Equity of Reproduction:

A reproduction of Amortizing Individual Fairness in Rankings

—

Thomas van Zwol, Sietze Kuilman, Emil Dudev, Nils Lehmann

The Paper

Asia J Biega, Krishna P Gummadi, and Gerhard Weikum.
Equity of attention: Amortizing individual fairness in rankings.
In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pages 405–414. ACM, 2018.

Problem

Disparity between ranking relevance and attention

- Define an unfairness metric

- Define a comparative quality metric for a new ranking

- Define an ILP to minimise the unfairness given a constraint on the ranking quality

The Mathematical Model

$$unfairness(\rho^1, \ldots \rho^m) = \sum_{i=1}^{n} |A_i - R_i| = \sum_{i=1}^{n} \left| \sum_{j=1}^{m} a_i^j - \sum_{j=1}^{m} r_i^j \right|$$

$$\text{minimize } \sum_{i=1}^{n} \sum_{j=1}^{n} |A_i^{l-1} + w_j - (R_i^{l-1} + r_i^l)| \cdot X_{i,j}$$

$$\text{subject to } \sum_{j=1}^{k} \sum_{i=1}^{n} \frac{2^{r_i^l} - 1}{log_2(j+1)} X_{i,j} \geq \theta \cdot IDCG@k$$

$$X_{i,j} \in \{0, 1\}, \forall_{i,j}$$

$$\sum_{i} X_{i,j} = 1, \forall_j$$

$$\sum_{j} X_{i,j} = 1, \forall_i$$

Our reproduction

We can affirm the unavoidable use of technical devices, and also deny them the right to dominate
us, and so to warp, confuse, and lay waste our nature.

-   Martin Heidegger

Data collection

Attempt to reproduce the experiment with the same data

- Synthetic dataset

- AirBnB

- StackExchange

Approach

Python library

- Model the ILP with PuLP

- Utilise the Gurobi solver through PuLP

- Simulate thousands of rankings and see how the unfairness progresses

Experiments

Synthetic dataset

100 subjects
Singular attention
Geometric attention

- Uniform

- Linear

- Exponential

Uniform dataset, singular attention

Figure 3: Model performance on the synthetic Uniform dataset. Attention singular.
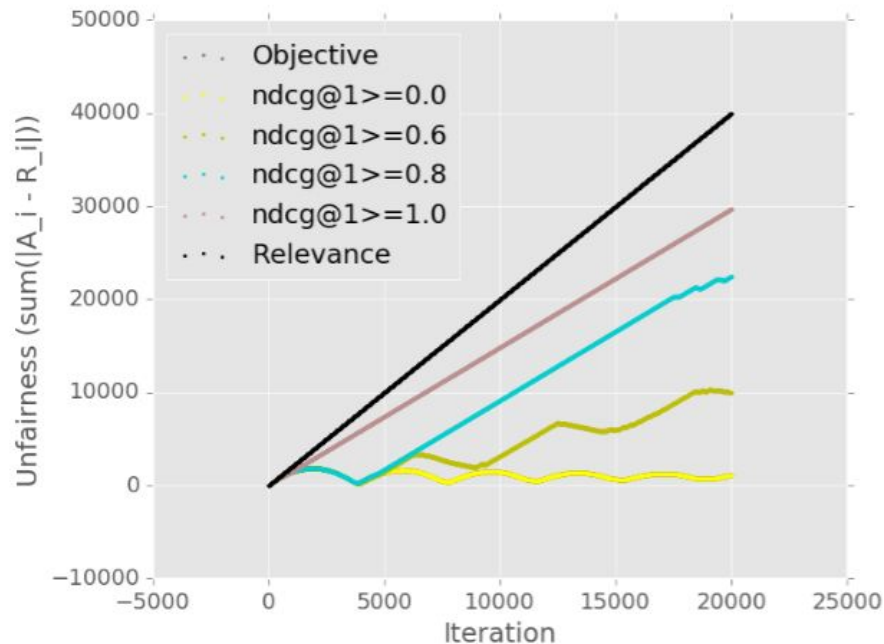
Uniform dataset, geometric attention

