# Assignment 6 - Dimension Reduction

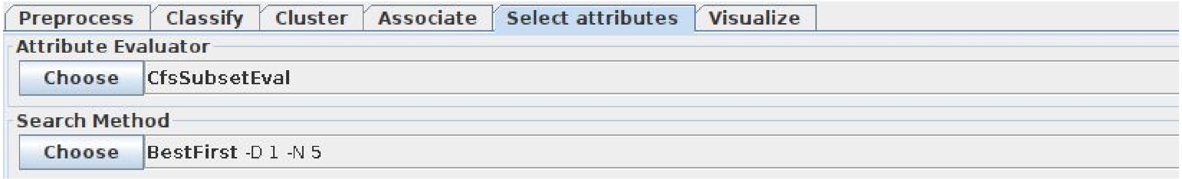
Assignment was solved using *Weka* 3.8.4.

## Feature selection

### Configuration

Classifier evaluator was set to CfsSubsetEval. From *Weka* docs:

*"Evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them."*



Attribute Evaluator
CfsSubsetEval

Search Method
BestFirst -D 1 -N 5

*Fig 1. Attribute Evaluator is set to CfsSubsetEval. Search Method is set to Best Method*

### Output

```
Selected attributes: 5,7,12,17 : 4
                     odor
                     gill-spacing
                     stalk-surface-above-ring
                     veil-color
```

The most discriminative features are odor, gill-spacing, stalk-surface-above-ring and veil-color.

## Plots

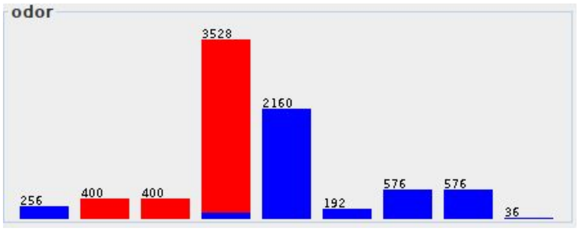


*Fig 2. Odor*

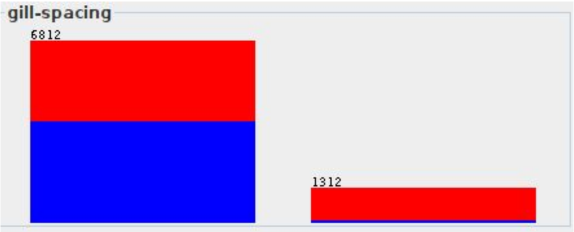


*Fig 3. Gill-spacing*

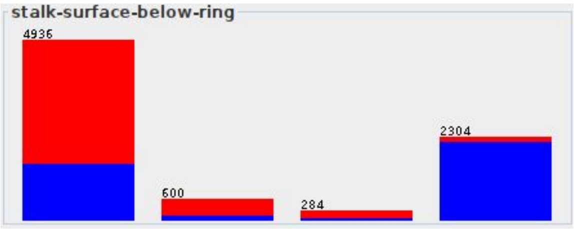


*Fig 4. Stalk-surface-above-ring*

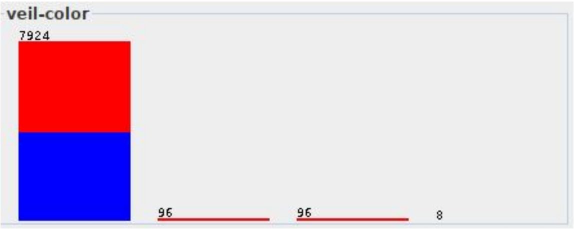
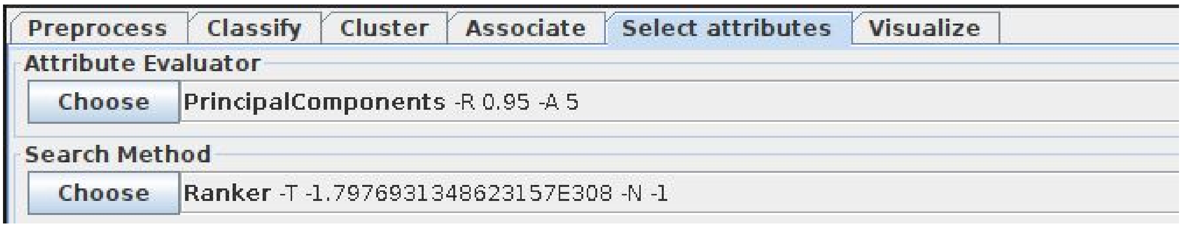


*Fig 5. Veil-color*

We can see from the plots that the feature selection is good.

# Principal Components Analysis

## Configuration



Attribute Evaluator
PrincipalComponents -R 0.95 -A 5

Search Method
Ranker -T -1.7976931348623157E308 -N -1

*Fig 6. Attribute Evaluator is set to PrincipalComponents. Search Method is set to Ranker*

## Output

### Transform back to original = true

```
Ranked attributes:
 1   110 habitat=l
 1    35 gill-color=n
 1    37 gill-color=p
 1    38 gill-color=w
 1    39 gill-color=h
 1    36 gill-color=g
 1    34 gill-color=k
 1    41 gill-color=e
 1    33 gill-size
 1    30 odor=m
 1    31 gill-attachment
 1    32 gill-spacing
```

### Transform back to original = false

```
Ranked attributes:
 0.9106     1
0.257stalk-surface-below-ring=k+0.256stalk-surface-above-ring=k+0.234rin
g-type=l+0.231odor=f-0.215ring-type=p...
```

The above combination of features explain the most variance in the dataset.

# PCA features vs feature selection

There are some overlaps. Both odor and gill-spacing occur in both.