A project report on

CHARACTER RECOGNITION USING DEEP LEARNING NEURAL NETWORK

Submitted in partial fulfillment of the requirements for the award of degree of

Master of Technology

In

Information System

Submitted by:

Ankit Tiwari

(2K15/ISY/04)

Under the guidance of

Dr. O.P. Verma

(Professor, Department of Computer Science and Engineering, DTU)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

Bawana Road, Delhi-110042

Certificate

This is to certify that **Ankit Tiwari (2K15/ISY/04)** has carried out the major project entitled "***Character Recognition using Deep Learning Neural Network*** " in partial fulfillment of the requirement for the award of Master of Technology Degree in Information System during session 2015-2017 at Delhi Technical University.

The major project bonafide piece of work carried out and completed under my supervision and guidance. To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University/Institute for the award of any degree or diploma.

Dr. O.P. Verma

Professor

Department of Computer Science and Engineering

Delhi Technological University

Delhi – 110042

Acknowledgments

I express my sincere gratitude towards my project mentor **Dr. O. P. Verma**, Professor, Department of Computer Science and Engineering, Delhi Technological University, Delhi, for providing valuable guidance and constant encouragement throughout the project. It is my pleasure to record my sincere thanks to him for his constructive criticism and insight without which the project would not have shaped as it has.

I thank God for making all this possible, my parent and friends for their constant support and encouragement throughout the project work.

Ankit Tiwari

Roll No. 2K15/ISY/04

M.Tech (Information System)

Department of Computer Science and Engineering

Delhi Technological University

Abstract

OCR stands for Optical Character Recognition and is the mechanical or electronic translation of images consisting of text into the editable text. It is mostly used to convert handwritten(taken by scanner or by other means) into text. Human beings recognize many objects in this manner our eyes are the "optical mechanism." But while the brain "sees" the input, the ability to comprehend these signals varies in each person according to many factors. Digitization of text documents is often combined with the process of optical character recognition (OCR). Recognizing a character is a normal and easy work for human beings, but to make a machine or electronic device that does character recognition is a difficult task. Recognizing characters is one of those things which humans do better than the computer and other electronic devices.

Contents

List of Figures

Chapter 1

Introduction

In character recognition(also optical character reader, character recognition) the of images of handwritten ,typed, or printed text is converted into machine-encoded text by the means of mechanical or electronic conversion , whether from a scanned document, a image of a document, a scene-image (for example the text on signs and billboards in a landscape image) or from subtitle text superimposed on an image (for example from a television broadcast). It is generally utilized as a type of data section from printed paper information records, regardless of whether international ID reports, solicitations, bank articulations, automated receipts, business cards, mail, printouts of static-information, or any appropriate documentation.[1] It is a typical technique for describing printed messages with the goal that they can be electronically altered, looked, put away more minimally, showed on-line, and utilized as a part of machine procedures, for example, intellectual figuring, machine interpretation, (extricated) content to-discourse, key information and content mining. character acknowledgment is a field of research in design acknowledgment, counterfeit consciousness and PC vision.Character recognition.

There are the basic types of foundation Character recognition algorithm, that may produce a ranked list of candidate characters.manly divided in two types

Matrix matching involves comparing an image to a stored glyph on a pixel-by-pixel basis; it is also called as "pattern matching", "pattern recognition", or "image correlation". This depends on the information glyph being effectively segregated from whatever remains of the picture, and on the put away glyph being in a comparable text style and at a similar scale. This technique works best with typewritten text and does not work well when new fonts are encountered. This is the technique the early physical image cell-based character recognition implemented, rather directly.

Feature extraction techniqueturns glyphs into "features" like lines, closed loops, line intersectionsand line direction.[2] The extraction features diminishes the dimensionality of the representation and create the algorithm of recognition process computationally efficient. These features are compared with an abstract vector-like representation of a character, which might reduce to one or more glyph prototypes. General strategies of highlight location in PC vision are appropriate to this kind of character acknowledgment, which is normally observed in "intelligent" handwriting cknowledgment and surely most present day character acknowledgment programming.neighbors algorithm is used such as the k-nearest neighbors algorithm as Nearest neighbor classifiersto compare image features with stored glyph features and choose the nearest match.

projects, for example, Cuneiform and Tesseract utilize a two-pass way to deal with character acknowledgment. The second pass is known as "adaptive recognition" and utilizations the letter shapes perceived with high certainty on the primary go to perceive better the rest of the letters on the second pass. This is profitable for strange text styles or low-quality outputs where the textual style is contorted (e.g. obscured or blurred).

1.1 BACKGROUND

These days neural system and deep learning strategies is the hotly debated issue in machine learning field and artificial intelligence .Neural net is fundamentally a system
of perceptions to mimic the human sensory system . Each perceptron act like a neuron which take some information, applying some calculation over that approaching info and return yield to next neuron in next level.