Figure 6: Model performance on the synthetic Uniform dataset. Attention geometric.
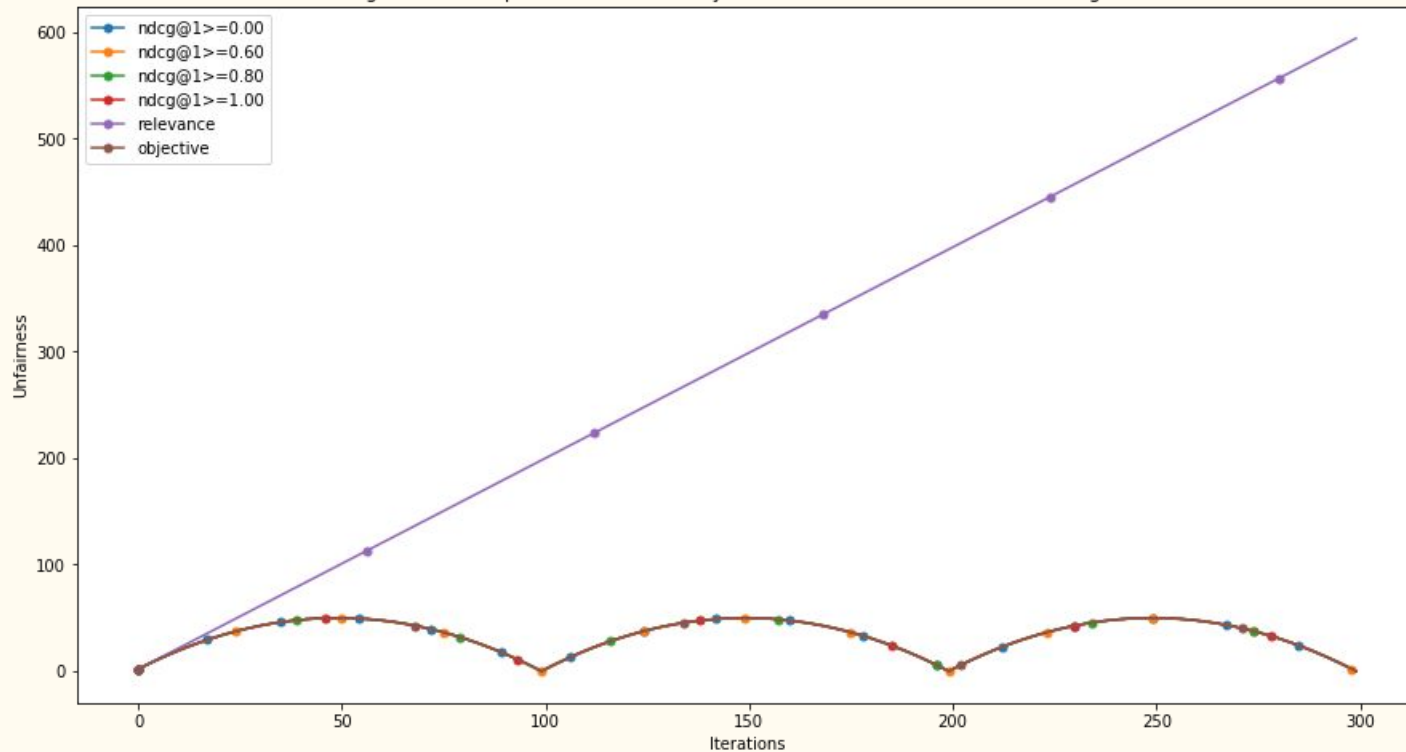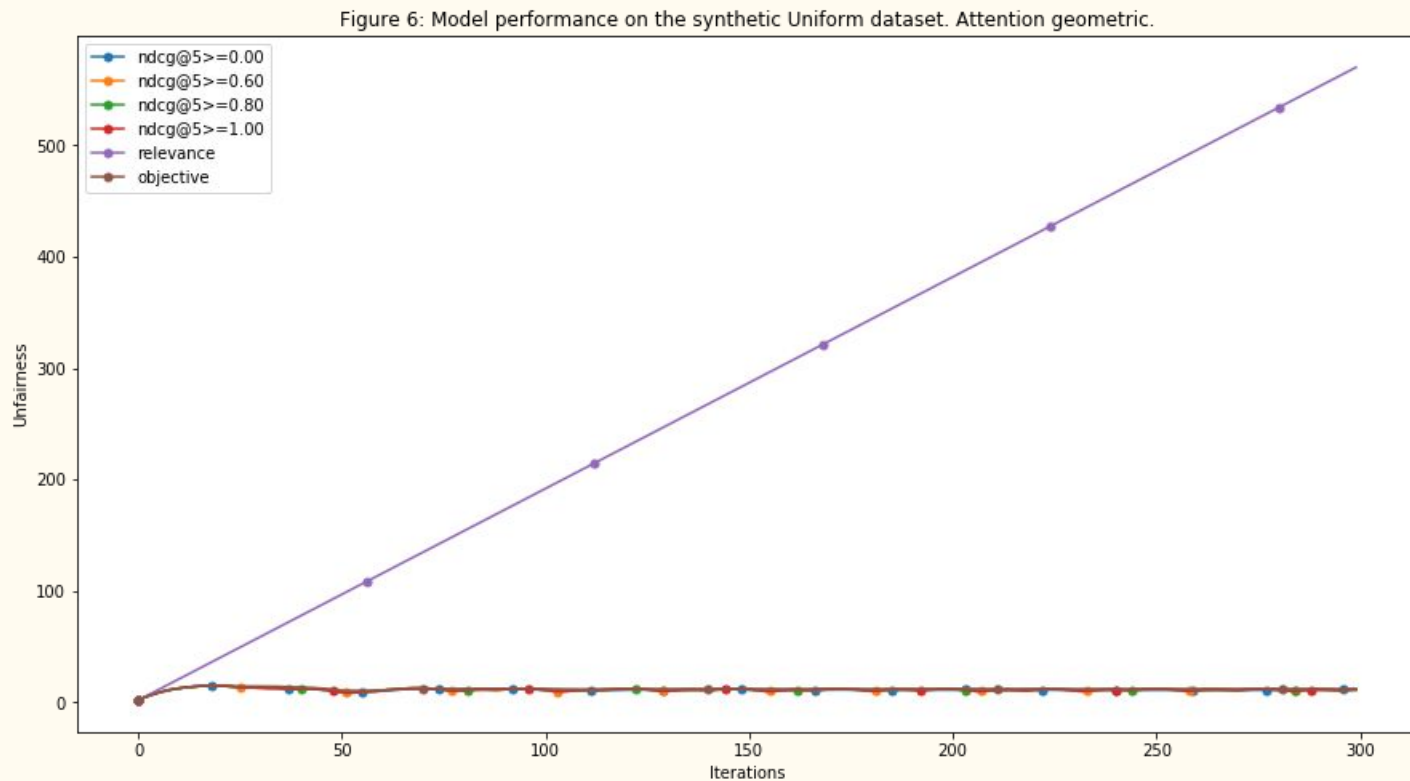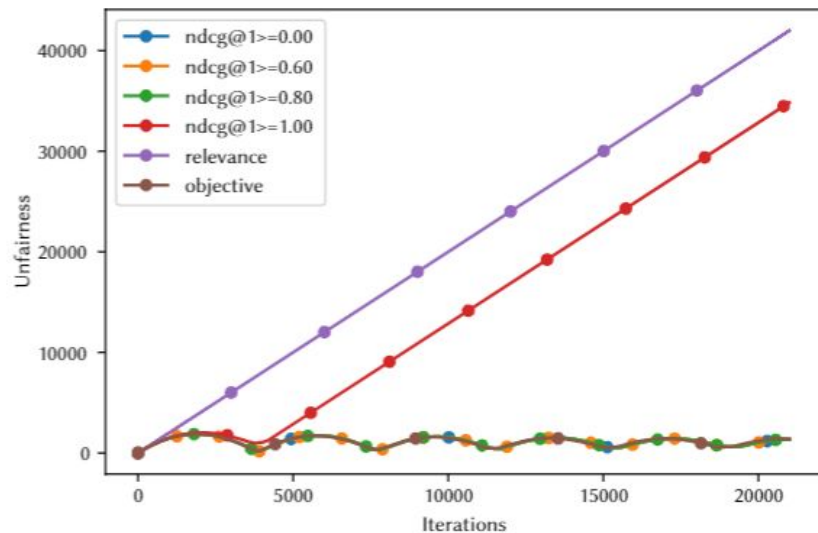
AirBnB dataset

Singular attention
Single-query
Multi-query

- Boston   (3924 subjects)

- Geneva   (2209 subjects)

- Hong Kong   (4513 subjects)

