

COMP47460 Tutorial

Dimension Reduction

1. In Weka, apply *filter-based feature selection* with Information Gain to identify the 3 most discriminating and 3 least discriminating features in the *Wine* dataset in the ARFF file provided.

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

- (i) only the 3 most discriminating features included
- (ii) only the 3 least discriminating features included

2. In Weka, apply *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?

Would it be appropriate to use either of the resulting feature subsets in conjunction with a Decision Tree classifier? Justify your answer.

3. Use Weka to apply *PCA feature transformation* to the *Diabetes* dataset in the ARFF file provided. (Note: Use the Ranker search method in the 'Select Attributes' tab).

How many Principal Components are selected by Weka?

Produce a visualisation of the 2 leading Principal Components (PCs).