NAME: K UGENDER
SR.NO: 21008
ugenderk@iisc.ac.in

Visual Neuroscience
Data Analytics Assignment 4

Prog: M-tech
Dept: CSA
2023-10-06

Task

The first task in the assignment is to compute the *Relative Entropy* distance between various pairs of images in the odd-one-out detection experiment. Additionally, we need to calculate the $L1$ distance and the *Average Search Delay*, using the firing rate data of the neurons. The second task involves finding the ratio of the AM and GM of the spreads of the products of the distance metrics.

1 Implementation

To calculate the average search time for a column "cïn the search time dataset with 24 columns (representing 4 sets, each with 6 groups) and a total of 144 rows (corresponding to 24 persons each shown 6 variations of an oddball-distractor pair), we use the following equation:

$$Avg\_Searchtime_c = \frac{1}{144}\sum_{i=1}^{144}(searchdata_c^i - baseline)$$

In the provided equation, the superscript ï"denotes the search data for the i^th row in the c^th column, represented as c in subscript. The baseline search time for this problem is $328ms$.
For measuring the relative entropy and $L1$ distances using the neuronal firing data, I have utilized the definitions outlined in the lecture slides. The neuronal firing data consists of 30 columns, categorized into 3 sets, each containing 6 groups, and one set with 12 groups. Following the instructions, I calculated a total of 18 relative entropy values for the first three sets, and for the last set, which comprises 12 columns, then applied a compound search strategy to obtain the distance measures.

1.1 Compound Search Strategy

In calculating the relative entropy (or L1 distance) values, let's consider the Bug-Worm pair as an example. For the column where the oddball is 'bug' and the distractor is 'worm,' the observation data can be derived from any of the following four cases:

1 Odd-one Bug - Distractor Worm

2 Odd-one Bug - Distractor Worm flipped

3 Odd-one Bug flipped - Distractor Worm

4 Odd-one Bug flipped - Distractor Worm flipped

All four of these cases have equal probabilities. Therefore, I calculated the relative entropy for each case and computed a simple average to obtain the final entropy value for the specific column. Consequently, for all 12 columns corresponding to the 4 sets, and obtain a total of $12 \times 4 \times 4 = 192$ data points.

1.2 Fitting Gamma Distribution

For fitting a Gamma Distribution, denoted as $\Gamma(\alpha, \beta)$, to the search times and estimating the shape α and rate β parameters of this probability distribution. We need to compare the Cumulative Distribution Function (CDF) of the original distribution with the CDF obtained from a subset of data points in the search time data.

1.2.1 Estimation of parameter

To estimate the shape parameter i have used the slop of the straight line plotted in the previous section, then I randomly selected half of the samples. These selected data points were then utilized to estimate the rate parameter. This estimation was carried out by minimizing the squared error loss of variance, which is a linear function of the mean. This approach was employed because, for the gamma distribution, the mean is given by $Mean = \frac{\alpha}{\beta}$, and the variance is given by $Variance = \frac{\alpha}{\beta^2}$. Consequently, the rate parameter β can be calculated as $\beta = Mean/Variance$.

1.2.2 Kolmogorov-Smirnov Statistic

In statistics, the Kolmogorov–Smirnov test, often referred to as the K-S test or KS test, is a non-parametric test used to quantify the distance between the empirical distribution function of a sample and the cumulative distribution function of a reference distribution, or between the empirical distribution functions of two different samples. I have obtained the Kolmogorov-Smirnov Statistic by comparing the CDF of the empirical gamma distribution and gamma distribution.

2 Plotting

This scatter plot illustrates the relationship between inverse search delay and the relative entropy distance measure. The blue straight line represents the best fit for the data points, achieved by minimizing squared error loss.

