Classify the glass samples using Gaussian Mixture Models (GMM)

Ville Sillanpää, k84338 - Lauri Viitanen, 338853
ville.sillanpaa@aalto.fi - lauri.viitanen@aalto.fi

March 31, 2015

1 Introduction

In this project we analyze a glass data set using both unsupervised and supervised machine learning approaches. The glass dataset [1] has been provided by the US Forensic Science Service and is available for download from the UCI machine learning repository. The dataset consists of samples for 6 types of glass (class label 4 is missing in the current data set); defined in terms of their oxide content (i.e. Na, Fe, K, etc). In particular, the dataset contains 214 samples characterized by 9 features and the samples have been categorized into 6 different classes of glass.

The goal of the project is to classify the glass samples using Gaussian Mixture Models (GMM) in unsupervised and supervised approaches. In the unsupervised approach we Use GMMs to find clusters from the full data set and compare the clustering results with the ground-truth class labels that are provided with the data set. In the supervised approach we fit a separate GMM for each class label and evaluate the classification performance using the fitted models on the test set. The dataset has been randomly divided into training (75class instances in the training and test sets. In addition, we compare the performance of the GMM model with the k-nearest neighbors classifier.

2 Methodology

3 Results

4 Discussion

[1] http://archive.ics.uci.edu/ml/datasets/Glass+Identification

Appendix A

Matlab code for asdf.