Distance Measures

1. Compute the Euclidean Distance between the following points: $X = \{1, 5, 4, 3\}$ and $Y = \{2, 1, 8, 7\}$

x_j	y_j	$x_j - y_j$	$(x_j - y_j)^2$
1	2	-1	1
5	1	4	16
4	8	-4	16
3	7	-4	16
			49

The Euclidean Distance between the two points is $\sqrt{49}$ i.e. 7.

2. Compute the Manhattan Distance between the following points: $X = \{1, 3, 4, 2\}$ and $Y = \{5, 2, 5, 2\}$.

	Case X	Case Y	Difference	|Diff|
Variable 1	1	5	-4	4
Variable 2	3	2	1	1
Variable 3	4	5	-1	1
Variable 4	2	2	0	0
				6

- The Manhattan Distance between the two points is 6.

3. Write a brief description of the Cosine Similarity. Explain how it is related to the Euclidean distance. Illustrate your answer with a sketch.

4. Write a brief description of the Mahalanobis Distance. Illustrate your answer with a sketch.

5. Compute the following distance metrics between the cases A, and B, described below. .

   (i) Euclidean Distance

   (ii) Squared Euclidean Distance

   (iii) Manhattan Distance

   (iv) Chebyshev Distance

$$A = \{5, 8, 2, 1, 14\}$$
$$B = \{3, 2, 9, 4, 7\}$$

   (v) Explain why the squared Euclidean distance may be used in preferences in to the Euclidean Distance.

6. Why do you standardize variables before carrying out a cluster analysis.

7. Explain why using the standardized value may not be suitable in some cases? Give another example of numeric transformation.