DATA REDUCTION I

- The idea of data reduction is to obtain the compressed representation of the original data set.
- The compressed representation can be obtained by:

1. Data Selection:

- Extracting only few meaningful features from the given data set.
- The objective of data selection techniques are to remove redundant and irrelevant features from X such that new features selected improves the end result/objective function.

2. Data Extraction:

- Discovering a new subset of features from the original feature space.
- The data extraction aims at discovering new set of m (where $m < n$) features X′ from X such that X′ captures essence of all features in X and yet improves the end result/objective function.

Suppose we are given n dimensional data set with feature set represented by X as following :

$$X = \{ x_1, x_2, x_3, \ldots\ldots\ldots\ldots, x_n \}$$

$$X = \{ x_1, x_2, x_3, x_4, x_5 \} \xrightarrow{\text{Conversion}} Y = \{ y_1, y_2, y_3, y_4, y_5 \}$$

$$x_2, x_4, x_5 \qquad\qquad y_2, y_4, y_5$$