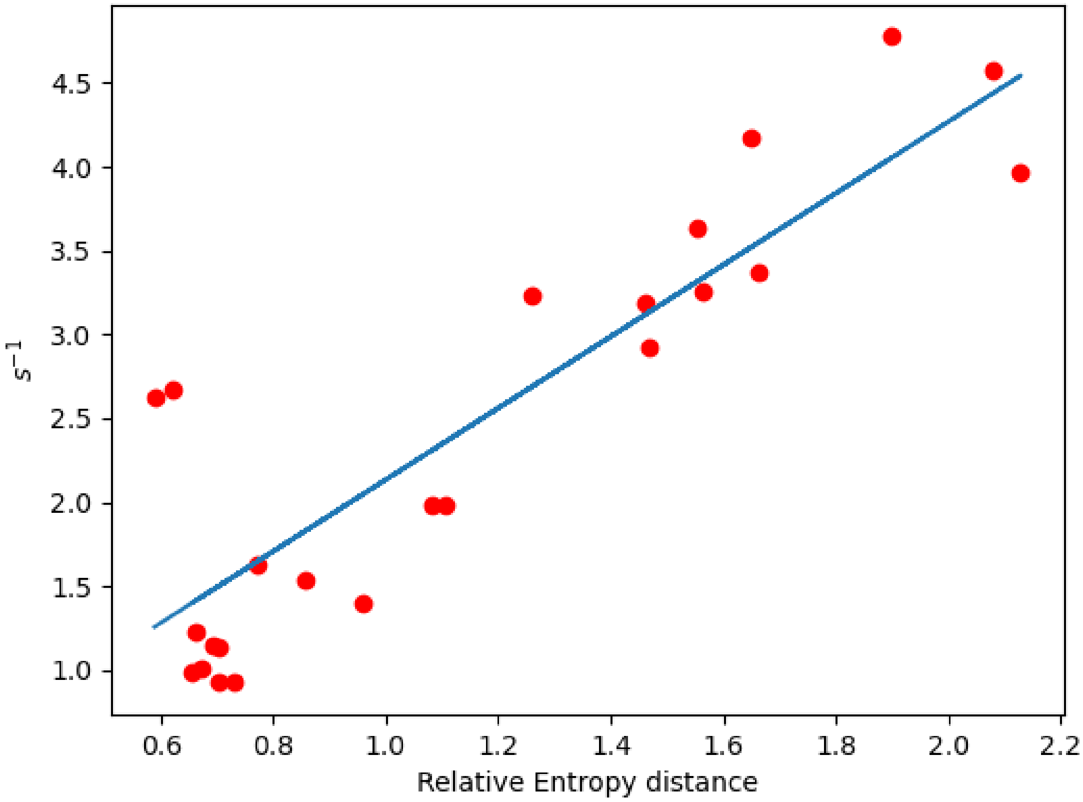

Fig. 1: Entropy

This scatter plot visually displays the correlation between inverse search delay and the L1 distance measure. The blue straight line serves as the optimal fit for the data points, achieved through the minimization of squared error loss.

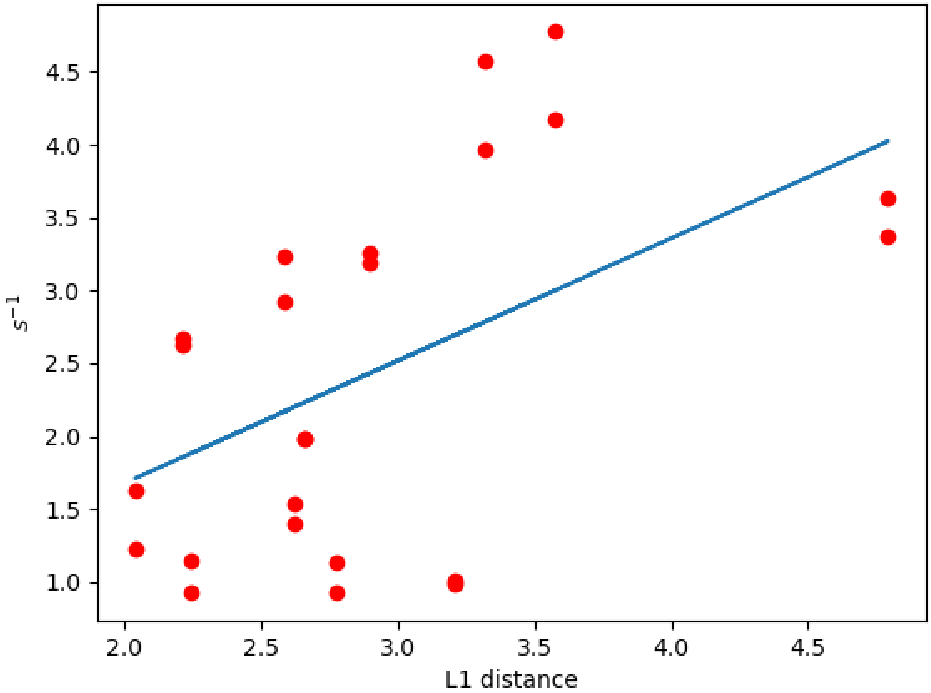

Fig. 2: L1

Here the scatter plot showcases the relationship between the mean and standard deviation of search delays. It reveals a linear trend, where the standard deviation varies almost proportionally with changes in the mean.

