

# Assignment report,

Machine Learning

MSc Applied Bioinformatics,

Marie Schmit

# Abstract

1. Enose samples are separated by their sensory scores. Sample 10F9 can be considered as an outlier. HPLC samples are less clearly clustered by samples, but three close groups still emerged. There are some potential outliers like 0F12 or 5F6.

Table PCA scatter plots of enose and HPLC grouped by sensors class

|  |  |
| --- | --- |
|  |  |

We try other analytical method to find better separation. First, we display the PCA plot with other methods (in 3D, in a biplot). We also use HCA. With this method, three clusters emerge, but they do not necessarily correspond to their sensory value. The previous outlier for enose (10F9) is no longer an outlier: it is now in the same cluster as F1a, which does not correspond to its sensory value. The same goes for HPLC, for which the clusters 2 and 3 are often mixed. This analysis is not the best.

|  |  |
| --- | --- |
|  |  |

We decide that the data that seem to stand out are not sufficiently so to be considered as outliers. We choose not to remove them to avoid the risk of introducing errors if they are significant.