Machine Learning Engineer Nanodegree

Capstone Proposal

Tushar Bansal

August 13th, 2017

Domain Background

Handwriting recognition is the ability of a computer to receive and interpret intelligible handwritten input from sources such as paper documents, photographs, touch-screens and other devices. The image of the written text may be sensed "off line" from a piece of paper by optical scanning (optical character recognition) or intelligent word recognition. Alternatively, the movements of the pen tip may be sensed "on line", for example by a pen-based computer screen surface, a generally easier task as there are more clues available.

Off-line handwriting recognition involves the automatic conversion of text in an image into letter codes which are usable within computer and text-processing applications. The data obtained by this form is regarded as a static representation of handwriting. Off-line handwriting recognition is comparatively difficult, as different people have different handwriting styles. And, as of today, OCR engines are primarily focused on machine printed text and ICR for hand "printed" (written in capital letters) text.

Problem Statement

Narrowing the problem domain often helps increase the accuracy of handwriting recognition systems. A form field for a U.S. ZIP code, for example, would contain only the characters 0-9. This fact would reduce the number of possible identifications.

Primary techniques:

Specifying specific character ranges

Utilization of specialized forms

Our Final goal is to predict the digit written in an image.

Datasets and Inputs

The MNIST database (Modified National Institute of Standards and Technology database) is a large database of handwritten digits that is commonly used for training various image processing systems. The database is also widely used for training and testing in the field of machine learning. It was created by "re-mixing" the samples from NIST's original datasets. The creators felt that since

NIST's training dataset was taken from American Census Bureau employees, while the testing dataset was taken from American high school students, it was not well-suited for machine learning experiments. Furthermore, the black and white images from NIST were normalized to fit into a 20x20 pixel bounding box and anti-aliased, which introduced grayscale levels.

The MNIST database contains 60,000 training images and 10,000 testing images. Half of the training set and half of the test set were taken from NIST's training dataset, while the other half of the training set and the other half of the test set were taken from NIST's testing dataset. There have been a number of scientific papers on attempts to achieve the lowest error rate; one paper, using a hierarchical system of convolutional neural networks, manages to get an error rate on the MNIST database of 0.23 percent. The original creators of the database keep a list of some of the methods tested on it. In their original paper, they use a support vector machine to get an error rate of 0.8 percent.

The set of images in the MNIST database is a combination of two of NIST's databases: Special Database 1 and Special Database 3. Special Database 1 and Special Database 3 consist of digits written by high school students and employees of the United States Census Bureau, respectively.

Solution Statement

I will use Support Vector Machine (SVM) algorithm with raw pixel features. The solution is written in Python with the use of sci-kit-learn easy to use machine learning library.

I use two approaches to SVM learning. First, uses classical SVM with RBF kernel. The theory behind is quite complicated, however, sci-kit-learn has ready to use classes for kernel approximation. We will use:

Nystroem kernel approximation

Fourier kernel approximation

Benchmark Model

I will use the Simple one-layer neural network as my primary benchmark model. Additionally, I will refer to many other models on Kaggle hand written digit recogniser competition. But, I will stick to SKLearn now. And try to achieve the best accuracy.

Evaluation Metrics

The goal in this is to take an image of a handwritten single digit and determine what that digit is.

For every ImageId in the test set, you should predict the correct label.

This is evaluated on the categorization accuracy of your predictions (the percentage of images you get correct).

Project Design

We will first download the MNIST database using sk-learn and preprocess it using sk-image and PCA for feature selection. I will use SVM with raw pixel features. Then will use SVM with RBF Kernel to achieve high accuracy. We will also plot confusion matrix which will help us visualise the differences between actual and predictions. Using Grid Search CV we will optimise our model which will choose best value of gamma and c.