Project-II by Group Sydney

Diego Antognini & Jason Racine
EPFL
diego.antognini@epfl.ch, jason.racine@epfl.ch

Abstract

This report provides a summary of the project two of the PCML class.

1 Binary Classification

1.1 Data Description

The train-data for binary classification consists of $N = 6000$ images. As input we have 2 representations of the image : *the histogram of oriented gradients* (HOG) $\mathbf{X}_{hog}$ and the *overFEAT ImageNet CNN features* (CNN) $\mathbf{X}_{cnn}$. Our input $\mathbf{X}$ is the concatenation of $\mathbf{X}_{hog}$ with $\mathbf{X}_{cnn}$.
Each input sample is a vector $\mathbf{x}_n$ with dimension $D = 42273$ (5408 for the HOG and 36865 for the CNN) and is the concatenation of $\mathbf{x}_n^{hog}$ with $\mathbf{x}_n^{cnn}$ The output ($\mathbf{y}$) represents the classification of the images, with 0 if it represents a car, a horse or a plane, 1 otherwise.

We also have test-data of size $N = XXX$ without their corresponding output. Our goal is to produce predictions for those data, as well as an approximation of the test-error using *Balanced Error Rate* (BER).

1.1.1 Histogram of Oriented Gradients

Histogram of oriented gradients is a feature used in computer vision in order to detect objects. To compute it, we first need to normalize the image, compute the gradients (of each pixel) for the different color channel and take those with the largest norm. Then we decompose the image in bins of size $w \times h$. For each of those bins, we compute an histogram of the orientation of the gradients using theirs angles and weighted by theirs magnitudes. The histogram has n bins from 0 to 180 degrees. We compute this histogram using 4 different normalizations.

In our case, the dimensions of this feature is $13 \times 13 \times 32 = 5408$, where the first 13 is the number of bins in the height, the second 13 the number of bins in the width. Finally 32 is 4×8 where the 4 is the different normalizations for the histogram and 8 the number of bins (each bin has a size of 22.5).

1.1.2 OverFEAT ImageNet CNN features

Those features are extracted from a convolutional network-based image features extractor OverFeat. They were trained on the ImageNet dataset (tens of millions images). The size is the output of the fifth layer of the neural network which is $1024 \times 6 \times 6 = 36864$. In our feature, we also have an extra dimension which can be ruled out.

1.2 Data visualization and cleaning