Digit Recognition System

CSL 402

Digital Image Analysis

By

Imroj Qamar (2011CS1011)

Rahul Kumar (2011CS1030)

Sahil Dabra (2011CS1033)

Table of Contents

Abstract

Handwritten digit recognition is the ability of the computer to interpret and classify handwritten digits taken inputs such as a documents, images, pen based computer screen surface. It has various applications such as Zip Code reader in US Mail Service, Input reading on a pen based computer screen surface. It is a subtype of Optical Character Recognition which uses the concepts of Object Recognition. **Object recognition** – in image processing is the task of finding and identifying objects in an image or video sequence.There are many techniques for the task of object recognition , some of them are employed in this project for the task of handwritten digit recognition .

Introduction

The process of Digit Recognition can be broken into 3 steps

1. Pre processing
2. Feature Extraction
3. Classification

Pre-processing is usually done to discard irrelevant information for achieving high speed and accuracy. It involves binarization, normalization, sampling, smoothing and denoising. Feature Extraction is done to highlight important information for the recognition model. Classification involves mapping of the extracted features to different classes to identify the different digits. Complete digit recognition systems involve formatting, segmentation and noise removal, feature extraction and recognition. This project deals mostly with the last stage classification .In this project we have implemented classification based on four techniques namely :

1. Artificial Neural Network (ANN)
2. Linear Classification
3. Naïve Bayes Classification
4. Support Vector Machine (SVM)

Classification Techniques Used :

1. Linear Classification

Theory : It is a probabilistic classification model. The probabilities describing the possible outcomes of a single trial are modelled, as a function of variables, using a [logistic function](http://en.wikipedia.org/wiki/Logistic_function). It models the relationship between independent and dependent variables using a binary predictor which applies logistic function on a real valued result from the dependent variables and then decides the class of output using value of logistic function .The logistic function usually used is Sigmoid function or sometimes simple sign checking (positive or negative).

F = w0 + w1x1 +…… wnxn = wtx

Logistic(F) -> (0,1)

Where x = x1……xn  is the feature of the object and the w are the learned parameters.

Approach :

In this problem we have to classify object into 10 classes , so it’s apparent that a simple binary classifier like Logistic Regression cannot be used . But with certain modification even logistic regression can be used for multiclass classification. We use the **one vs rest** technique where a single classifier is trained per class to distinguish that class from all other classes. Prediction is then performed by predicting using each binary classifier, and choosing the prediction with the highest confidence score.

Select : yi arg max(1….k) fk (x)

Work Performed/Experiment :

To train the logistic regression we have used an error term and varied the parameter wi to achieve the minimum error :

Error is given by : yilog(sigmoid(wtx))+(1-yi) log(1-sigmoid(wtx))