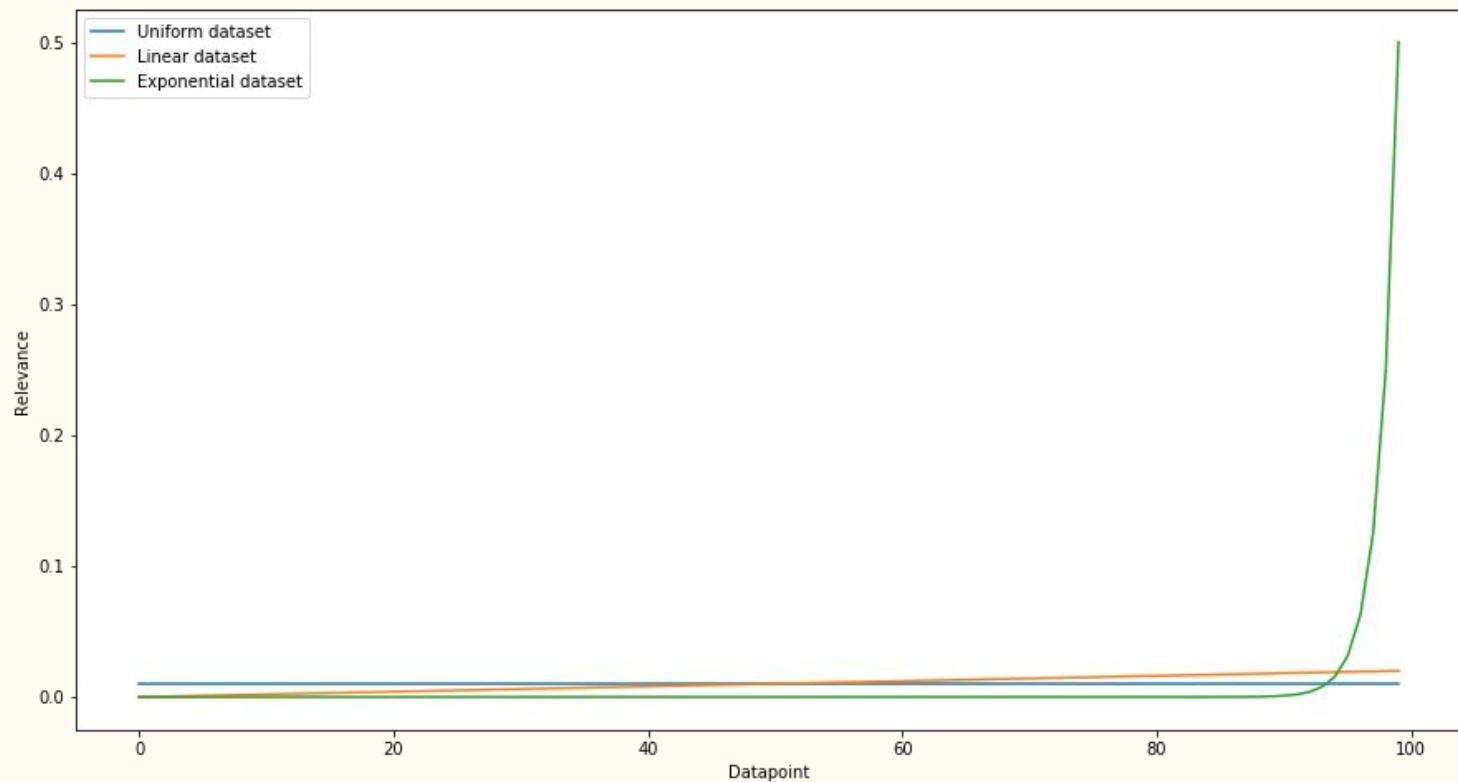# Boston dataset, single-query
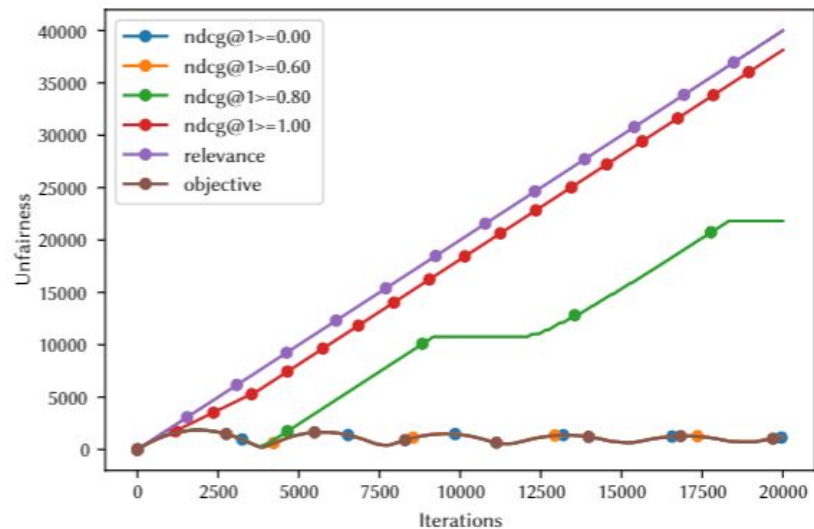
Boston dataset, multi-query
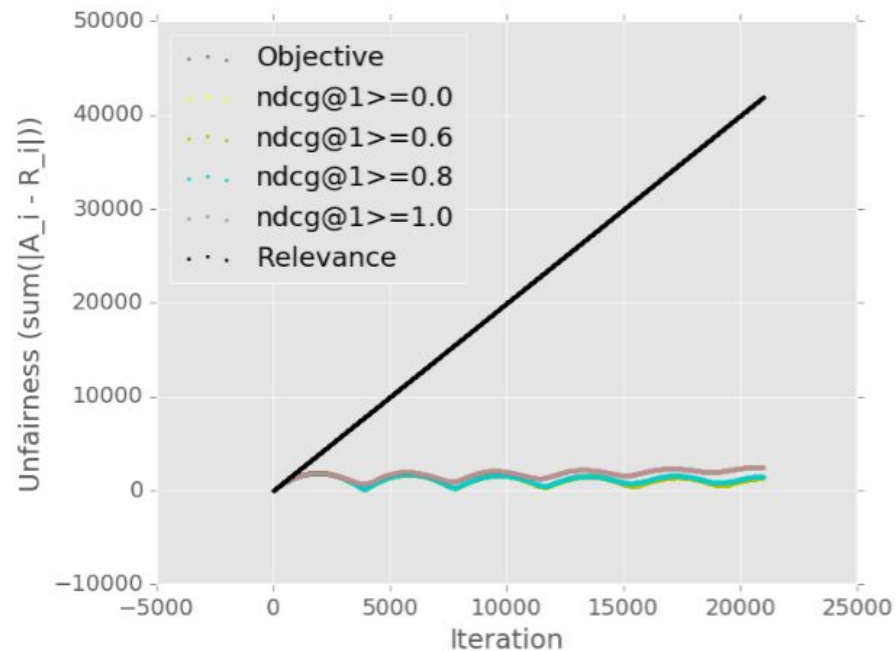
Conclusion

- Synthetic dataset - completely reproduced
