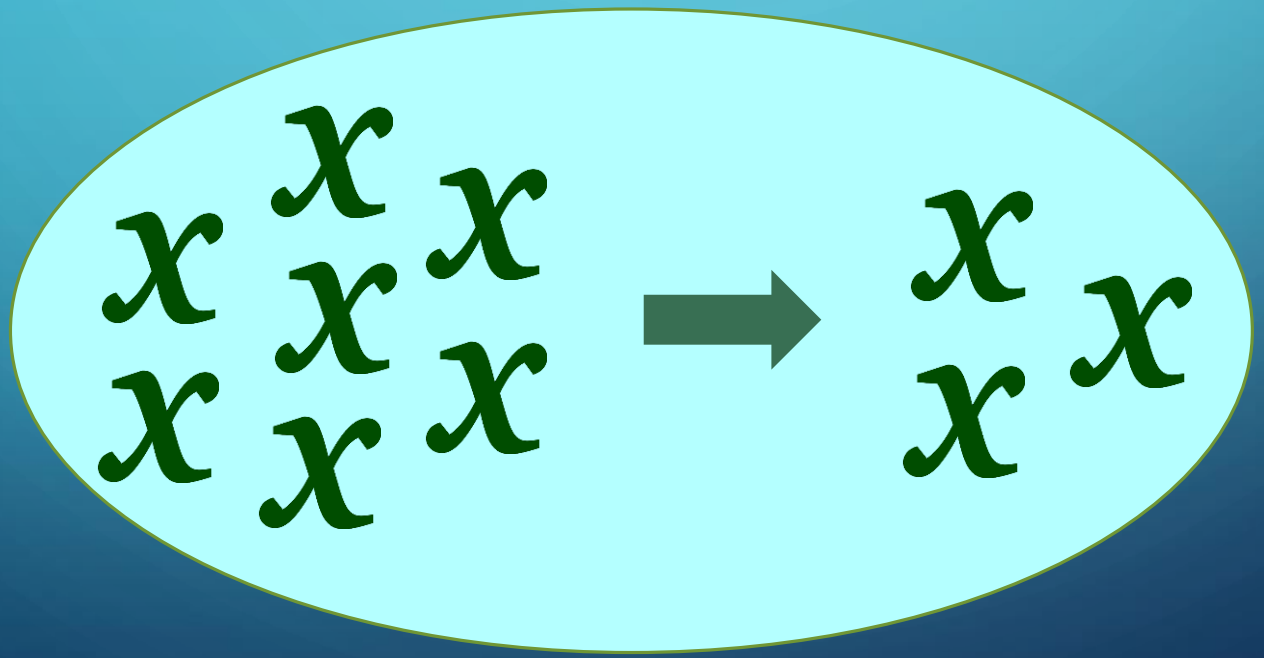
FEATURE SELECTION

DEFINITION

FEATURE SELECTION: DEFINITION

Feature selection is the process of selecting a subset of relevant features (variables, predictors) for use in machine learning model building.



WHY SHOULD WE SELECT FEATURES?

- Simple models are easier to interpret
- Shorter training times
- Enhanced generalisation by reducing overfitting
- Easier to implement by software developers
- Reduced risk of data errors during model use
- Variable redundancy
- Bad learning behaviour in high dimensional spaces

FEATURE SELECTION: PROCEDURE

A feature selection algorithm can be seen as the combination of a search technique for proposing new feature subsets, along with an evaluation measure which scores the different feature subsets.

- Computationally expensive
- Different feature subsets render optimal performance for different machine learning algorithms
  - **Different methods of feature selection**

FEATURE SELECTION: METHODS

- Filter methods
- Wrapper methods
- Embedded methods