Dimension Reduction Lab Sheet

1. In **scikit learn**, apply *filter-based feature selection* with Information Gain to identify the 3 most discriminating and 3 least discriminating features in the *wine.csv* dataset.

   Based on these results, assess the 10-fold cross-validation classification accuracy of a 1-Nearest Neighbour classifier with:

   (i) only the 3 <u>most</u> discriminating features included

   (ii) only the 3 <u>least</u> discriminating features included

2. Using **mlxtend** identify informative feature subsets by applying *wrapper-based feature selection* to the *wine* dataset using a 3-Nearest Neighbour classifier and the following search strategies:

   (i) forward sequential search
   (ii) backward elimination search

   Which common features were selected by both search strategies?