Assignment: Artificial Neural Networks & Deep Learning

14 de mayo, 2019

Data sets: Caltech 101 Description

In the web page:

http://www.vision.caltech.edu/Image_Datasets/Caltech101/

there are pictures of objects belonging to 101 categories. About 40 to 800 images per category. Most categories have about 50 images. Collected in September 2003 by Fei-Fei Li, Marco Andreetto, and Marc 'Aurelio Ranzato. The size of each image is roughly 300 x 200 pixels.

Questions:

There are 2 sections: questions about NN architectures and questions about Deep Learning architectures. Notice that the first section is the task previously proposed in class.

NN architectures

- Select five categories from caltech101 collection. Try to select balanced categories with representative number of images.
- Implements a classifier fed with the HOG descriptors that learns the category to which a given image belongs. In HOG use 3 cells and 9 orientations. Compare the performance of two NN architectures:
  1. Layer 1: 81 units (input layer), Layer 2: 10 units (hidden layer), Layer 3: 5 units (output layer),
  2. Layer 1: 81 units (input layer), Layer 2: 50 units (hidden layer), Layer 3: 25 units (hidden layer), Layer 4: 5 units (output layer).

Question 1 : Reshape grayscale images to 32x32. Vectorize images by rows, to get a 1024-vector representation of each image. Implement a NN with two hidden layers that uses such a vectors as input data for learning image categories. Tune the number of neurons in each hidden layer using a grid search procedure.

Question 2 : Compare the three NN model at least in two characteristics: the performance and CPU time.

Deep Learning architectures:

- Categories: airplanes (800), Motorbikes (798) and Faces (435) are large categories. For each category reshape grayscale images to 32x32. Split the dataset in two halves (training and test). From the training part, fit deep learning models according with the following architectures:

CNN architecture:

- Convolution layer with filters = 20, kernel_size = c(5,5), activation = 'relu',
  - Pooling layer with: pool_size = c(2, 2)
- Convolution layer with filters = 40, kernel_size = c(3,3), activation = 'relu',
  - Pooling layer with: pool_size = c(2, 2)
- Fully connected layer with: units = 128, activation = 'relu'
- Output layer with: units = num_classes, activation = 'softmax'

Question 3 : Assess the performance of the CNN predicting the categories of test images and obtain the confusion matrix.

Question 4 : Implement an script that classifies a new image that belongs to one of the three categories: airplanes, motorbikes and faces, using the trained CNN network. The script should accept as input an image in png format, resize the image to 32 x 32 and apply the CNN to get the classification.

Autoenconder architecture:

```
1. Layer 1: 1024 units (input layer),
2. Layer 2: 100 units (hidden layer),
3. Layer 3: 50 units (hidden layer),
4. Layer 4: 100 units (hidden layer),
5. Layer 5: 1024 units (output layer)
```

Question 5: For all images, represent the deep features in layer 3 using PCA or t-SNE to visualize the categories representation in the plot.

Important remarks

- You should do an R markdown (or R latex) report as dynamic as you can. Try to use r code into paragraph.
- Use relative paths instead of absolute paths to read / write files, to make it easier to run the code outside of your computer.

Delivery / Deadline

A zip file including the data set, the Rmd (or Rsw) file used as template for the report and output reports in pdf and html files.

Deadline: 3th of June, 2019

Score of each question (a revisar)

- Questions 1 & 2 (40%)
- Questions 3 & 4 (30%)
- Question 5 (20%)
- Dynamic report quality (10%)