

Data Description:

The data contains features extracted from the silhouette of vehicles in different angles. Four "Corgie" model vehicles were used for the experiment: a double decker bus, Cheverolet van, Saab 9000 and an Opel Manta 400 cars. This particular combination of vehicles was chosen with the expectation that the bus, van and either one of the cars would be readily distinguishable, but it would be more difficult to distinguish between the cars.

Domain:

Object recognition

Context:

The purpose is to classify a given silhouette as one of three types of vehicle, using a set of features extracted from the silhouette. The vehicle may be viewed from one of many different angles.

Attribute Information:

- All the features are geometric features extracted from the silhouette.
- All are numeric in nature.

Learning Outcomes:

- Exploratory Data Analysis
- Reduce number dimensions in the dataset with minimal information loss
- Train a model using Principal Components

Objective:

Apply dimensionality reduction technique – PCA and train a model using principal components instead of training the model using raw data.

Steps and tasks:

1. Data pre-processing - Understand the data and treat missing values (Use box plot), outliers (15 points)
2. Understanding the attributes - Find relationships between different attributes (Independent variables) and choose carefully which all attributes have to be a part of the analysis and why (15 points)

3. Use PCA from scikit learn and elbow plot to find out reduced number of dimension (which covers more than 95% of the variance) - 20 points
4. Use Support vector machines to classify the class(y) of vehicles and find the difference of accuracy with and without PCA on predictors(X). 20 points
5. Optional - Use grid search (try C values - 0.01, 0.05, 0.5, 1 and kernel = linear, rbf) and find out the best hyper parameters and do cross validation to find the accuracy.

References:

- [Book on PCA](#)