**Knowledge Discovery and Data Mining (CS 513 A)**

Midterm Solution

Prof. Khasha Dehnad

Student : Devila Bakrania

CWID : 10457590

#1 (10 Points)

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

Solution1: Distance function can only be a proper distance function if it follows the following properties:

1. d(x,y) >= 0 & d(x,y) = 0, if and only if x=y
2. d(x,y) = d(y,x)
3. d(x,z) <= d(x,y) + d(y,z)

Now, let us consider the three points in a coordinate system as p(0, 0, 0), q(0, 1, 0), r(0, 1, 1), and s(1, 1, 1)

**Case 1:** Calculating distance between p(0,0,0) and q(0,1,0) using above function

d(p,q) = |(0-0)²+(0-1)² +(0-0)²| =|0+1+0| = 1

**d(p,q) = 1**

**Case 2:** Calculating distance between p(0,0,0) and r(0,1,1) using above function

d(p,r)=|(0-0)²+(0-1)² +(0-1)²| =|0+1+1| = 2

**d(p,r) = 2**

**Case 3:** Calculating distance between p(0,0,0) and s(1,1,1) using above function

d(p,s)=|(0-1)²+(0-1)² +(0-1)²| =|1+1+1| = 3

**d(p,s) = 3**

**Case 4:** Calculating distance between q(0,1,0) and p(0,0,0) using above function

d(q,p)=|(0-0)²+(1-0)² +(0-0)²| =|0+1+0| = 1

**d(q,p) = 1**

**Case 5:** Calculating distance between q(0,1,0) and r(0,1,1) using above function

d(q,r)=|(0-0)²+(1-1)² +(0-1)²| =|0+0+1| = 1

**d(q,r) = 1**

**Case 6:** Calculating distance between q(0,1,0) and s(1,1,1) using above function

d(q,s)=|(0-1)²+(1-1)² +(0-1)²| =|1+0+1| = 2

**d(q,s) = 2**

**Case 7:** Calculating distance between r(0,1,1) and p(0,0,0) using above function

d(r,p)=|(0-0)²+(1-0)² +(1-0)²| =|0+1+1| = 2

**d(r,p) = 2**

**Case 8:** Calculating distance between r(0,1,1) and q(0,1,0) using above function

d(r,q)=|(0-0)²+(1-1)² +(1-0)²| =|0+0+1| = 1

**d(r,q) = 1**

**Case 9:** Calculating distance between r(0,1,1) and s(1,1,1) using above function

d(r,s)=|(0-1)²+(1-1)² +(1-1)²| =|1+0+0| = 1

**d(r,s) = 1**

**Case 10:** Calculating distance between s(1,1,1) and p(0,0,0) using above function

d(s,p)=|(1-0)²+(1-0)² +(1-0)²| =|1+1+1| = 3

**d(s,p) = 3**

**Case 11:** Calculating distance between s(1,1,1) and q(0,1,0) using above function

d(s,q)=|(1-0)²+(1-1)² +(1-0)²| =|1+0+1| = 2

**d(s,q) = 2**

**Case 12:** Calculating distance between s(1,1,1) and r(0,1,1) using above function

d(s,r)=|(1-0)²+(1-1)² +(1-1)²| =|1+0+0| = 1

**d(s,r) = 1**

**Verifying the validity of the distance formula properties on the calculated by above given function**

**Property 1: d(x,y) >= 0 & d(x,y) = 0, if and only if x=y**

Here, d(p,q)>=0, d(p,r)>=0, d(p,s)>=0, d(q,p)>=0, d(q,r)>=0, d(q,s)>=0, d(r,p)>=0, d(r,q)>=0, d(r,s)>=0, d(s,p)>=0, d(s,q)>=0, d(s,r)>=0

**Hence, Property 1 is satisfied for all cases.**

**Property 2: d(x,y) = d(y,x)**

From case 1 and 4, d(p,q) = d(q,p)

From case 2 and 7, d(p,r) = d(r,p)

From case 3 and 10, d(p,s) = d(s,p)

From case 5 and 8, d(q,r) = d(r,q)

From case 6 and 11, d(q,s) = d(s,q)

From case 9 and 12, d(r,s) = d(s,r)

**Hence, Property 2 is satisfied for all cases.**

**Property 3: d(x,z) <= d(x,y) + d(y,z)**

From case 3, case 1, case 5 and case 9,

d(p,s) <= d(p,q)+p(q,r)+p(r,s)

3 <= 1 + 1 + 1 ——-> **eq.1**

Now, from case 10, case 12, case 8 and case 4,

d(s,p) <= d(s,r) + d(r,q) + d(q,p)

3 <= 1+ 1+ 1 ———> **eq.2**

Hence, For both above **eq.1** and **eq.2, Property 3 is satisfied**

From above analysis, Property 1, Property 2 and Property 3 all are satisfied. **Hence, given function is a proper distance function.**

**8**(10 Points)

**The following table shows the population and the actual current prevalence rate of COVID19 in the US, Italy and Spain.**

**Considering only the three countries (US, Italy and Spain) use the table to answer the following questions:**

1. **Estimate the number of cases in the US, Italy and Spain.**
2. **Given that a person is living in the US, what is the probability that the person is infected with COVID19.**
3. **Given that a person is diagnosed with the COVID19, what is the probability that the person lives in the US.**

|  |  |  |
| --- | --- | --- |
|  | **Population** | **Prevalence** |
|  | rounded to nearest | Cases |
|  | **Million** | **Per Million** |
| **US** | 331 | 381.24 |
| **Italy** | 60 | 1463.97 |
| **Spain** | 47 | 1590.24 |

**Solution #8 :**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Population** | **Prevalence** | **Total Cases** |
|  | rounded to nearest | Cases |  |
|  | **Million** | **Per Million** |  |
| **US** | 331 | 381.24 | 126190.44 |
| **Italy** | 60 | 1463.97 | 87838.2 |
| **Spain** | 47 | 1590.24 | 74741.28 |
|  |  |  | Total= 288769.92 |

1. The number of cases in US, Italy and Spain = 126190.44 + 87838.2 + 74741.28 = 288769.92 cases
2. Given that a person is living in the US, the probability that the person is infected with COVID19 = (126190.44/331000) % = 0.0381%
3. Given that a person is diagnosed with the COVID19, the probability that the person lives in the US.=(126190.44/288769.92)% =43.7%