Traffic Signs Classification

Andrea Laruelo

Nov 2017

INTRODUCTION

- A driver may miss some of the traffic signs on the road
    - overcrowding of neighbouring vehicles,
    - lack of concentration,
    - may be problematic and can lead to car accidents.

- Continuous growth of vehicle numbers around the world → problem expected to grow.

- Traffic Sign Recognition Systems aim to detect and identify road signs.
    - can be implemented on the automobile
    - inform road users about regulations, directions, warnings or potential risks on the road.
    - Examples: stop, “closed road”, speed limits.

- Vast amount of publications

- Emerging market (ADAS) / autonomous driving

- Releases of large traffic signs datasets → benchmarking of methods.

IMAGES

- Several datasets available

- German Traffic Sign Dataset
    - complete and
    - small enough to be feasible to work with.
    - can be downloaded from: http://benchmark.ini.rub.de/?section=gtsrb&subsection=dataset.

- 50.000 RGB images (40.000 for training and 12.000 for testing)

- 43 classes

- image sizes vary from 15x15 to 250x250 pixels

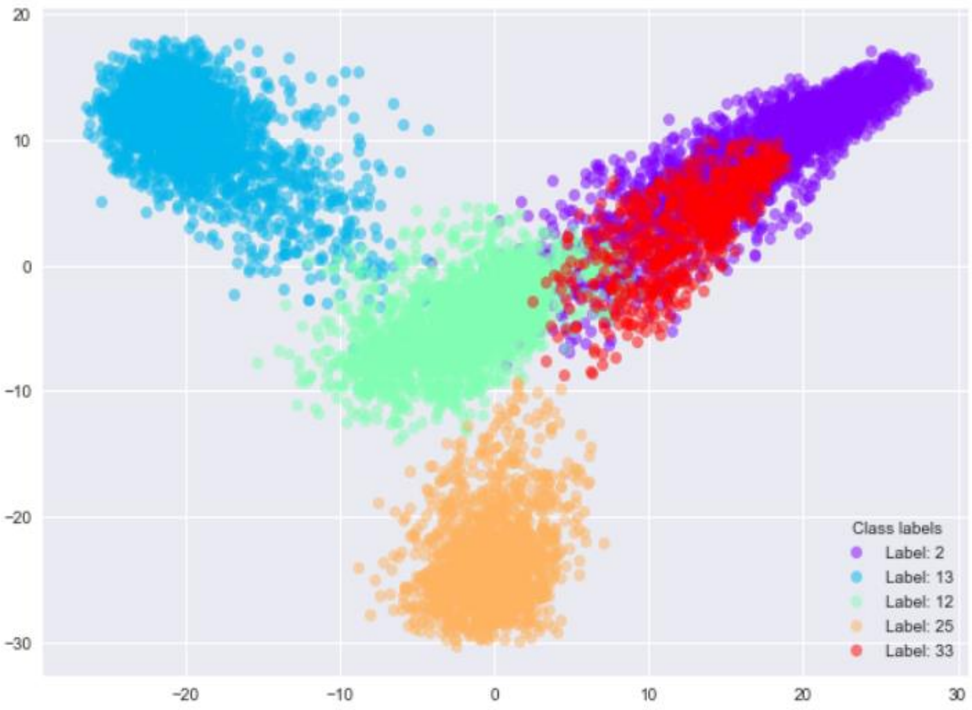
- images are not necessarily square.

- Image annotations are provided in CSV files containing the filename, the size, the coordinates of the bounding box containing the sign and the class label.

