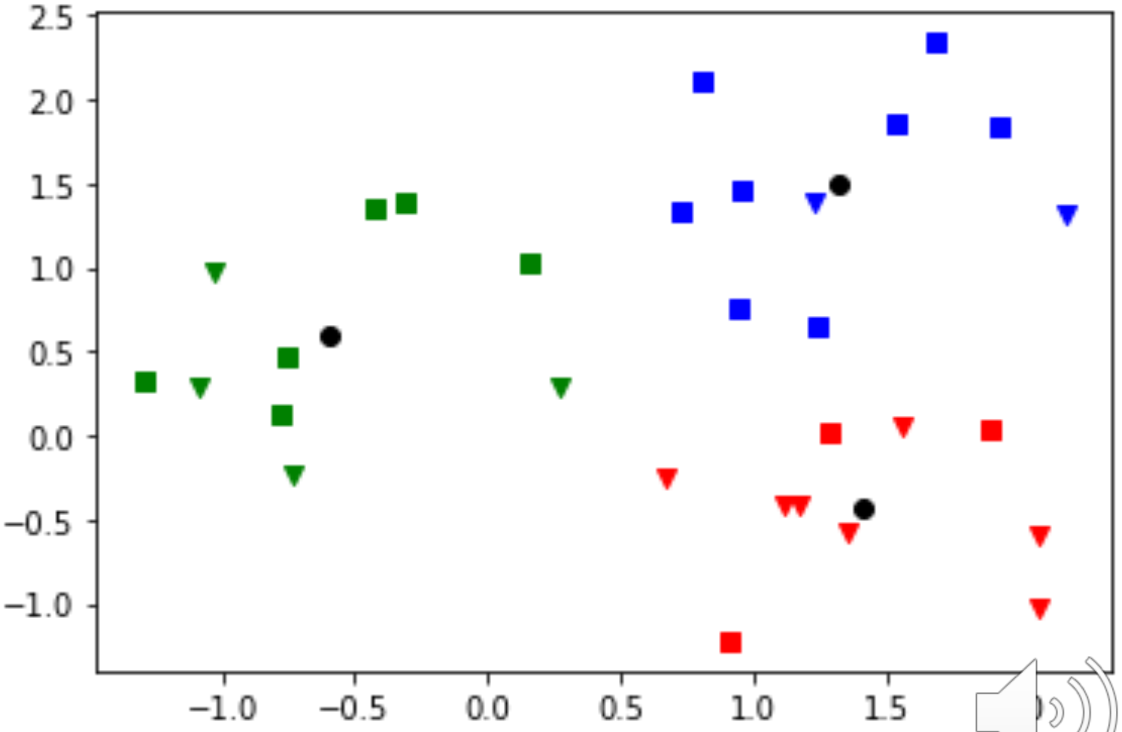
Clustering Costs:

– Squares: 8.52

– Triangles: 5.27

Social Fairness:

$$\max\left(\frac{8.52}{17}, \frac{5.27}{13}\right) = 0.501$$



Silhouette Difference

- For $m = 2$ protected groups and k clusters, the silhouette difference is the difference in silhouette scores for each protected group. For a given group, the silhouette score is:

$$\frac{d_b - d_a}{\max(d_a, d_b)}$$

where d_b is the mean nearest-cluster distance and d_a is the mean intra-cluster distance. The silhouette difference is defined as:

$$\frac{S_a - S_b}{2}$$

Silhouette Score and Silhouette Difference are bound between -1 and 1.

Silhouette Difference

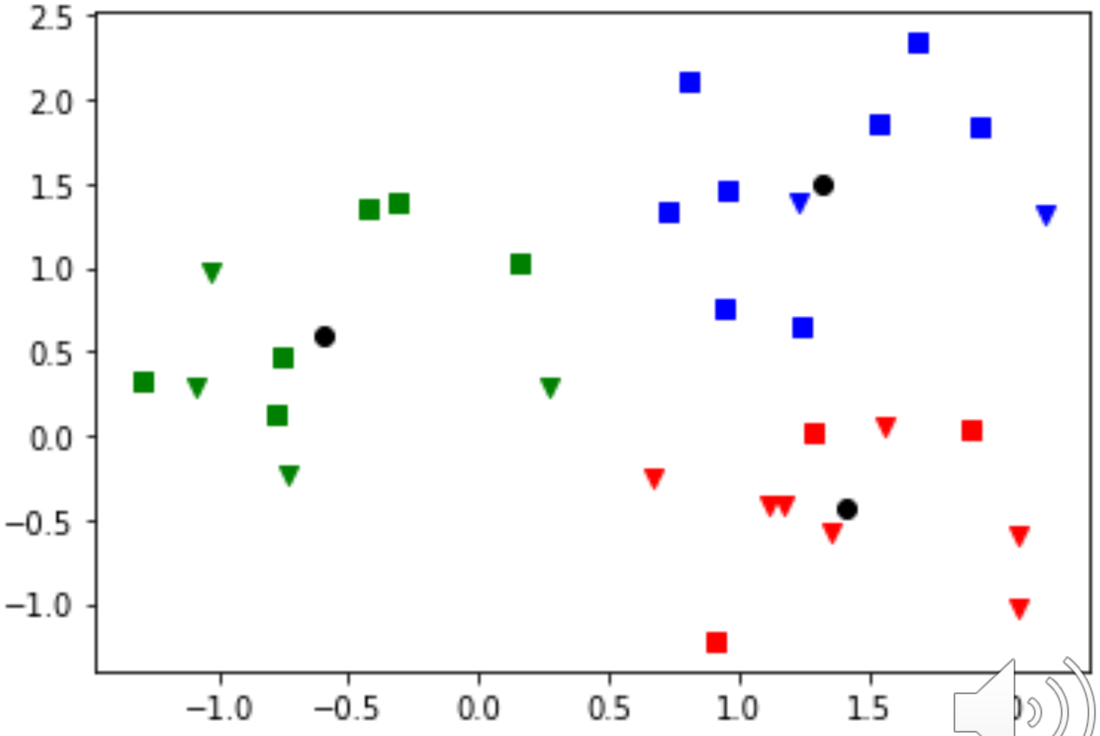
Average Silhouette Scores:

– Squares: 0.480

– Triangles: 0.552

Silhouette Difference:

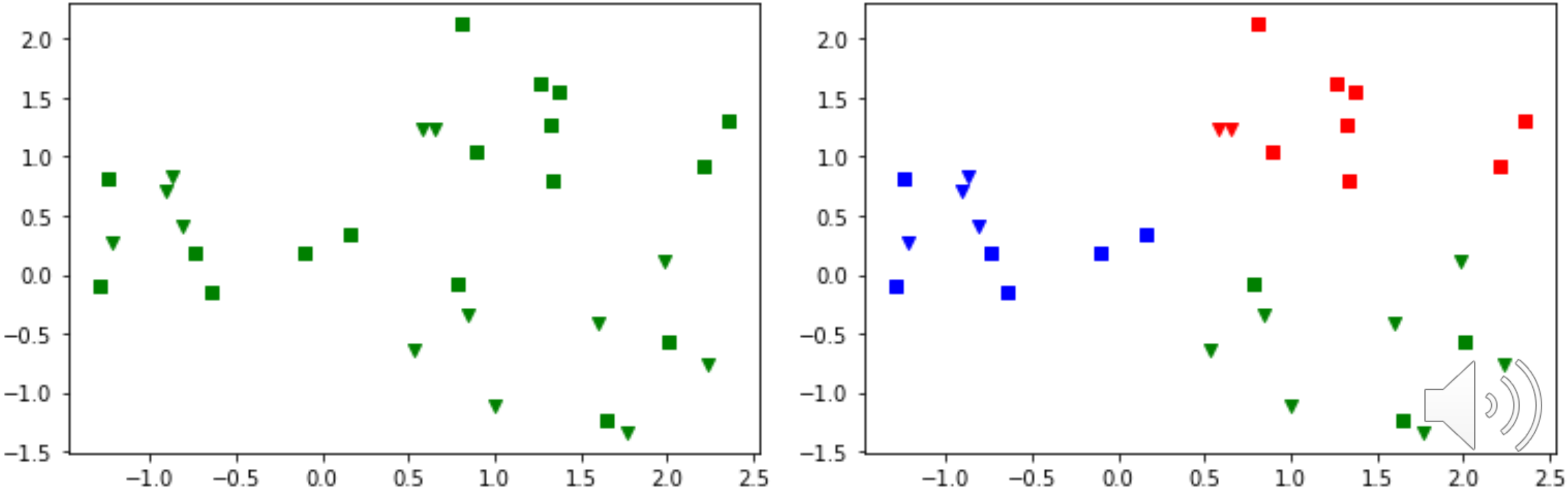
$$\frac{0.480 - 0.552}{2} = -0.036$$



Toy Example

– Use k-means to label data into 3 clusters

– 2 protected groups: Triangles and Squares



Toy Example (Cont.)

Metric	Value
Cluster Balance	0.462
Cluster Distribution KL Divergence	0.378
Cluster Social Fairness Ratio	1.022
Cluster Silhouette Difference	0.002

Individual Fairness Metrics

Proportionality

- For n data points and k clusters, any $\frac{n}{k}$ points are entitled to form their own cluster if there is another centre that is closer in distance for all $\frac{n}{k}$ points.

Aggregate Fairness

- Requires the distance of a point from its centroid to be at most α times the average distance of the points in its cluster to its centroid.

Conclusion

– Defined several group and individual fairness metrics in the context of clustering.

– Demonstrated how to measure fairness and bias using group fairness metrics.

Contents