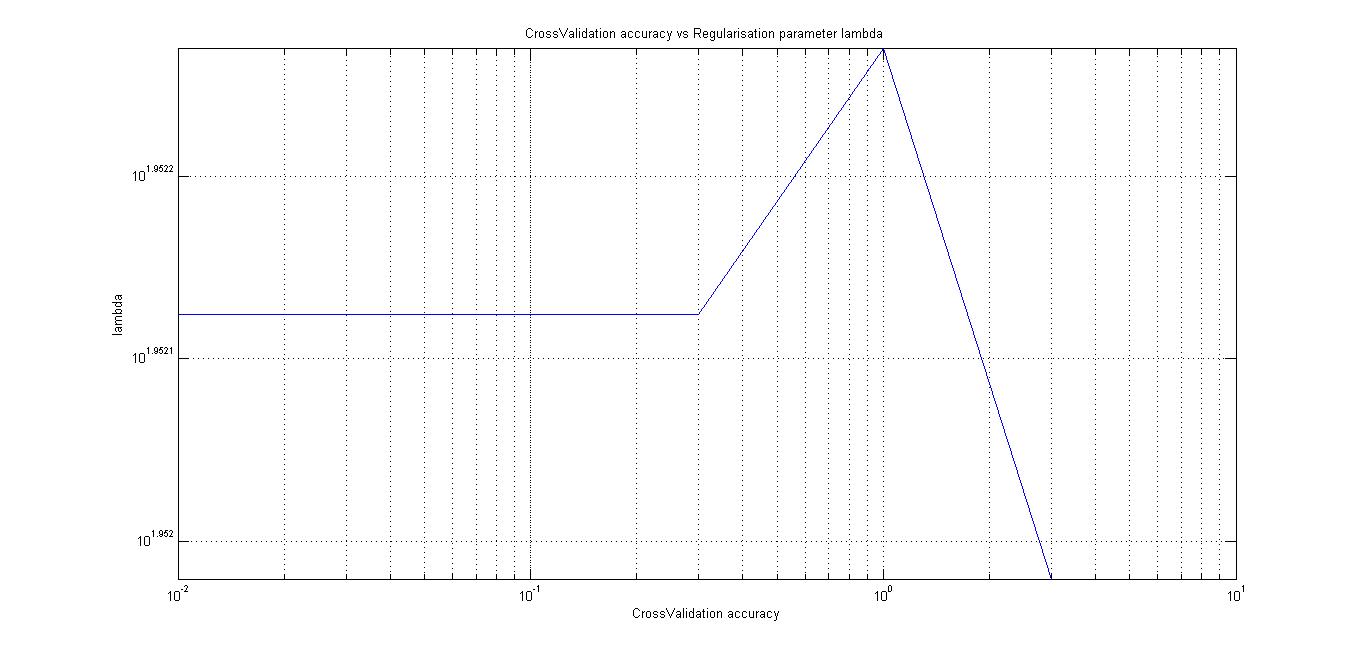
But as it is common with curve fitting it is possible that the solution can be overfitting even though it gives minimum error hence we introduce regularisation term lambda also in the error.

So given the error equation we minimise it by using the technique of Gradient Descent .IN this technique we update wi by the derivative of wi until convergence .

wi  = wi  - dE/dwi

Experiment :

After implementing the algorithm we are left with the task of parameter tuning . Since we are using regularisation parameter lambda we try to find the optimal value of lambda so as to maximise the accuracy . Hence accuracy vs lambda graph is plotted :



From the graph it can be seen that the maximum accuracy is achieved at lambda = 1 . Hence lambda =1 is used for training the parameters .

1. Artificial Neural Network

(a)Theory : Artificial Neural Network(ANN) is a network based model that maps sets of input data to different output classes. It has an input layer, output layer and hidden layers which have non-linear activation function at each node so that the decision regions are bounded by smooth lines instead of curves. The learning process of ANN is conducted with error Back Propagation learning algorithm derived from generalized delta rule. In Back Propagation, the weights or the parameters associated with each node is changed on every input based on the amount of error in output compared to the result .

