

How do you calculate the Canberra Distance in Python, and can you provide an example?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How do you calculate the Canberra Distance in Python, and can you provide an example?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=163750>

The Canberra Distance is a measure of similarity between two sets of data, typically used in data analysis and machine learning. It takes into account the differences and similarities between individual data points in the two sets, rather than just the overall difference between them. In Python, the Canberra Distance can be calculated using the scipy library's "canberra" function. This function takes two arrays as inputs and returns the Canberra Distance between them. An example of calculating the Canberra Distance in Python would be:

```
import numpy as np
from scipy.spatial.distance import canberra

# create two arrays of data
A = np.array()
B = np.array()

# calculate the Canberra Distance between A and B
distance = canberra(A, B)

# print the result
print(distance)

# output: 0.5
```

This means that the Canberra Distance between the two arrays is 0.5, indicating a moderate level of similarity between the two sets of data.

Calculate Canberra Distance in Python (With Example)

The Canberra distance between two vectors, A and B, is calculated as:

$$\text{Canberra distance} = \sum |A_i - B_i| / (|A_i| + |B_i|)$$

where:

A_i : The i th value in vector A **B_i : The i th value in vector B**

For example, suppose we have the following two vectors:

$A = B =$

We would calculate the Canberra distance between A and B as:

Canberra Distance = $|2-5|/(2+5) + |4-5|/(4+5) + |4-7|/(4+7) + |6-8|/(6+8)$
Canberra Distance = $3/7 + 1/9 + 3/11 + 2/14$
Canberra Distance = 0.95527

The Canberra distance between these two vectors is 0.95527.

The following example shows how to calculate the Canberra distance between these exact two vectors in Python.

Example: Calculating Canberra Distance in Python

First, let's create a NumPy array to hold each of our vectors:

```
import numpy as np
```

```
#define two arrays
```

```
array1 = np.array()  
array2 = np.array()
```

Next, we can use the `canberra()` function from the SciPy package in Python to calculate the Canberra distance between the two vectors:

```
from scipy.spatial import distance
```

```
#calculate Canberra distance between the arrays  
distance.canberra(array1, array2)
```

```
0.9552669552
```

The Canberra distance between the two vectors is 0.95527.

Notice that this value matches the one we calculated earlier by hand.

Additional Resources

The following tutorials explain how to calculate other common distance metrics in Python: