**Topic:**

**Measures of similarity and Dissimilarity:**

1. Definition of similarity and Dissimilarity :
2. Transformation
3. Measures between data objects with simple attribute
4. Measures between data objects with multiple attributes
5. Proximity measures in binary attributes
6. **Example Problem:**
7. **Calculate the measures of similarity and dissimilarity for given binary vectors x and y**
8. **Cosine similarity b) Jaccard c) Eucledian**
9. **X =(1,0,0,1) Y=(0,1,1,0)**
10. **Cosine Similarity:**

**Cosine Similarity Measure for given binary attributes**

**i.e; x= (1,0,0,1) and y=(0,1,1,0)**

**cos (x,y) = ;**

**Now we have to find the values of x.y , │x│,│y│**

**For , x.y = (1\*0+0\*1+0\*1+1\*0) = 0**

**│x│=√1\*1+0\*0+0\*0+1\*1 = √2 = 1.414 (consider only x values)**

**│y│= √0\*0+1\*1+1\*1+0\*0 =√2 = 1.414 (consider only y values)**

**Now substitute these values x.y, │x│,│y│ in the above formula**

**cos (x,y) = ;**

**cos(x,y) = 0/1.414\*1.414 = 0;**

**The measure of cosine similarity of given binary attributes is 0.**

1. **Jaccard :**

Jaccard **Similarity Measure for given binary attributes**

**i.e; x= (1,0,0,1) and y=(0,1,1,0)**

**Jaccard Co-efficient = Number of similar matching attributes**

|  |
| --- |
| **Total Number of attributes** |

**i.e; = f11**

|  |
| --- |
| **f00+f01+f10+f11** |

**f00 = Where x=0 and y=0**

**f01 = Where x=0 and y=1**

**f10 = Where x=1 and y=0**

**f11 = Where x=1 and y=1**

**Based on theses constraints now check the values of f00, f01,f10,f11**

**Now coming to step 1 : Compare the values of x and y which are similar like**

**X(1, 0, 0, 1)**

**Y(0, 1, 1, 0)**

**By the comparison of x and y values none of the values are matched, so now you have to consider the value of f00 is zero.**

**Ie., f00=0**

**Step2: Now find the value of f01and again compare the x and y values, by comparing the values of a x and y only two values are matched, so the value of f01=2**

**Step3: Now find the value of f10 and again compare the x and y values, by comparing the values of a x and y only two values are matched, so the value of f10=2**

**Step4: Now find the value of f11 and again compare the x and y values, By the comparison of x and y values none of the values are matched, so now you have to consider the value of f11 is zero.**

**Step 5: Now values of f00,f01,f10,f11 substitute in above formula**

**Jaccard Co-efficient f11**

|  |
| --- |
| **f00+f01+f10+f11** |

**i.e; = 0**

|  |
| --- |
| **0+2+2+0** |

**i.e; = 0**

|  |
| --- |
|  |
| **4** |

**= 0**

**Jaccard coefficient measure of x and y is 0**

**In the similar way calculate the simple matching coefficient**

**Formula=**

**i.e; = f00+ f11**

|  |
| --- |
| **f00+f01+f10+f11** |

1. **Eucledian Distance**

**Dissimilarity Measure for given binary attributes**

**i.e; x= (1,0,0,1) and y=(0,1,1,0)**

**x= ( 1, 0, 0, 1)**

**x1 y1 z1 p1**

**y= ( 0, 1, 1, 0)**

**x2 y2 z2 p2**

**Eucledian Distance**

**=**

**=**

**= 2**