- Unbalanced data set

PRE-COMPUTED FEATURES

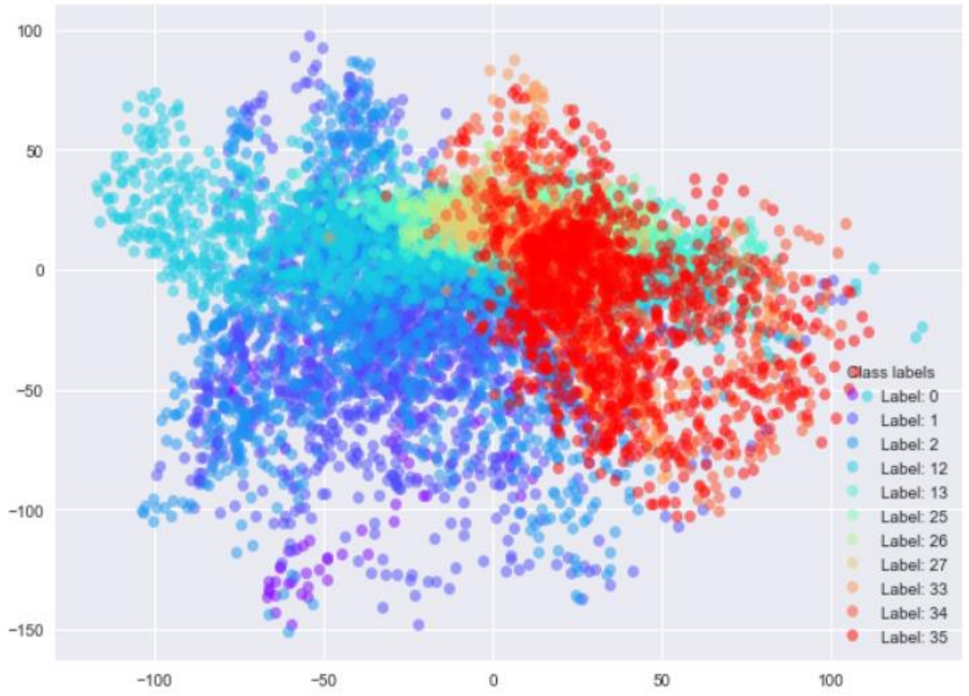
- Histograms of Oriented Gradient (HOG):
  - proposed by Dalal and Triggs (2005) for pedestrian detection
  - based on gradients of color images, different weighted and normalized histograms are calculated
  - provide a good representation of the data set
  - the first two principal components provide already a quite clear separation of different sign shapes.



- Labels 0,1,2: speed limit (blue)
- Label 12: diamond (light blue)
- Label 13: inverted down (turquoise)
- Labels 25,26,27: up triangle (green)
- Labels 33,34,35: round blue signals (red)