- AirBnB dataset - trends reproduced
- Algorithm - successfully implemented
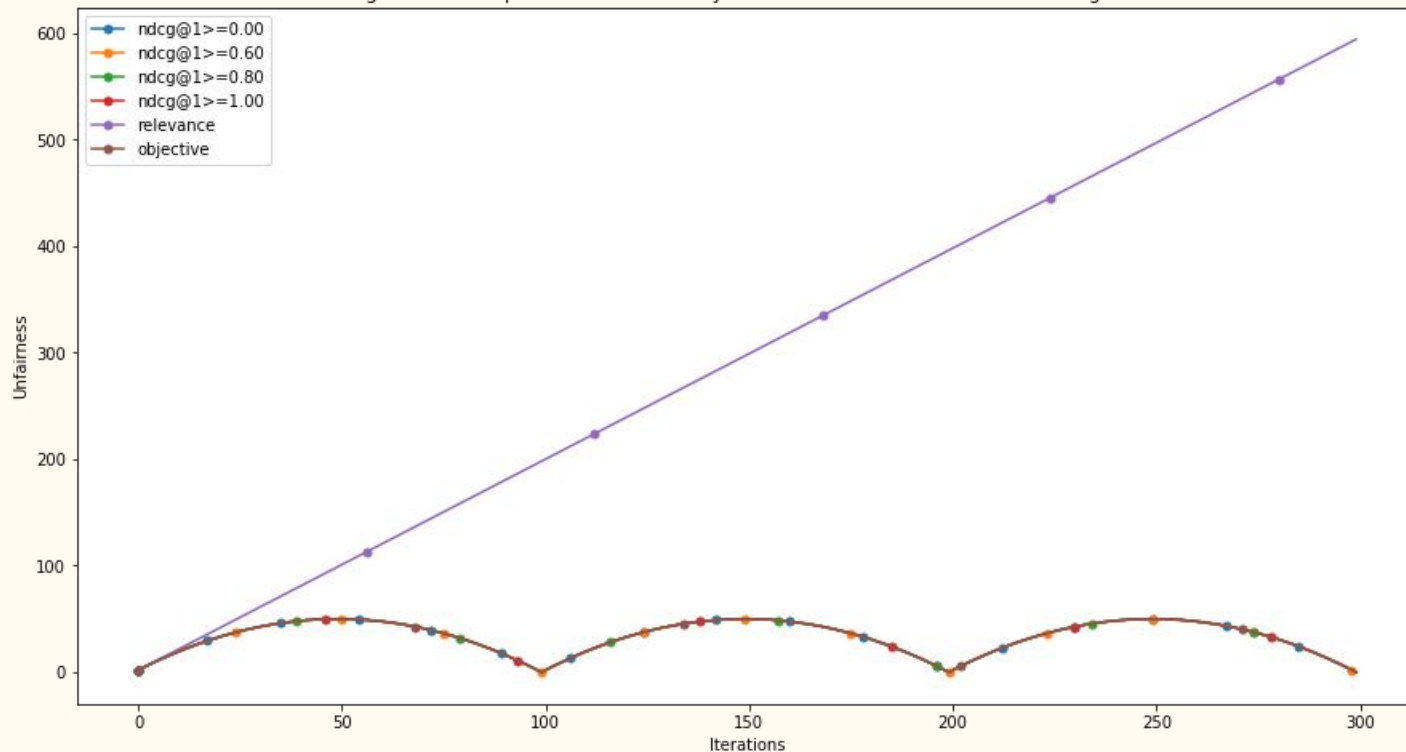- Actual unfairness reduction - under debate
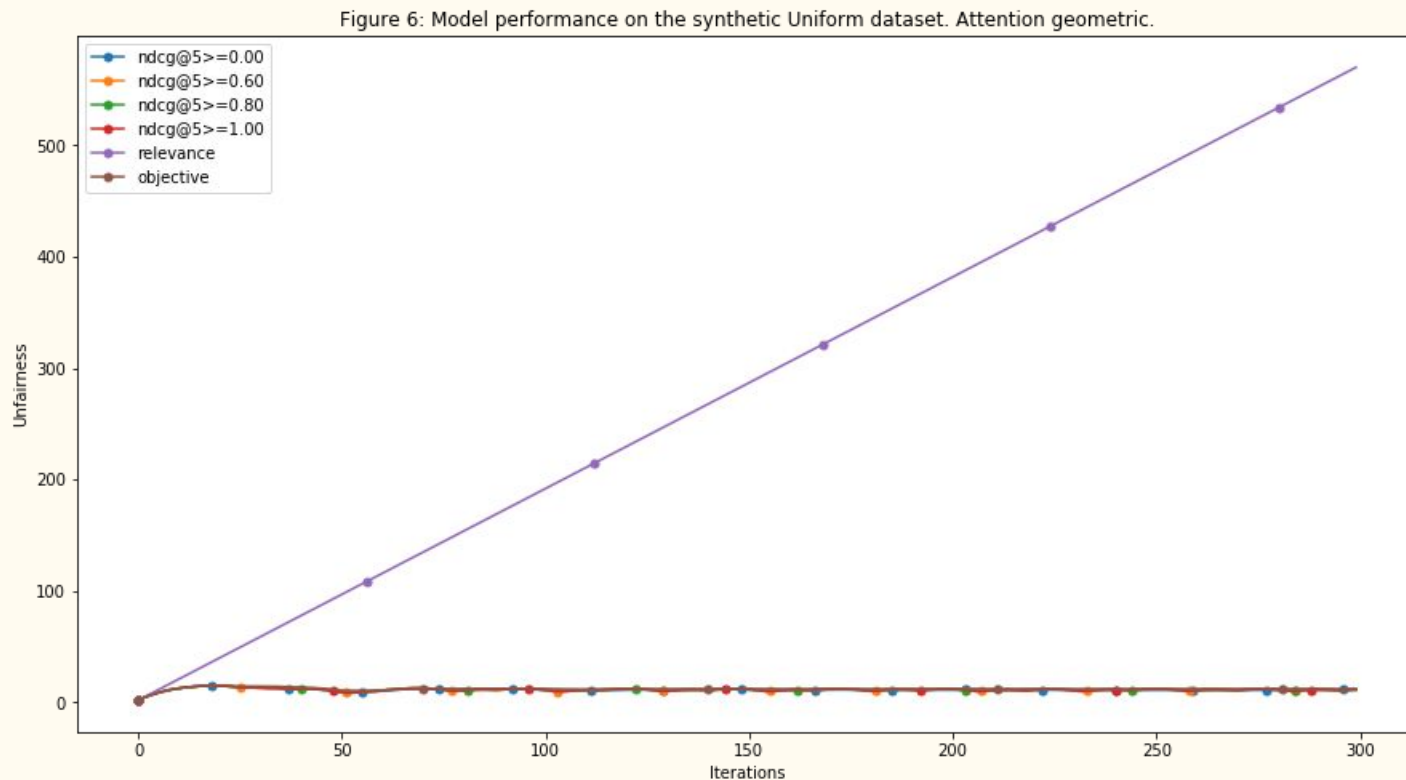
Synthetic dataset

