Final project
Image recognition

Guillaume Wisniewski
(d'après une idée d'A. Allauzen)
guillaume.wisniewski@limsi.fr

March 2018

Abstract
The goal of this project is to design an image recognition system. This is a 'research project': the description of the work to do is voluntarily fuzzy and you will have to use your creativity to solve the problem completely while showing us that you have understood the main concepts of statistical learning.

The work has to be done in pairs. The project will be evaluated by a (written) report and an oral presentation.

1 Overview

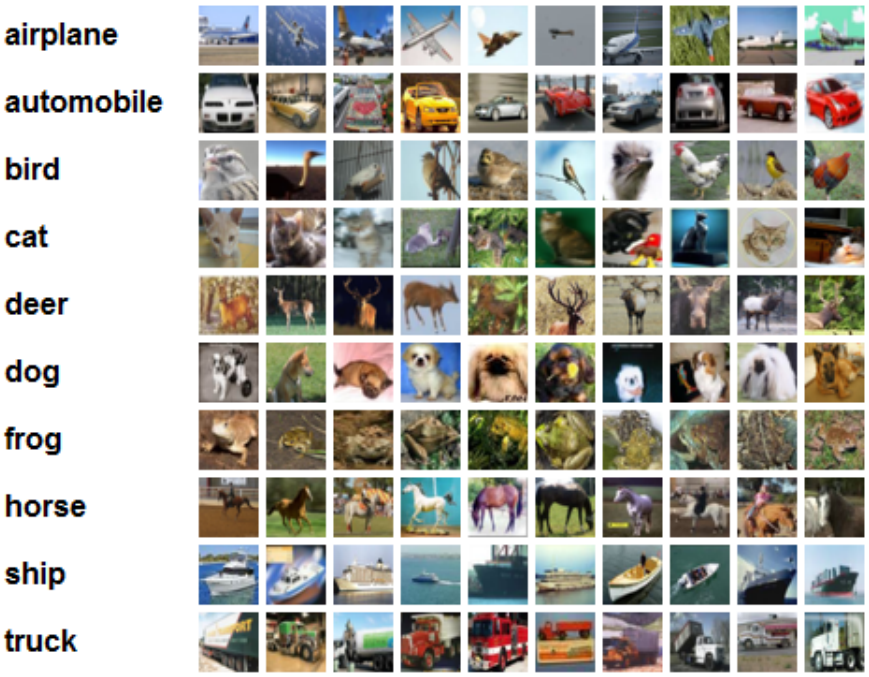

Figure 1: Example of images to recognize

This project aims at developing an image recognition system able of automatically identifying the object represented on an image. Figure 1 shows several examples of images and their corresponding class. The main idea of the project is to use the k-mean algorithm to find automatically the most relevant representation of an image; a 'standard' classification method like the perceptron can then predict the class. This approach was originally proposed in:

> An Analysis of Single-Layer Networks in Unsupervised Feature Learning, A. Coates, H. Lee et A. Y. Ng, Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, 2011

This article can be downloaded from: `http://jmlr.csail.mit.edu/proceedings/papers/v15/coates11a/coates11a.pdf`

2 The Cifar-10 dataset

In all our experiments, we will use the CIFAR-10 dataset[1] that gather images of 10 different kinds of object (dog, truck, aircraft, ...).

[1] `http://www.cs.toronto.edu/~kriz/cifar.html`.

You should first familiarize yourself with the database format (how are the data stored?), the type of images represented (is the task difficult for a human? for a computer? what error rate can we expect to achieve? ...) and the 'general' statistics of the dataset (how many classes are there? how many examples by class are there? what is the class distribution? ...).

CIFAR-10 is a standard dataset that has been used to test new machine learning algorithms / approaches for many years. It is very easy to find many articles reporting the performance achieved by new ML algorithm as well as standard baselines. What is the error rate achieved by the latest ML algorithms? Can we expect to achieve such performance in this project?

Your final report must include a summary of these observations and a discussion on the task interest.

Be carefull: The CIFAR-10 dataset is much larger than the datasets we have been using during the different labs. You should start working with a smaller dataset (e.g. by reducing the number of classes or the number of examples) if you do not want to loose too much time loading the dataset or debugging your learning algorithm.

3 Naive approach

A very simple way to solve the task at hand is to apply a classification algorithm directly to the images (i.e. by considering that each image is represented by a vector describing its pixel values).

How to represent pixels ? What are the performances achieved ? Which classes are predicted correctly? which ones are systematically not recognized? Why?

It is also possible to use the k-mean algorithm to cluster the images. What distance can we use? Can this clustering be used to identify objects? If yes, What performance do you get?

4 Learning the feature vector

We will now implement the method proposed by A. Coates *et al.* (see reference in the §1). This method is summarized in Figure 1 of the article. It is fundamental you understand it.

The main idea of this approach is to automatically build a representation of the image by a vector of real numbers. To do this, we start by extracting from the training set a (very) large number of *patches*. A patch is simply a small part of an image. By multiplying patches, we can hope to capture some important characteristics (e.g., a ear or an eye). It is then possible to represent an image from the patches that compose it.

The first step is to build a *dictionary* of N patches:

- each picture is divided in 4 patches;
- we gather together all the patches of the training set;

- we use the k-means algorithm to find N representative of this dataset; they will make our 'patch dictionary'

A new picture can then be described as follow:

- the picture is divided in 4 patches;
- we compute the distance between each patch and each element of the dictionary;
- each patch is represented by a vector of N binary features; the i-th component is equal to 1 if and only if the center of the i-th cluster is the closest to the patch;
- the picture is described by concatenating the vectors of the 4 patches.

This representation is used to train a standard classifier.

5 Work to do

The project has to be done in pairs. You have to :

- answer the various questions on the subject;
- implement and test the naive approach;
- implement and test the proposed approach;
- test the impact of the different parameters of the method,
- analyze the results obtained and dictionaries found.

Your work will be evaluated by a 20 page reports and an oral defense.

The report have to be sent to Guillaume Wisniewski (`guillaume.wisniewski@limsi.fr`) and Véronique Ventos (`veronique.ventos@lri.fr`) before April 13th 2018 at midnight. The oral defense will take place April 18th 2018.