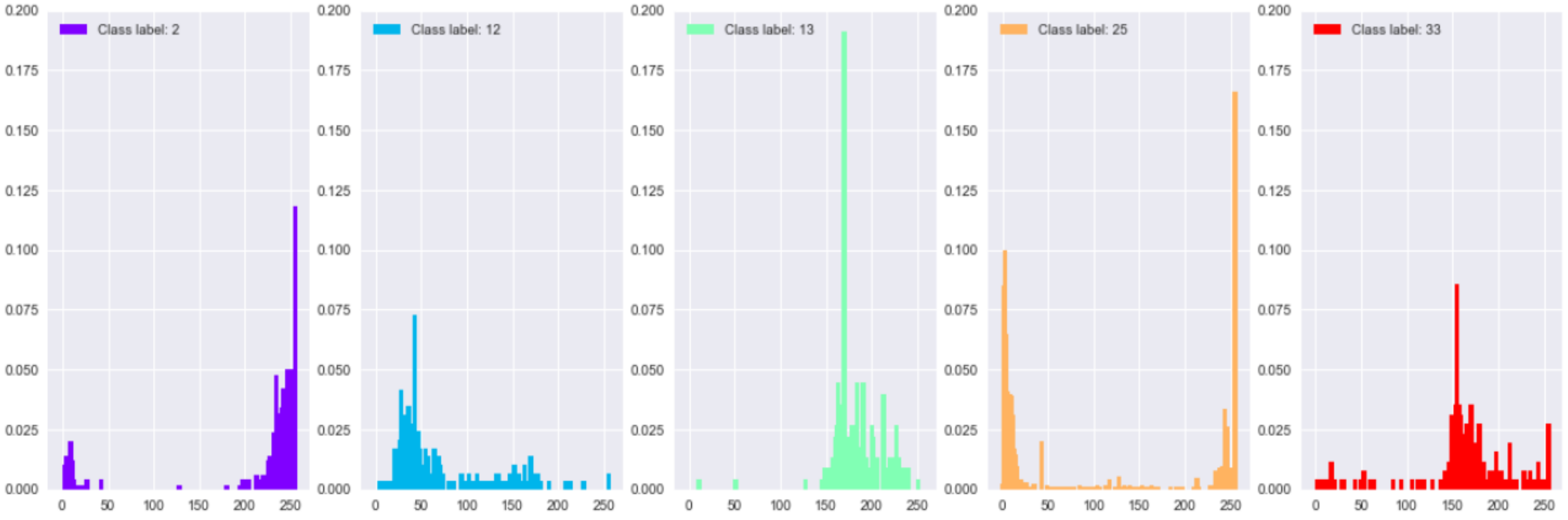
PRE-COMPUTED FEATURES

- Haar-like features
- For each image, 5 different types of Haar-like features were computed in different sizes for a total of 12 different features.
- The overall feature vector contains 11,584 features.
- The figure below that HAAR features also provide a certain degree of class separation.



PRE-COMPUTED FEATURES

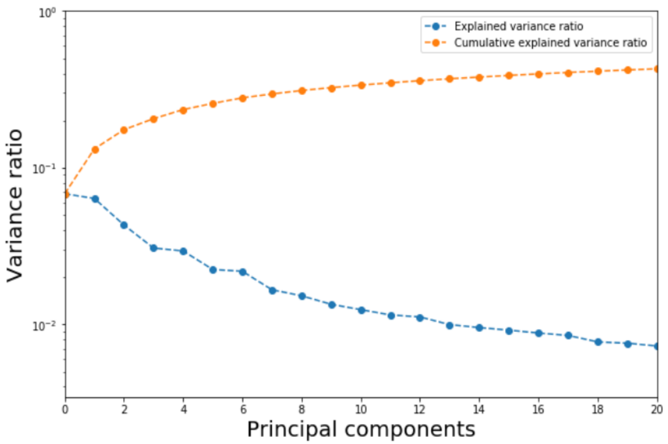
For each image in the training set, a 256-bin histogram of hue values (HSV color space) is delivered.



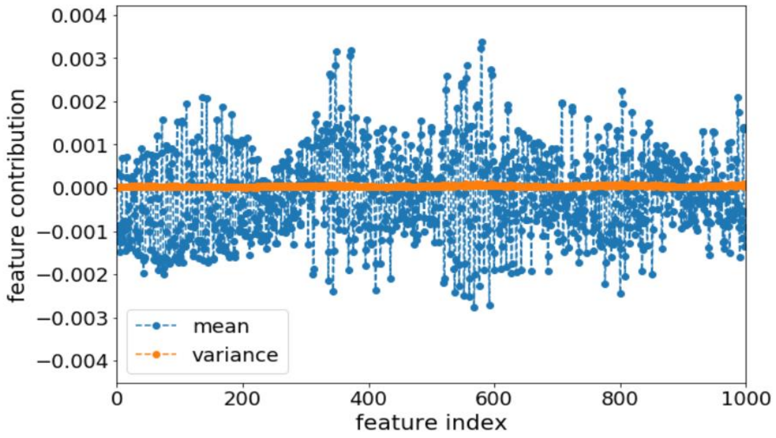
FEATURES USED FOR CLASSIFICATION

1. Feature selection:

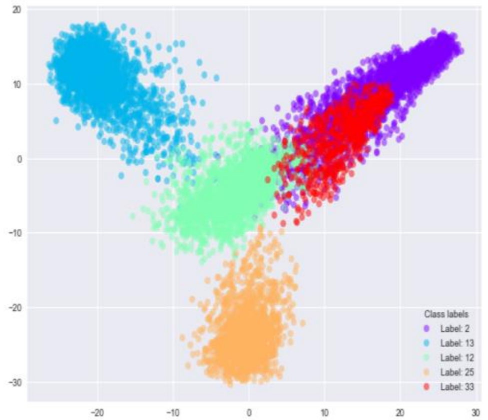
- features provided have shown to provide useful insights about the data
- However, the total number of features is around ~17000.
- PCA feature selection:
    - identify the ‘most important’ variables in the original feature space
    - Features that contribute most to the most important principal components.
    - a dimension that has not much variability cannot explain much of the data.
    - normalize the data before applying PCA in order to prevent that features having higher values dominate



According to the cumulative explained variance, by keeping only 20 principal components we can describe more than the 80% of the full variance.