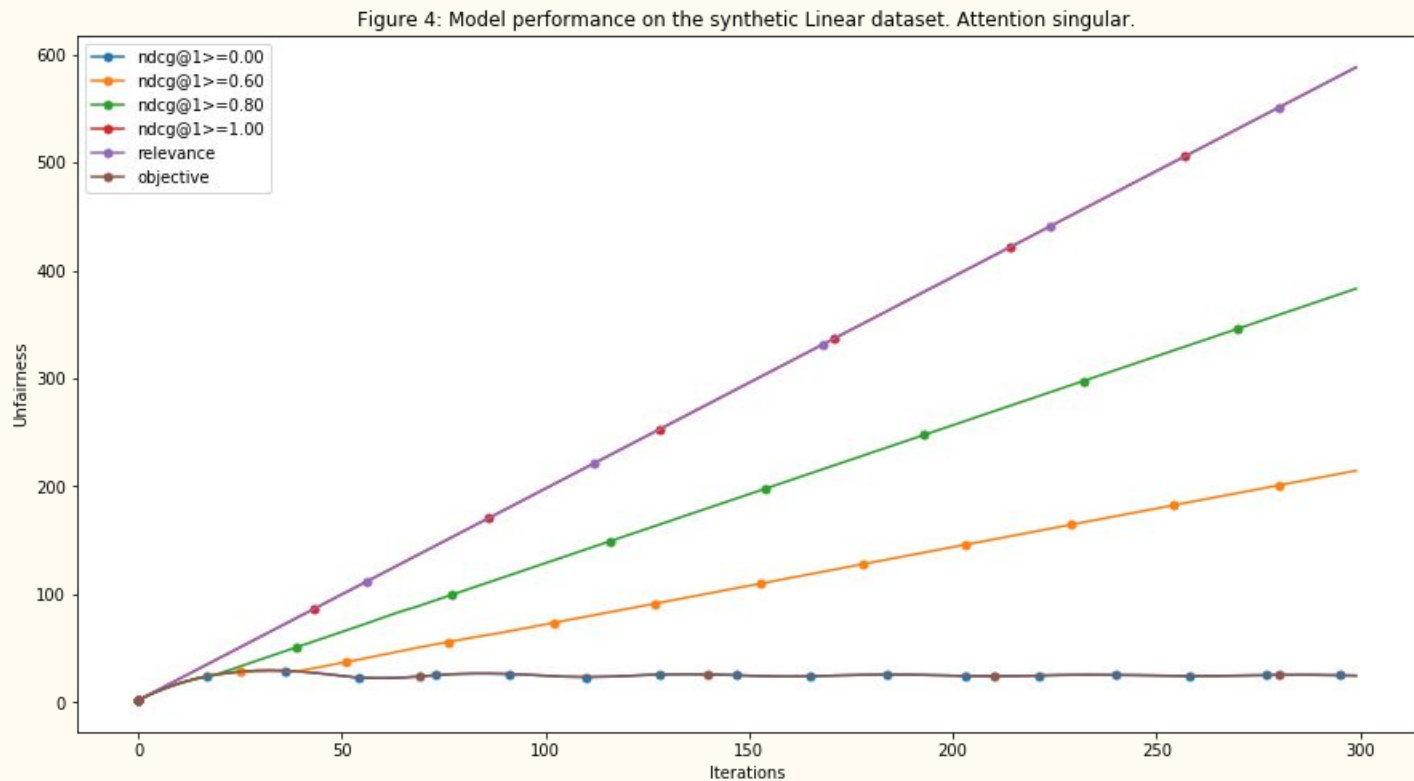# Uniform dataset, singular attention

Figure 3: Model performance on the synthetic Uniform dataset. Attention singular.
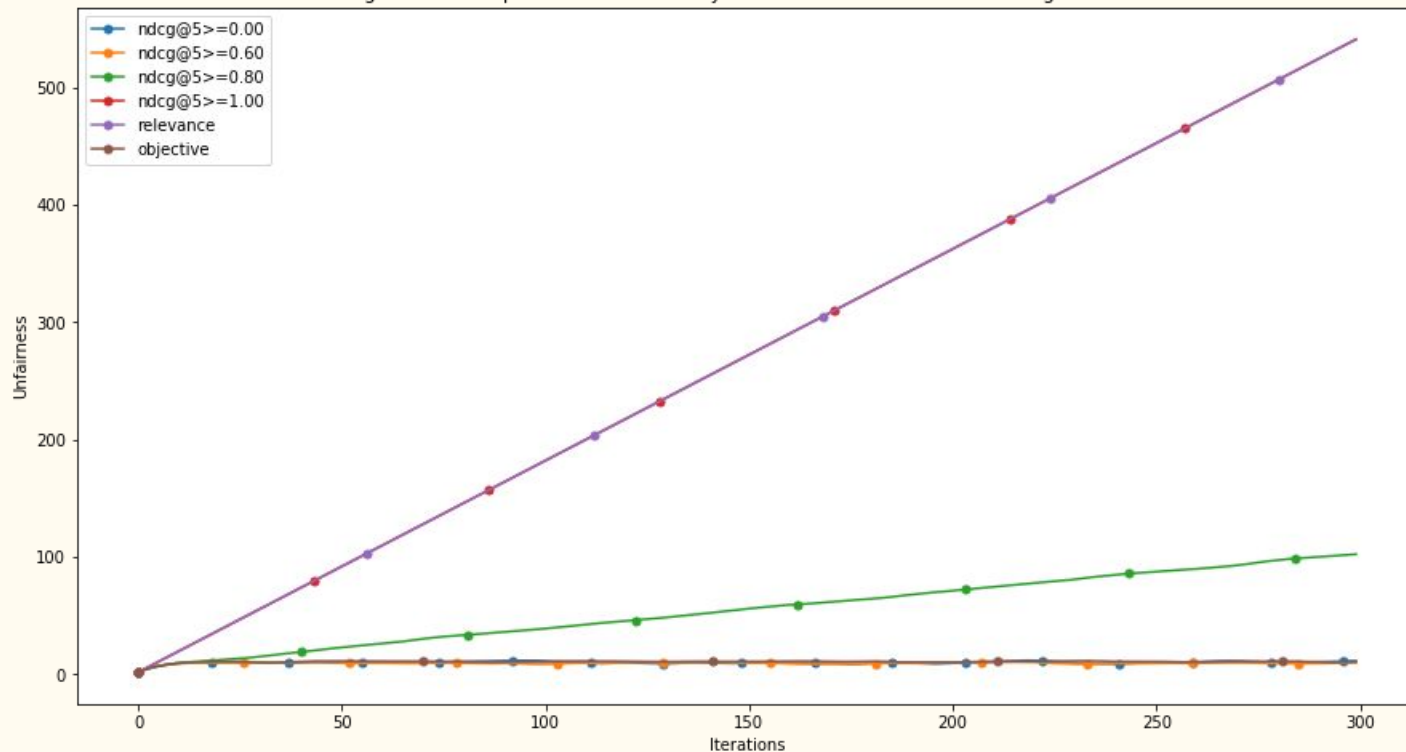
Uniform dataset, geometric attention

Figure 6: Model performance on the synthetic Uniform dataset. Attention geometric.

Linear dataset, singular attention

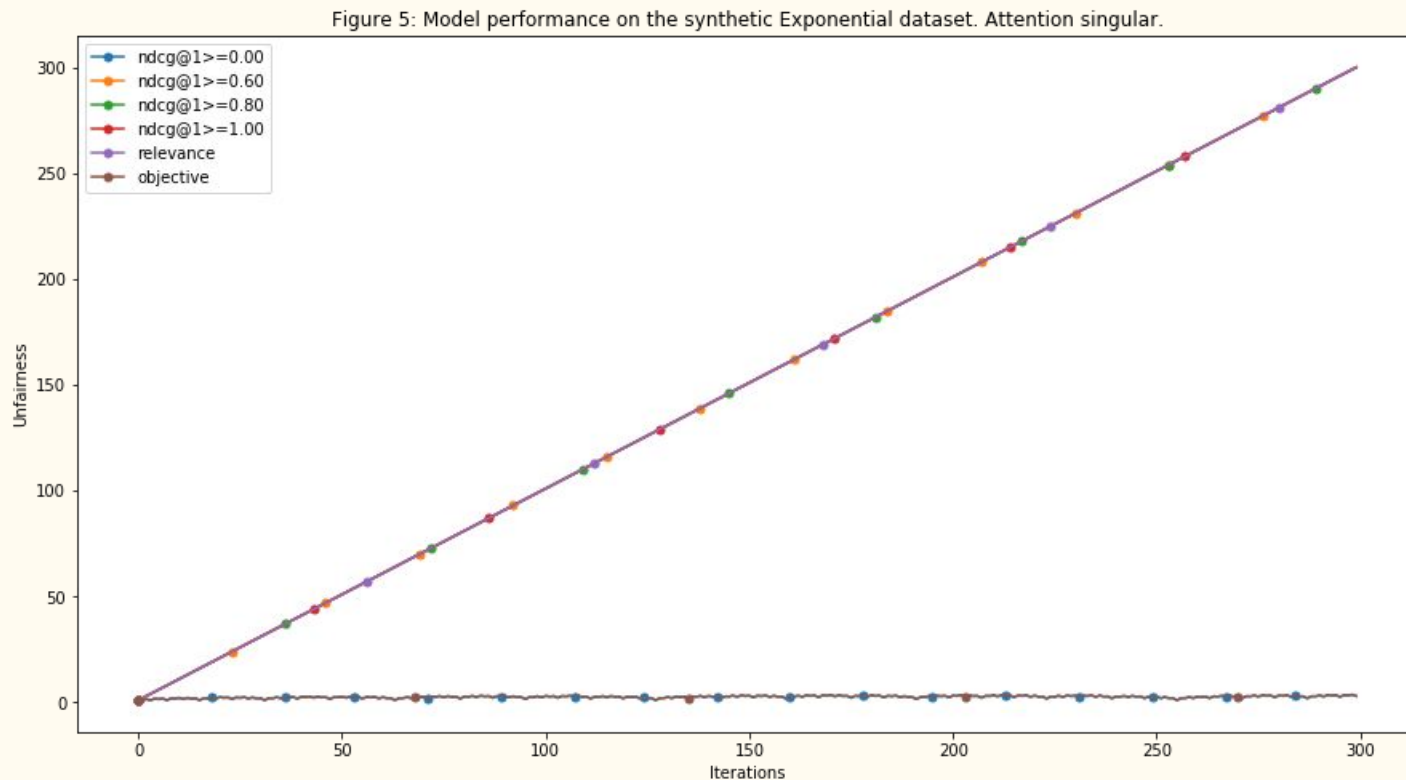Figure 4: Model performance on the synthetic Linear dataset. Attention singular.

