

CS423: Data mining

Assignment 1: Feature extraction

February 1, 2019

Instruction

In this assignment, you will be using a subset of CIFAR-10 image dataset to train a Support Vector Machine (SVM) classifier so that you can use the model to predict an unseen image belonging to one of ten possible classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck.

Unfortunately, the images you have are still images in RGB format. However, we have learnt from the class that there are several ways to convert an image into its useful vector representation. The simplest way is simply reshaping the image matrix into a vector. Of course, a more sophisticated methods do exists. It is your task to experiment with various image feature extraction methods including the ones you have seen from the class. You can also use others image features that were not discussed in the class. The goal is to learn the SVM classifier so that it can predict unseen images as accurately as possible.

- SVM will learn from the first 4000 images in the folder.
- SVM will be evaluated using the last 1000 images in the folder.
- To set the standard for measuring accuracy, we shall use default setting for SVM classifier.
- The success of feature extractions is measured by classification accuracy on the last 1000 images.
- Code template is provided in the hw1.zip file.

Useful Julia packages

LIBSVM, Image, ImageFeatures, Statistics: <http://juliaimages.github.io/latest/>

What to hand in ?

A zipfile containing

1. A report explaining your feature extraction algorithm in details. (HW1_5xxxxxxxx.pdf)
2. A source code of your feature extraction algorithm. (HW1_5xxxxxxxx.jl)

Email the zipfile to jakramate.b@cmu.ac.th using '[CS423-HW1]' prefix in your email title.