Keep features with largest absolute mean for the classification task.



2. HOG2 features:

In a second step, we have chosen to perform classification using the HOG2 features. This selection is based in the good results reported using HOG2 features in previous studies [7]. Results from PCA and HOG2 are then compared.

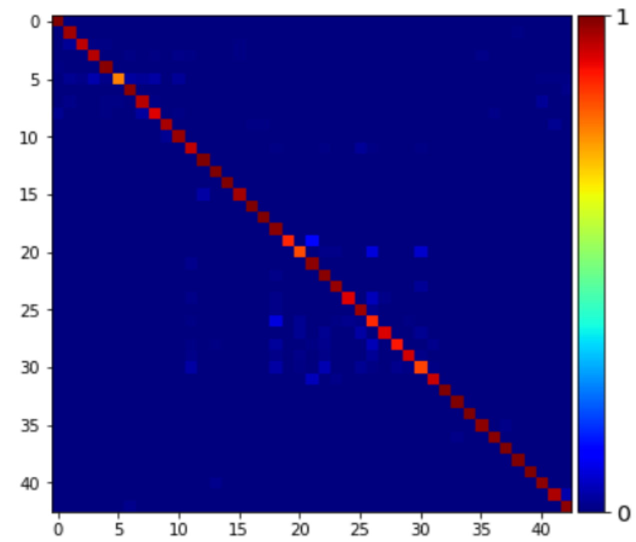
LINEAR DISCRIMINANT ANALYSIS

- linear classification technique
- finds the axes that maximize the separation between multiple classes
- provides surprisingly good results in practice despite its simplicity (Hastie et al., 2001).

Results using HOG2 features:

- Training data set: 98.96%
- Test data set: 95.68%

Confusion Matrix:



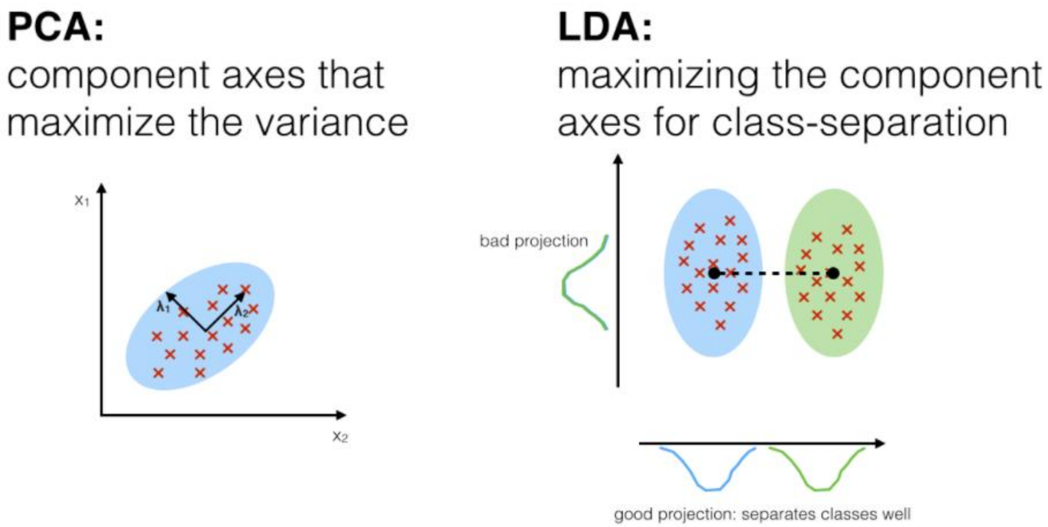


Figure from http://sebastianraschka.com/Articles/2014_python_lda.html

LOGISTIC REGRESSION

- one of the most popular supervised classification algorithms
- mostly used for solving binary classification problems
- can also be used for multiclass problems
- has shown to be an efficient model for multiclass classification while being much simpler and faster than other classifiers
- results comparable to the results from LDA on (the HOG2 features).
- Results on PCA selected features are very poor:
    - This can be explained by the fact PCA finds out which features are important for best describing the variance in a data set, so it may hide features that contribute little to the variance but that are significant class differentiators.

