

CSE 454 – DATA MINING
ASSIGNMENT #04

GÖKHAN HAS
161044067

ASSIGNMENT REPORT

LECTURER : DR. BURCU YILMAZ

1 -) Which technique has given better results in terms of f1 score? (filter feature selection or wrapper feature selection) Was it expected?

The filter methodology uses the selected metric to identify irrelevant attributes and also filter out unnecessary columns in your models. It provides the option to isolate selected dimensions that enrich your model. The columns are sorted based on the calculation of feature scores. By choosing and applying the right features, you can potentially increase the accuracy and efficiency of your classification models. **However, in this application, no model is used in a natural method.**

The Wrapper methodology considers the selection of feature sets as a search problem, where different combinations are prepared, evaluated and compared to other combinations. A predictive model is used to evaluate a combination of features and assign model performance scores. **Since a model is called in it, the result should be expected to be better than the Filter method.**

Filter methods are said to be much faster as they are not trained in wrapper methods. It can also be said that wrapper methods are computationally expensive. While filter methods cannot find the best subset of properties in most cases, wrapper methods try to find the best subset in all cases. Using this feature of the wrapper method may be more suitable for the model. In my own tests, the wrapper method appears slightly ahead. However, if we need to make a comparison according to this situation, we can say that the wrapper method gives better results as F1 score.

2 -) Which technique has given better results in terms of f1 score? (PCA or LDA)? Was it expected?

Both LDA and PCA are linear transformation techniques: LDA is a supervised whereas PCA is unsupervised – PCA ignores class labels. In contrast to PCA, LDA attempts to find a feature subspace that maximizes class separability. Remember that LDA makes assumptions about normally distributed classes and equal class covariances.

In the context of the appearance-based paradigm for object recognition, it is generally believed that algorithms based on LDA (Linear Discriminant Analysis) are superior to those based on PCA (Principal Components Analysis). In this communication, we show that this is not always the case. We present our case first by using intuitively plausible arguments and, then, by showing actual results on a face database. Conclusion is that when the training data set is small, PCA can outperform LDA and, also, that PCA is less sensitive to different training data sets.

I can say that I have a problem with LDA while trying it myself. Because I have seen that it gives much much lower results than much PCA. I think this is not an expected situation.

3 -) Have the filter feature selection and wrapper feature selection technique given similar set of features? Which attributes are different?

Since I had a problem while applying LDA, the maximum value of 19 took parameters. So I had to do all my operations by shrinking the dataset. I have observed that different clusters are actually formed in the big data. However, when I shrunk the dataset I found that all the features were the same.

So I will talk about the differences in big data. In particular, I observed that the values that are very different, that is, for each data column, are very different.

4 -) Which technique has given better results? (feature selection or dimension reduction)? Was it expected?

The difference is that the set of features made by feature selection must be a subset of the original set of features, and the set made by dimensionality reduction doesn't have to (for instance PCA reduces dimensionality by making new synthetic features from linear combination of the original ones, and then discarding the less important ones). This way feature selection is a special case of dimensionality reduction.

```
NAIVE BAYES METHOD, F1 SCORE : 0.8400267605287759
```

```
-----
```

```
PCC : [0.7300285113754563, 0.415185299845205, 0.7426355297258344, 0.7089838365853909]
```

```
FILTER FEATURE SELECTION, F1 SCORE : 0.8464231938568985
```

```
-----
```

```
WRAPPER METHOD, F1 SCORE : 0.8495431748549247
```

```
-----
```

```
PCA TOOL, F1 SCORE : 0.9562616126377763
```

```
-----
```

```
LDA TOOL, F1 SCORE : 0.6236878722101631
```

SOURCE

<https://www.explorium.ai/blog/demystifying-feature-selection-filter-vs-wrapper-methods/>

<https://dl.acm.org/doi/10.1109/34.908974>

<https://stats.stackexchange.com/questions/137100/what-is-the-difference-between-feature-selection-and-dimensionality-reduction>