

CS786 Assignment 4

Dataset and Instructions

I am giving you a dataset that contains 70 judgments a subject has made about the size of hypothetical people based on their weight (in kilos) and height (in inches). The subject has categorized people into three categories - small, average, and large.

The dataset `X.csv` contains the 70 actual judgments made by the subject as a 70×3 matrix:

- The first column contains weights.
- The second column contains heights.
- The third column contains the category label assigned by the subject (small = 1, average = 2, large = 3).

I am also giving you a test set `y.csv` of 10 more weight-height combinations as a 10×2 matrix (with the same column interpretations). I want you to tell me what a generalized context model (GCM) would predict this subject's category labels to be, assuming:

1. The subject is polite and is far more likely to call someone *average* than *large*.
2. The subject is more likely to use weight than height to make category judgments about size.

Questions

1. **Q1.** (30 points)

Implement a Generalized Context Model (GCM) encoding these assumptions and provide quantitative predictions for the test set. Submit both the code and category responses for the data points.

2. **Q2.** (40 points)

I am also sharing with you John McDonnell's Python implementation of Anderson's Rational Model of Categorization (`rational.py`). Modify the code to obtain category predictions for the data I have shared with you.

3. **Q3.** (30 points)

For both the GCM and Rational Model of Categorization (RMC), show empirically using the dataset I've shared that both models assume exchangeability of data, i.e., the order in which data enters the model does not affect the category labels of the model for any given subset of data.