**CS 59000/STAT 46700 – Topics in Data Science**

**Homework 1-Solution**

1. Given two points **a**=(6,5,4) and **b**=(3,2,1) calculate the following measures between **a** and **b**
2. Euclidean Distance
3. Manhattan Distance
4. Minkowski distance (choose order of the norm p=3)

*Solution: We have*

1. *the Euclidean distance given by*
2. *The Manhattan distance given by*
3. *the Minkowski distance of norm 3 given by*
4. Consider a dataset 2, 4, 6, 8, 8, 10, 12, 14. Express the data in standardized form using
5. Minmax approach in (0,1)
6. Z-score

*Solution: For the given data*

1. Calculate the entropy of a weighted six-sided dice such that three sides of the dice have 1/6 chances of facing, two sides of the dice have a 1/12 chance of facing up, and one side has a 1/3 chance of facing up.

*Solution: We have the following probability structure*

|  |  |
| --- | --- |
| *Outcome* | *Probability* |
| *1* | *1/6* |
| *2* | *1/6* |
| *3* | *1/6* |
| *4* | *1/12* |
| *5* | *1/12* |
| *6* | *1/3* |

*Therefore,*

1. For the dataset given below, find the approximate entropy H (Passed). This data describes whether students pass or not (Y for yes or N for no), based on their past CGPA scores (H for high, A for average, and L for Low) and whether they prepared or not (Y or N).

|  |  |  |
| --- | --- | --- |
| CGPA | Prepared | Passed |
| H | N | Y |
| H | Y | Y |
| A | N | N |
| A | Y | Y |
| L | N | N |
| L | Y | Y |

*Solution:*