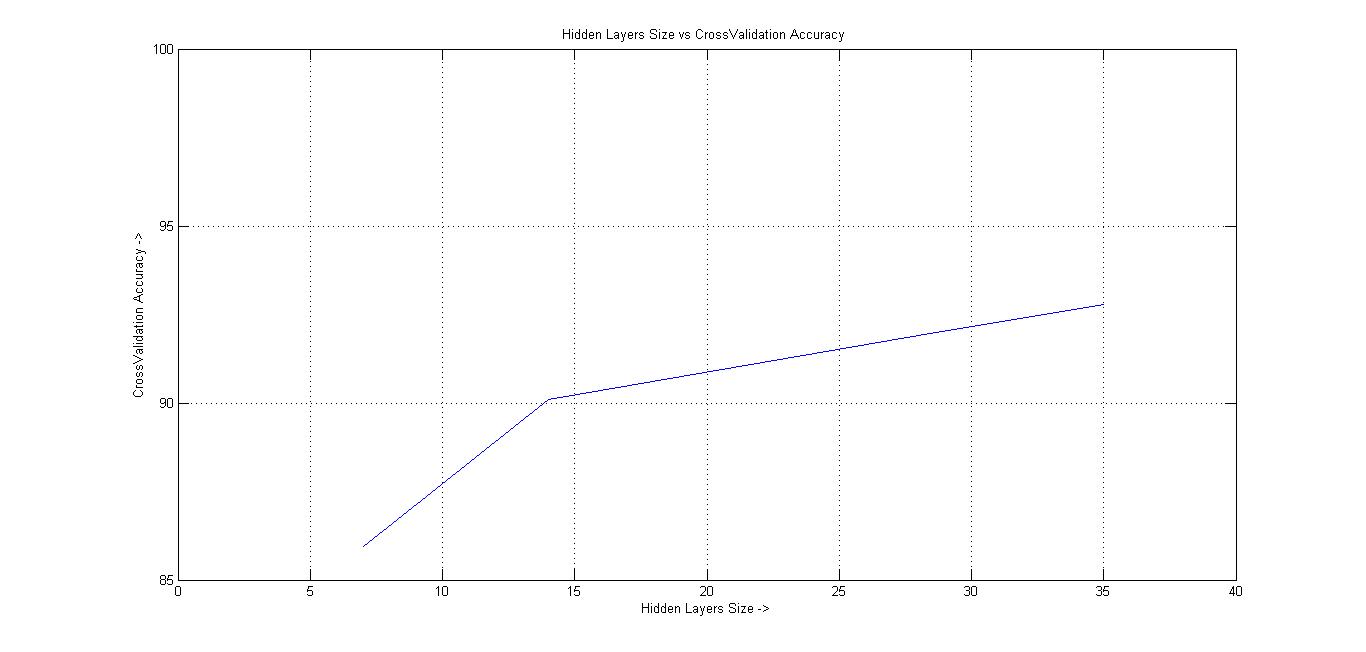
The probabilities of classes are calculated by Forward propagation where value of hidden layers are calculated using sum of linear combination of weights of input layer .Similarly final layer values are found using linear comination of weights of hidden layers .

Approach :

As the ANN given the sigmoid value for each class it has no problem with multiclass classification . In the back propogation algorithm the error is minimised by using Stochastic Gradient descent method.

Work Performed / Experiment :

The tunable parameter in ANN is the size of the hidden variables . Hence we change the size of hideen ayer and check its effect on the cross validation accuracy of the system.



IT can be easily seen from the figure that as the hidden layer size increases the accuracy increases . But the limitation of increasing hidden layer is that the computation time increases highly with increase in hidden ayer size . Hence we stop at hidden layer size = 35 as a trade off with time .

1. Support Vector Machine

(a)

Theory :

Support vector machine (SVM) is [supervised learning](http://en.wikipedia.org/wiki/Supervised_learning) model with associated learning [algorithms](http://en.wikipedia.org/wiki/Algorithm) to analyze data and recognize patterns, used for [classification](http://en.wikipedia.org/wiki/Statistical_classification) and [regression analysis](http://en.wikipedia.org/wiki/Regression_analysis). It is a non-[probabilistic](http://en.wikipedia.org/wiki/Probabilistic_classification) [binary](http://en.wikipedia.org/wiki/Binary_classifier) [linear classifier](http://en.wikipedia.org/wiki/Linear_classifier). It builds a model which is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. SVM can also perform non-linear classification using the kernel trick. It constructs a set of hyperplanes in a infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the [generalization error](http://en.wikipedia.org/wiki/Generalization_error) of the classifier.A SVM may be linear /non linear depending on the kernel used .

