CS – 513: Knowledge Discovery and Data Mining

MIDTERM FALL 2024

**1. Is the following function a proper distance function? Why? Explain your answer. Measure the distance between (0, 0, 0) and (0, 1, 0).**

Solution:

Yes, the given function is a proper distance formula. It satisfies the properties of metrics.

* Non-Negativity: d (x, y) ≥ 0 for all x and y, and d(x, y) = 0 if and only if x = y.

(xi – yi)4 always greater than or equal to 0. Taking the fourth root of (xi – yi)4 is always greater than or equal to 0. If d (x, y) = 0, that means (xi – yi)4 = 0 for all values of i, which means xi = yi.

* Symmetry: If d (x, y) = d (y, x) is true. Then it is symmetric.

We know that xi – yi and yi – xi only differs by a sign. The root ¼ would eliminate that negative sign too. Then the given function holds good for this condition.

* Triangle Inequality: If d (x, y) ≤ d (x, z) + d (z, y) for all x, y, z. Then it satisfies it.

Let x, y, z be points in Real number domain. Then, the distance between x and y is:

The goal is to show: d(x, y) ≤d(x, z)+d(z, y)

Using ai=xi−zi and bi=zi−yi​, we can express the distance as:

By Minkowski, we have:

≤ +

The first term on the right-hand side is d(x, z) and the second term is d(z, y). Hence:

d(x ,y) ≤ d(x, z)+d(z, y)

Hence proved that the given distance function is proper.

The distance between (0, 0, 0) and (0, 1, 0) is

The distance between [0, 0 ,0] and [0,1, 0] is 1.

2.**An employee of a company is traveling to either England, Italy, or Spain. The employee can travel to only one country. There is a 50% chance the employee will go to England and a 20% chance to Italy. Assume the chances of contracting COVID to be proportional to the prevalence of the disease in each country, given in the table below. For example, the chances of contracting COVID in England is 1200/1,000,000.**

|  |  |
| --- | --- |
|  | **Prevalence** |
|  | Cases |
|  | **Per Million** |
| **England** | 1200 |
| **Italy** | 1500 |
| **Spain** | 1600 |

* **What are the chances that the employee will contract COVID while travelling?**
* **Assume that the employee has traveled to Europe and contracted COVID, what is the probability that he/she traveled to England?**

Solution:

Given: P(E) = 0.5, P(I) = 0.2,

so P(sample set) = P ( E) + P(I) +P(S) = 1; P(S) = 1 - 0.5 - 0.2 = 0.3

P(C/ E) = 1200/1000000 = 0.0012, P(C/ I) = 1500/1000000 = 0.0015,

P(C / S) = 1600/1000000 = 0.0016.

1. Probability of employee getting COVID is,

P(C) = P (C /E). P (E) + P (C / I). P(I) + P(C/S). P(S)

P(C) = (0.0012 \* 0.5) + (0.0015 \* 0.2) + (0.0016 \* 0.3)

P(C) = 0.0006 + 0.0003 + 0.00048 = 0.00138.

The probability of employee getting COVID is 0.00138 or 0.138%.

1. Probability of the employee travel to England given get COVID is,

Apply Bayes,

So, probability of being in England given COVID is 43.478%.