Linear dataset, geometric attention

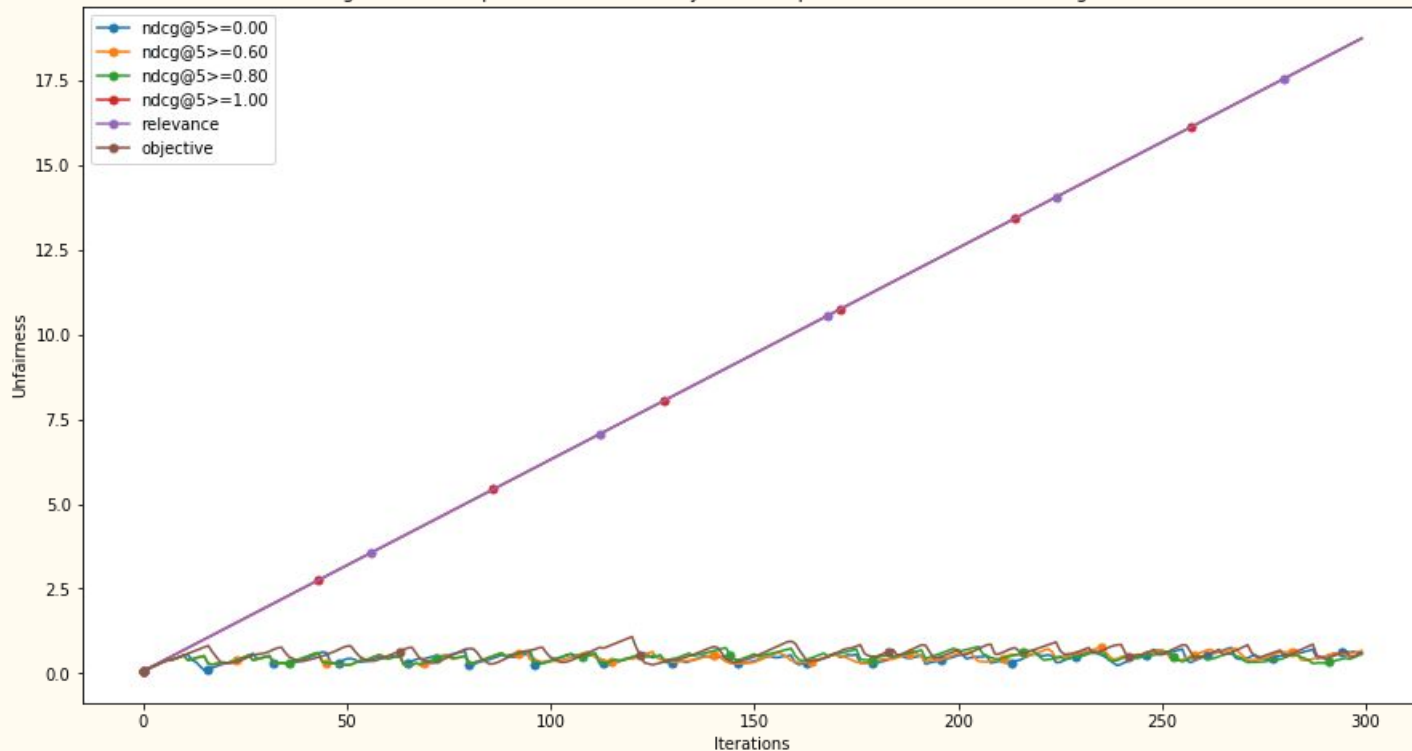Figure 7: Model performance on the synthetic Linear dataset. Attention geometric.
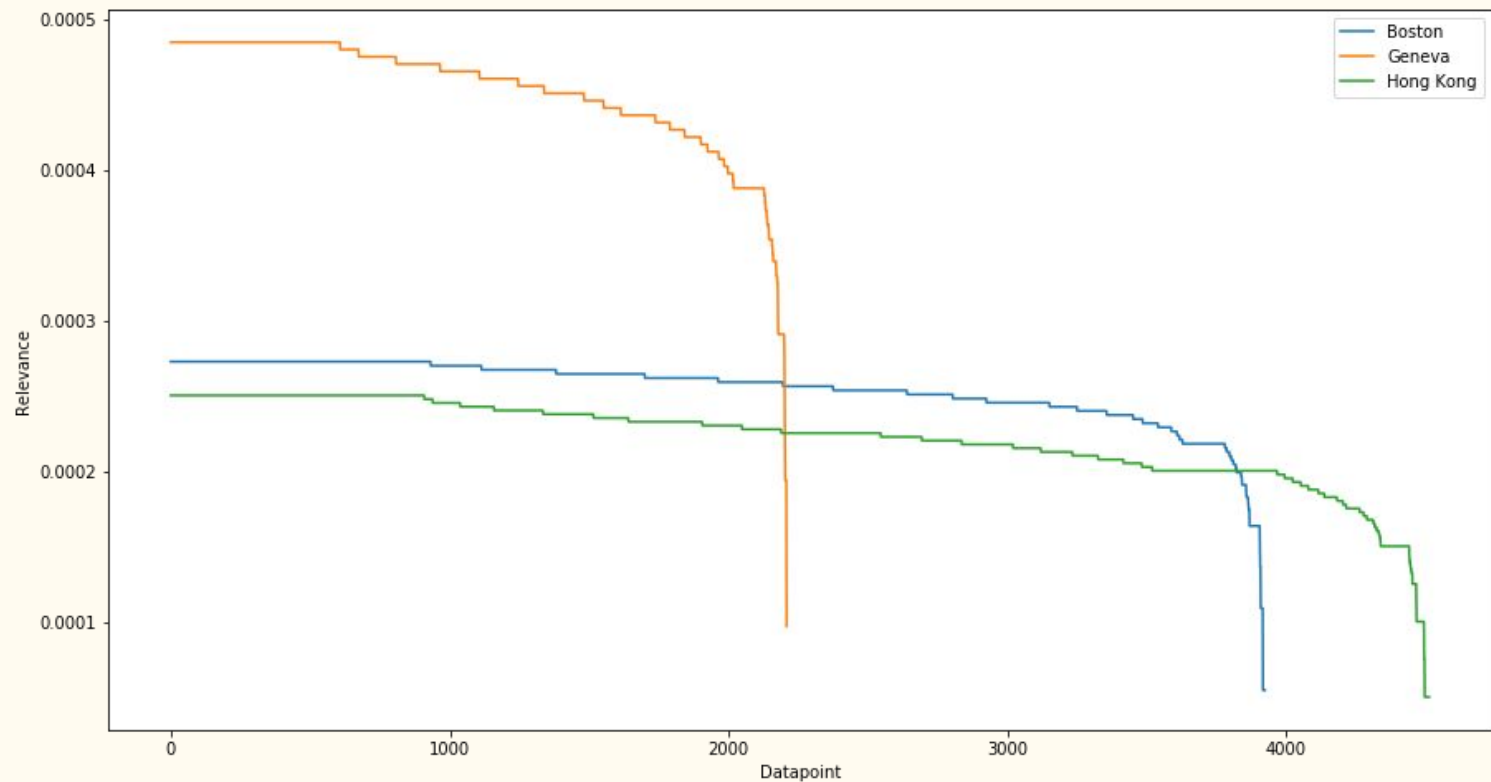
Exponential dataset, singular attention

Figure 5: Model performance on the synthetic Exponential dataset. Attention singular.

