Filter based methods

* SelectKBest: It selects k highest scoring features based on a function and removes the rest of the features. Lets take an example of California Housing dataset.

import numpy as np

from sklearn.datasets import fetch\_california\_housing

…………

dowload data and select a subset X\_california[:2000, :], y\_california[:2000]

Calculate Shape of feature matrix before feature selection

Select 3 most important features, since it is a regression problem, we can use only mutual\_info\_regression or f\_regression scoring functions only. Calculate shape of feature matrix after feature selection

* SelectPercentile: This is very similar to SelectKBest from previous section, the only difference is, it selects top

Percentile of all features and drops the rest of the features.

Similar to SelectKBest, it also uses a scoring function to decide importance of features.Let's use the california housing price dataset also for this API and calculate top 30 percentile of features.

* Similarly apply GenericUnivariateSelect also for same number of features on the same dataset.