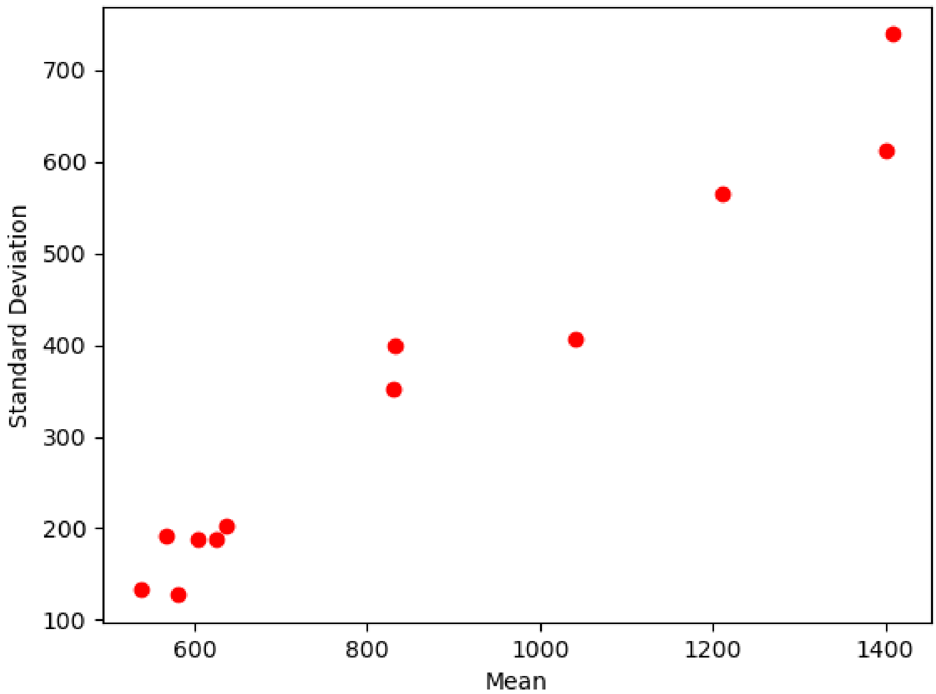

Fig. 3: Mean vs Standard Deviation

The cumulative distribution function (CDF) of the gamma distribution. The blue curve is generated from the remaining data samples, while the red curve represents the CDF of the gamma distribution calculated using the estimated shape and rate parameters.

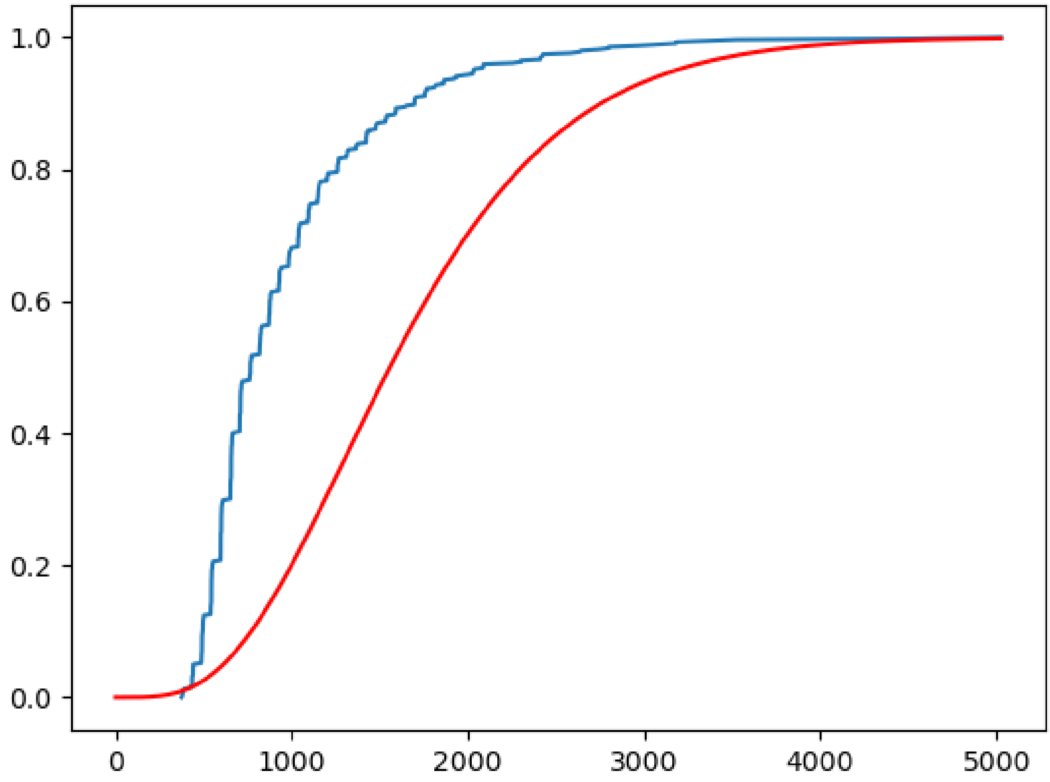

Fig. 4: CDF of Gamma Distribution

3 Results

3.1 Question 1

The AM/GM measure of spread for the products:
search delay $\times$ relative entropy $= 1.042$
search delay $\times$ $L1$ distance1.128
Relative entropy is a better distance metric compared to L1 distance because it measures the spread of data better.

3.2 Question 2

The estimated parameters are as follows:
The estimated value for the shape parameter is 3.222.
The estimated value for Rate parameter is 0.002.

3.3 Question 3

The Kolmogorov-Smirnov statistic obtained is as follows:
$statistic = 0.547$
$P\_value = 1.555 \times e^{-8}$