(b) Approach :

Since the SVM classifies data into two categories , digit recognition is not possible with simply one SVM . Hence we used one against one approach in which we train multiple SVMs for each possible pair of digits and while prediction use the trained SVMs and select the class with whch maximum SVMs are in favour .

c)Word Done / Experiment :

We tried to implement this SVM using different kernels and found different results for each which are tabulated below .

|  |  |
| --- | --- |
| Kernel Used | Accuracy |
| Linear | 90.97% |
| Gaussian (rbf) | 30.31% |
| Polynomial (degree 1) | 91.39% |
| Polynomial (degree 2) | 93.28% |
| Polynomial (degree 3) | 80.29% |

(d) Naïve bayes :

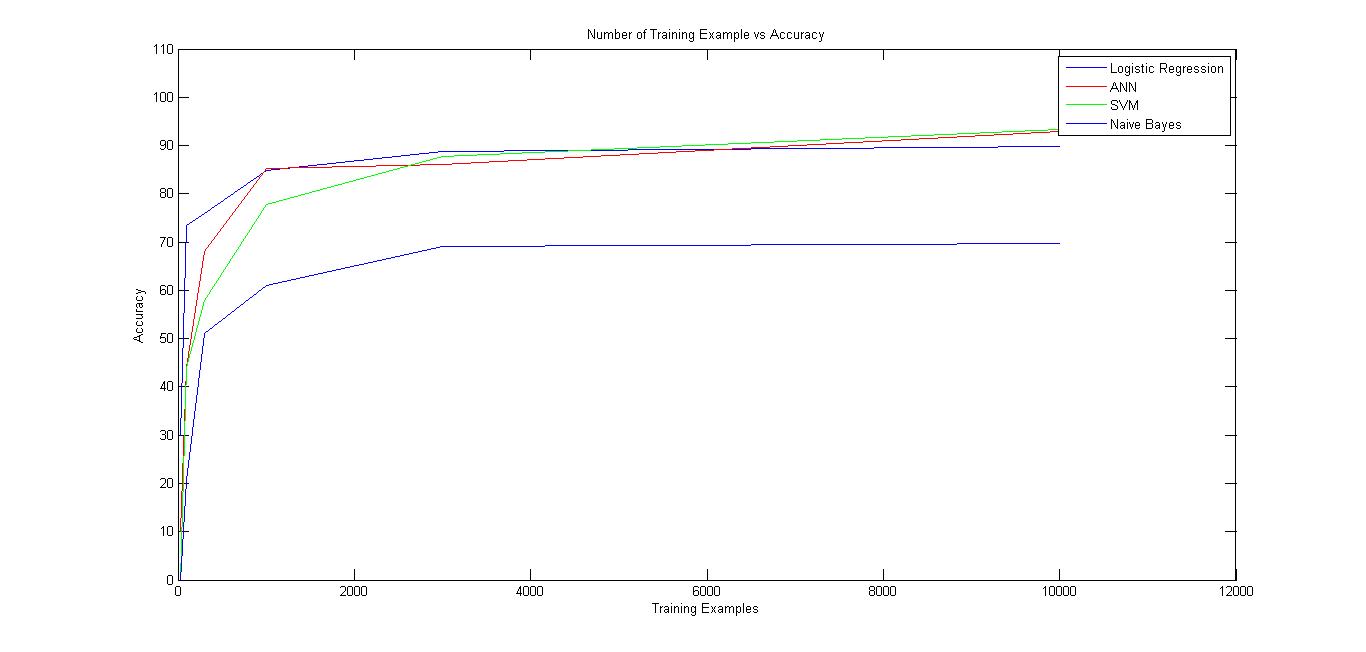
Theory :

http://en.wikipedia.org/wiki/Naive\_Bayes\_classifier

Comparison of Different Techniques :

