**CS612 - Lab 5**

1. For two binary data:
   * 1. Hamming distance = number of different bits between the two vectors
     2. Jaccard Similarity = J = (number of matching presences) / (number of attributes not involved in the 00 matches
     3. The Simple Matching Coefficient (SMC) distance is: Hamming distance / number of bits.

Compute the Hamming distance, the Jaccard similarity, and SMC between the following three binary vectors.

**X1** = 1111010001

**Y1** = 0100011000

**Hamming Distance= 5, Jaccard Similarity= 2/7=0.28 SMC=0.5**

**X2** = 1101110001

**Y2** = 0110011100

**Hamming Dinstance = 7, Jaccard Similarity = 2/9=0.22 , SMC=0.3**

**X3** = 1111110001

**Y3** = 1110011100

**Hamming Dinstance=5, Jaccard Similarity = 4/9=0.44 , SMC=0.5**

1. For the following vectors, x and y, calculate the indicated similarity or distance measures.
   * x = (1, 1, 1, 1), y = (2, 2, 2, 2) **cosine=1, correlation=undefine, Euclidean=2**
   * x = (0, 1, 0, 1), y = (1, 0, 1, 0) **cosine=0, correlation= -1, Euclidean=2, Jaccard=0**
   * x = (0,−1, 0, 1), y = (1, 0,−1, 0) **cosine=0, correlation=0, Euclidean=2**
   * x = (1, 1, 0, 1, 0, 1), y = (1, 1, 1, 0, 0, 1) **cosine=0.75, correlation=0.251, Jaccard=0.6**
   * x = (2,−1, 0, 2, 0,−3), y = (−1, 1,−1, 0, 0,−1) **cosine=0 and correlation=0**
2. Suppose you have a graph with 5 dimensions as shown below.

1

2

2

3

1

4

4

3

6

5

8

7

P5

P2

P3

P4

P1

**Hamming Distance:**

P1(1,1) P2(3,3) P3(5,2) P4(6,1) P5(8,4)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **P1** | **P2** | **P3** | **P4** | **P5** |
| **P1** | **0** | **4** | **5** | **5** | **10** |
| **P2** | **4** | **0** | **3** | **5** | **6** |
| **P3** | **5** | **3** | **0** | **2** | **5** |
| **P4** | **5** | **5** | **2** | **0** | **5** |
| **P5** | **10** | **6** | **5** | **5** | **0** |

**Euclidian Distance:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **P1** | **P2** | **P3** | **P4** | **P5** |
| **P1** | **0** | **2.8** | **4.1** | **5** | **7.6** |
| **P2** | **2.8** | **0** | **2.2** | **3.6** | **5.1** |
| **P3** | **4.1** | **2.2** | **0** | **1.4** | **3.6** |
| **P4** | **5** | **3.6** | **1.4** | **0** | **3.6** |
| **P5** | **7.6** | **5.1** | **3.6** | **3.6** | **0** |

P1(1,1) P2(3,3) P3(5,2) P4(6,1) P5(8,4)

Supremum distance

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **P1** | **P2** | **P3** | **P4** | **P5** |
| **P1** | **0** | **2** | **4** | **5** | **7** |
| **P2** | **2** | **0** | **2** | **3** | **5** |
| **P3** | **4** | **2** | **0** | **1** | **3** |
| **P4** | **5** | **3** | **1** | **0** | **2** |
| **P5** | **7** | **5** | **3** | **2** | **0** |

Find out the “Hamming Distance”, “Euclidean Distance” and “Supremum Distance” between any two points

**What to submit:**

* **A file that includes the result of Q1, Q2, and Q3. Call the file it Lab5**
* **Place Lab4 file in a folder and call the folder based on your name and your partner’s name (ex:Jack-and-Nancy-Lab5)**
* **Zip the folder and only one person in the team is required to submit the work.**