Exponential dataset, geometric attention

Figure 8: Model performance on the synthetic Exponential dataset. Attention geometric.
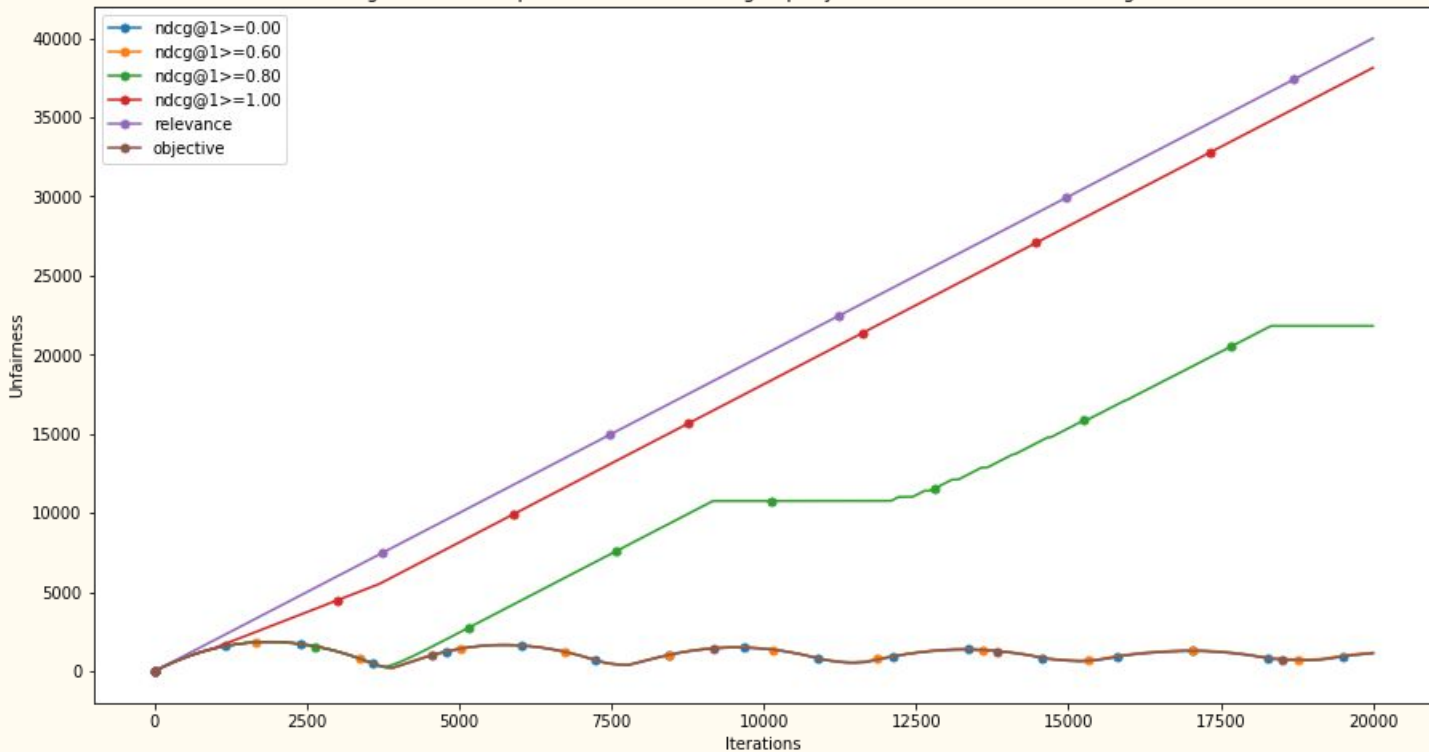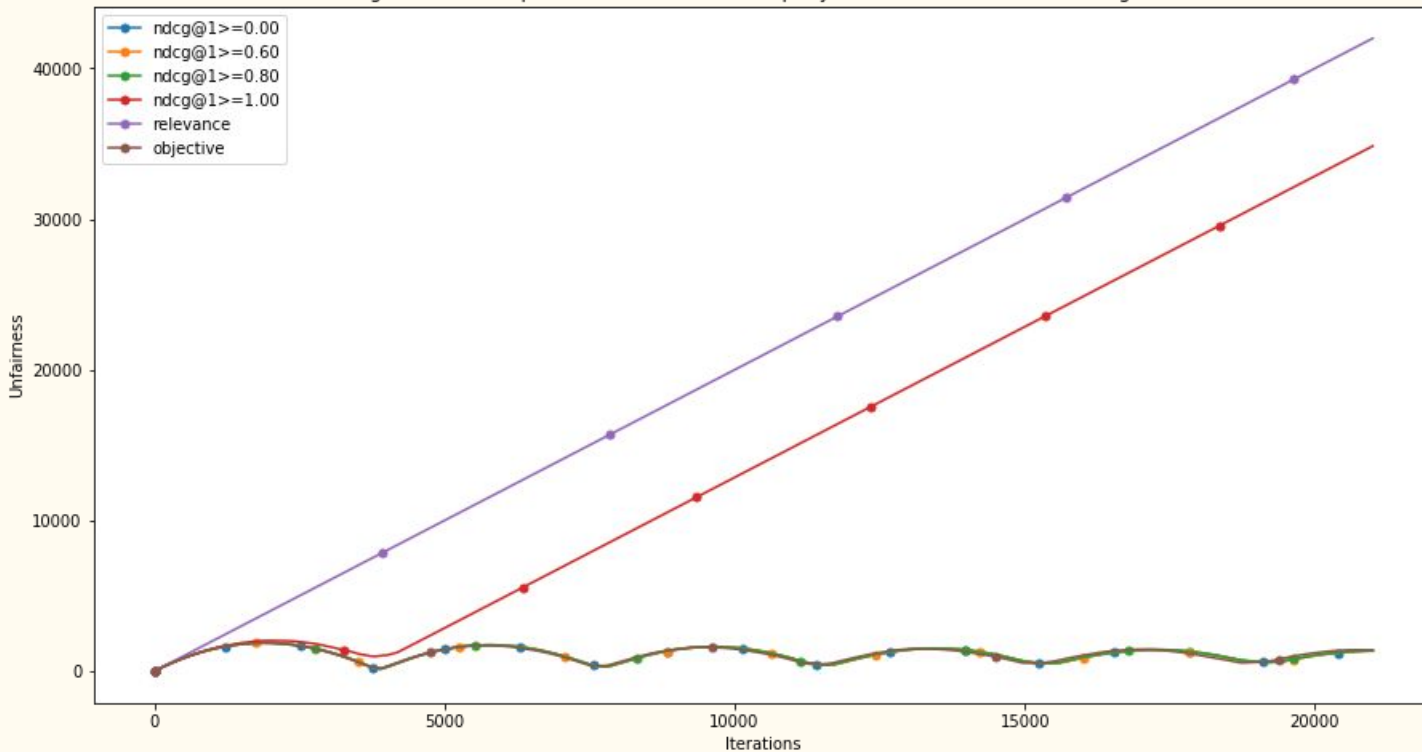
AirBnB Relevances

Boston dataset, single-query

Figure 10: Model performance on the single-query Boston dataset. Attention singular.
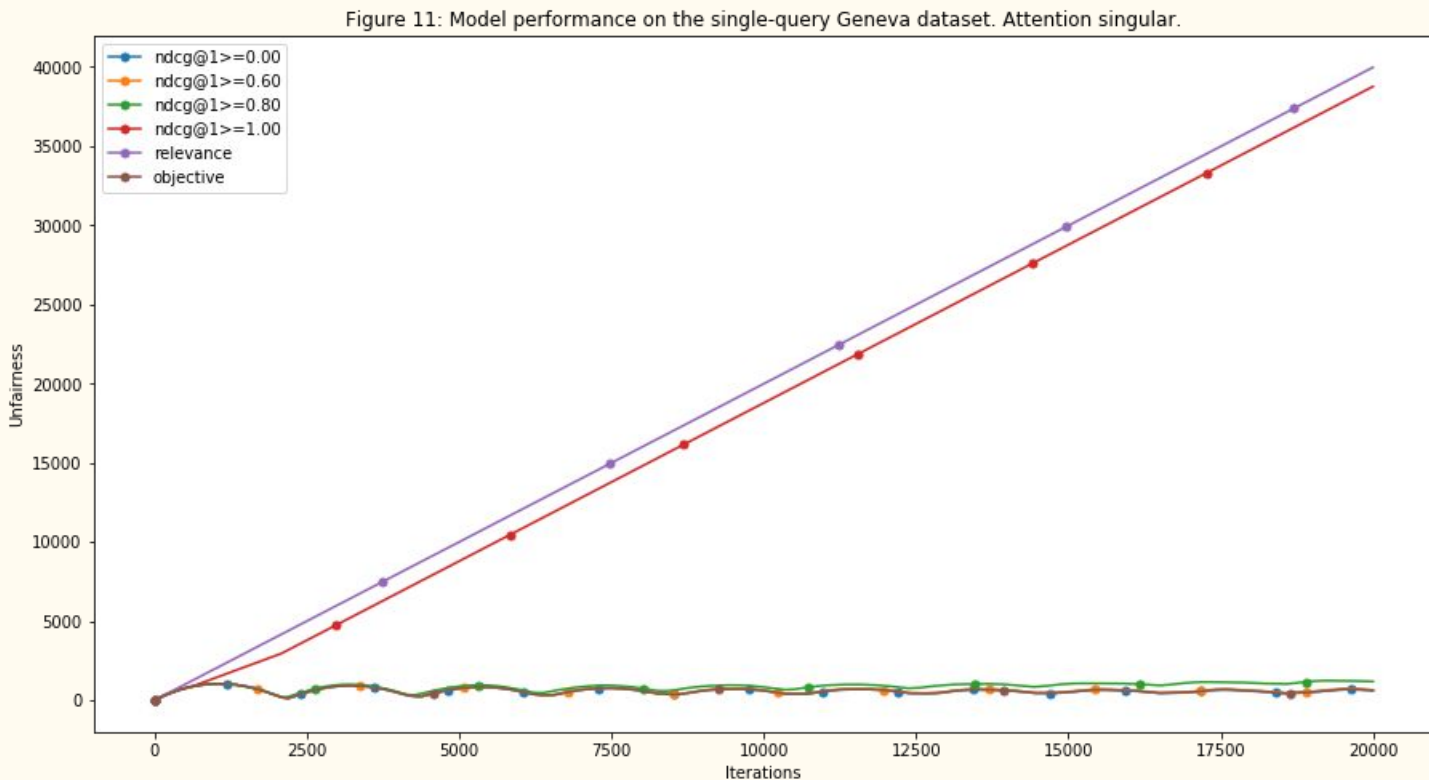
Boston dataset, multi-query