- In view of this poor results using PCA selected features, the following classification model are also tested on HOG2 features.

Results using PCA:

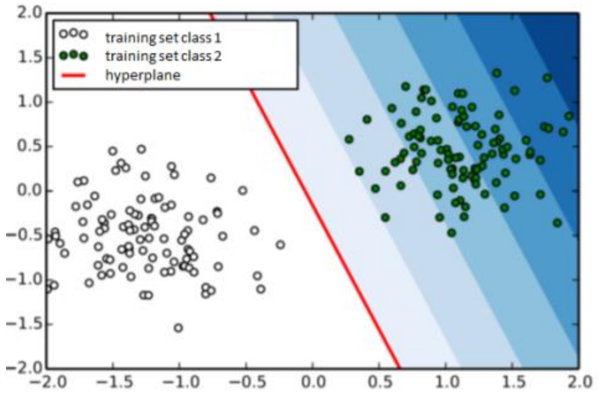
- Training data set: 12%
- Test data set: 10.6%

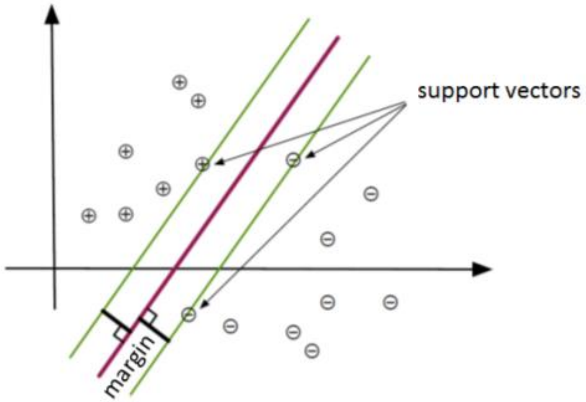
Results using HOG2 features:

- Training data set: 99.89%
- Test data set: 95.64

SVM

- Support vectors machine models try to find the optimal hyperplane/surface that separates the classes
- The minimal distance from the separating hyperplane to the closest training point is called margin.
- The training samples that lie on the margin define the border between the two classes and are referred to as support vectors.





Results using HOG2 features:

- Training data set: 85.12%
- Test data set: 80.92%

Results using **standardized** HOG2 features:

- Training data set: 99.98%

Test data set: 95%

SVM methods are very sensitive to the magnitude of the features.

RANDOM FORESTS

- Introduced by Breiman and Cutler in 2001.
- To classify a sample, the classification of each random tree in the forest is taken into account.
- The class label of the sample is the one with the majority of the votes.
- Achieve state-of-the-art performance in many multi-class classification applications.
- In [Zaklouta et al. 2011], the performance of random forest classifiers is studied for different parametrizations.

Results using HOG2 features:

- Training data set: 99.63%
- Test data set: 96.01%

- Best results compared to the previous methods.
- The reason for this may be that Random Forests are intrinsically suited for multiclass problems, while SVM is intrinsically two-class.
- No sensitive to hyperparameters tunning (except the number of trees)
- On the contrary, there are a lot of very sensitive parameters to be turned in SVMs, like the kernel or the regularization penalties.
- Summarizing, SVM requires a much more careful analysis and tend to provide good results if there are fewer outliers in the dataset, while Random forests are less sensitive to hyperparameters and (if there is enough data available) usually come up with a pretty good model.

FUTURE WORK

Deep learning

- Traffic sign detection is a challenging task

- In this study we have used well known machine learning algorithms

- Traditional methods rely pre-computed features that may be latter on selected according to different criteria → approach has not good enough for complex problems.

- Increasing interest on deep learning approaches.

- Deep learning models:
    - circumvent the feature extraction step
    - learn by themselves what features they should focus on

- Deep learning has shown to be able to provide very good performance in complex problems such as image classification.

- Therefore a natural next step would be to use a deep learning approach to classify our traffic signs dataset.