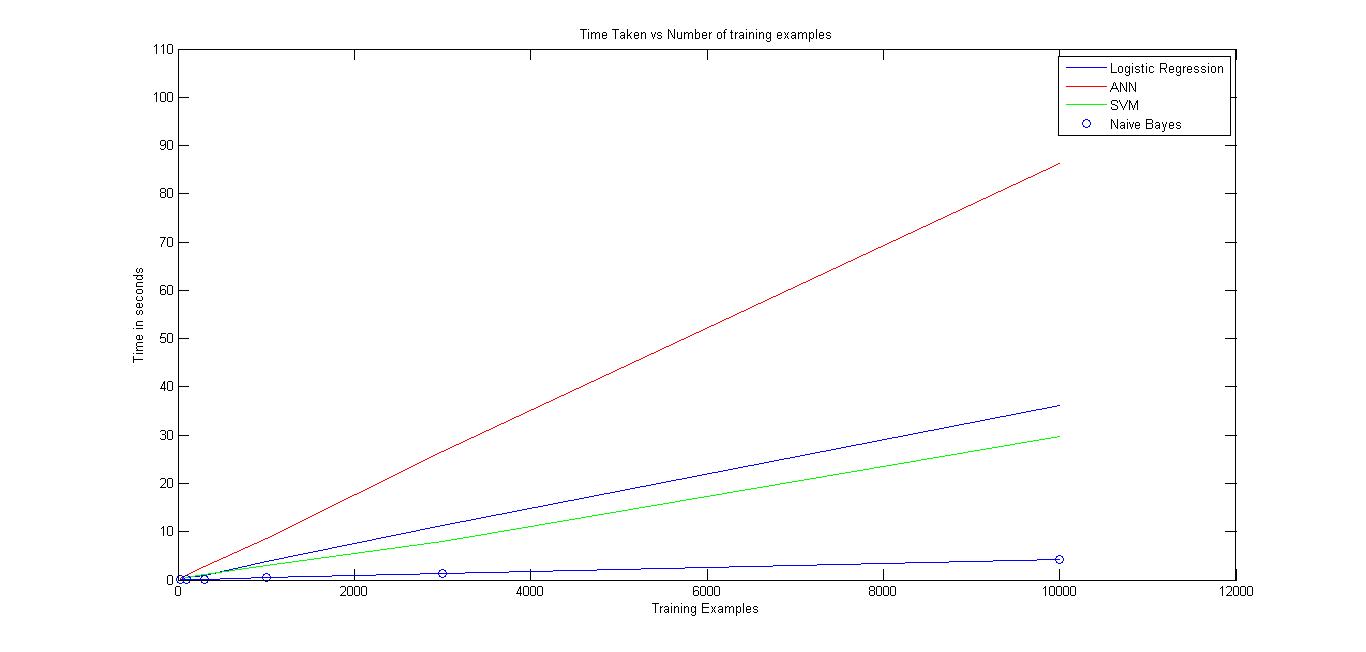
Since many approaches to the problem of Digit Recognition . It is important to compare these techniques .

First of all we compare the relative accuracy of these techniques with respect to each other .



As it can be see from the figure that maximum accuracy is achieved by SVM ,closely followed by ANN and then logistic regression and then Naïve Bayes .

Now we compare the time taken by different technques :



As it can be seen from the graph the training of Naïve Bayes is remotely fstest from others , folloed by SVm , then logistic regression and the slowest is Artificial Neural Network .

Conclusion :

To summarize the project , we tabulate the accuracy at best parameter for each technique when trained at 10000 training example .

|  |  |  |
| --- | --- | --- |
| Technique | Accuracy (%) | Time( in s) |
| Logistic Regression | 89.80 | 36.11 |
| Artificial Neural Network | 92.83 | 86.36 |
| Support Vector Machine | 93.40 | 29.71 |
| Naïve Bayes | 69.76 | 4.14 |

When Accuracy and Time both is taken into account it is pretty much evident that Support Vector machine perfroms the best for this given problem .

Source Code Used :

The source code loadImage/label is a function for reading the .idx file format and is taken from UFLD,Stanford resources .

Future Improvement :

References :

Database Used at : <http://yann.lecun.com/exdb/mnist/>

Others :

.