Figure 13: Model performance on the multi-query Boston dataset. Attention singular.
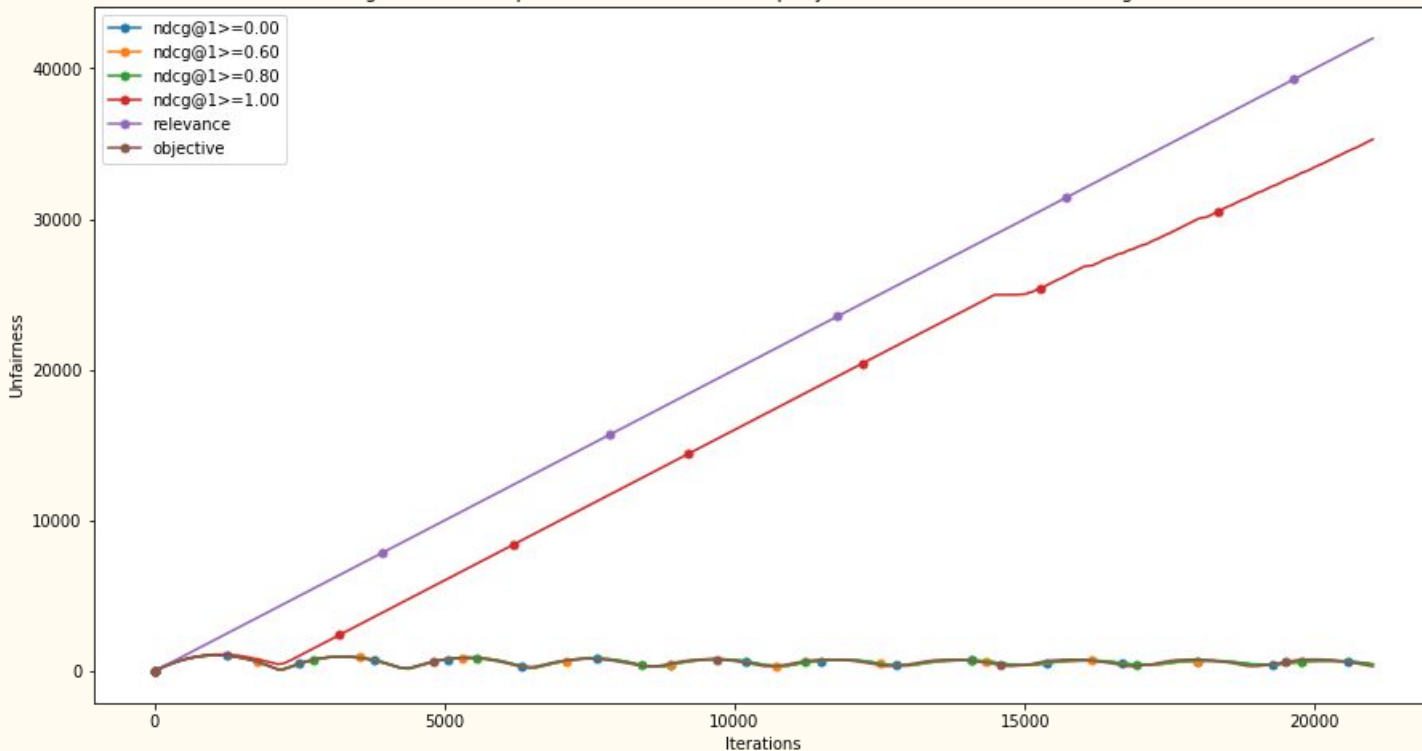
Geneva dataset, single-query

Figure 11: Model performance on the single-query Geneva dataset. Attention singular.
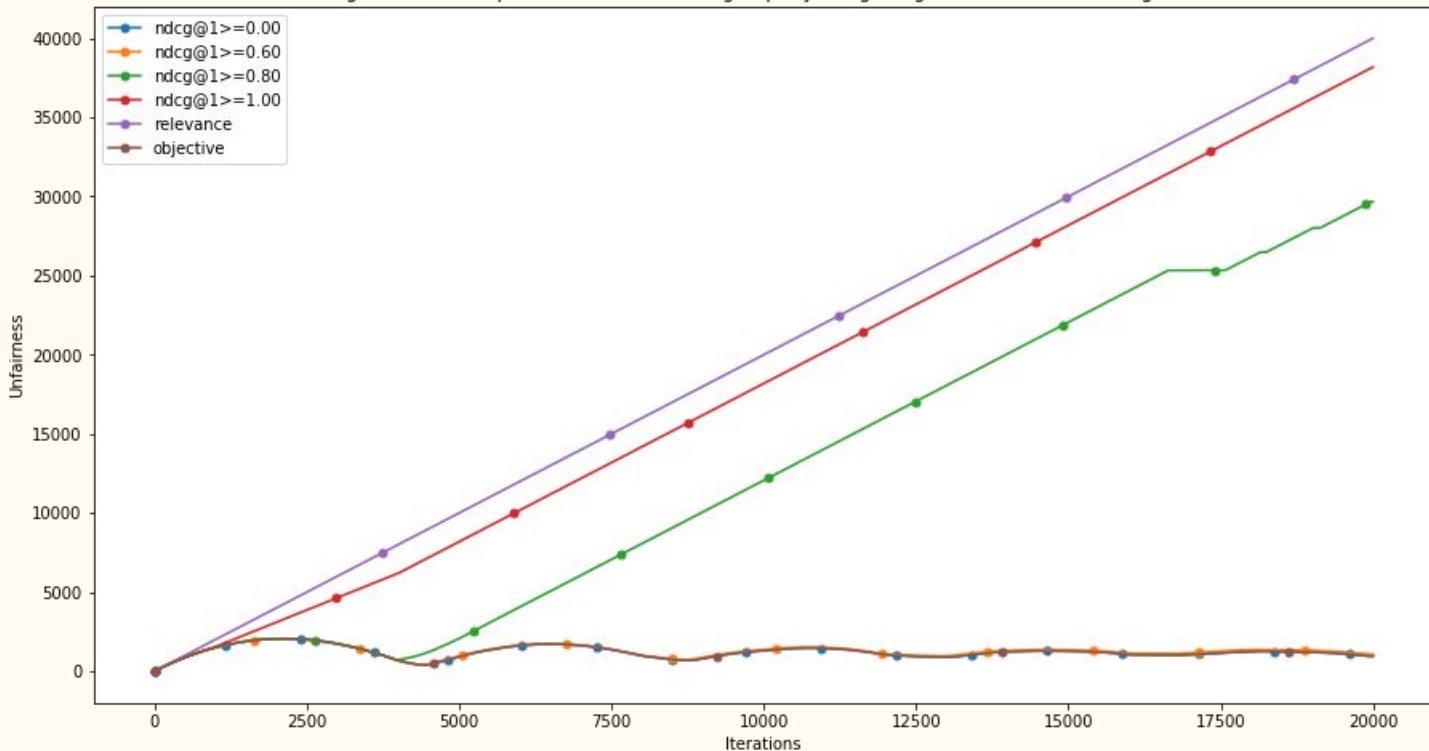
Geneva dataset, multi-query

Figure 14: Model performance on the multi-query Geneva dataset. Attention singular.

Hong Kong dataset, single-query

Figure 12: Model performance on the single-query Hong Kong dataset. Attention singular.

Hong Kong dataset, multi-query

Figure 15: Model performance on the multi-query Hong Kong dataset. Attention singular.