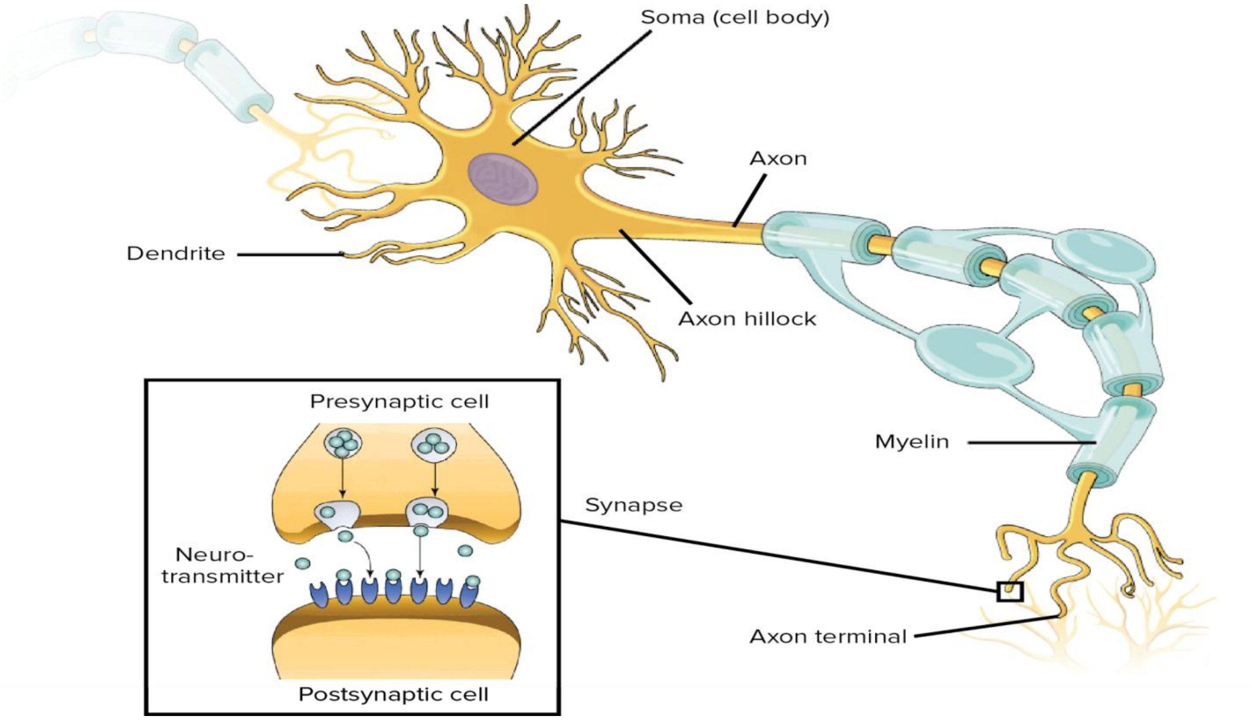


Fig 1: Neuron Representation

Artificial neural system emulates the conduct of human cerebrum. There is an assortment of neural system frameworks yet here we utilize two frameworks in this recommendation 1. Feed-forward Networks 2. Back propagation algorithm and convolution network (CNN). [3]

This portrayal has no less than three layers of neuron, each layer of neuron have loads of neuron. Every neuron does some calculation from the info and exchange the yield to next layer.

- Neural system framework comprise of info layer and one yield layer
- Neurons are associated with forward and in reverse layer neuron's to pass data
- Neurons in same layer are not associated with each other so at same level there is no correspondence and data passing happens.

Perceptrons

Fundamentally, perceptron is a regulated learning calculation at first intended for parallel yield frameworks, whose yield is the 0 or 1. The essential thought behind this calculation is, we have

a layered framework have various neurons at beginning and concealed layers yet we just single neuron on the yield layer. Which give the yield as 0
or 1. NNs were first introduced in 1957 at the Cornell Aeronautical Laboratory by Frank Rosenblatt.
This is scientific portrayal of neural model, assume we have a few data sources (x1, x2, x3, x4 ... x N) and having yield (y). To demonstrate the significance of each info Rosenblatt presented the idea of weighted system. So presented weights (w1, w2, w3, w4 ...wN). The yield created by this framework is either 0 or 1. This calculation decide yield, in light of whether the weighted entirety is 0 or 1.

$$output = \left\{ \begin{matrix} 1: & if\ w * x + b > 0 \\ \\ 0: & if\ w * x + b <= 0 \end{matrix} \right\}$$

This sort of neural framework is chiefly sent for paired order, Its is not reasonable for multi-class arrangement issue where input has a place with more than maybe a couple classes.[4] To devise this calculation for multi-class order we need to build the quantity of neurons in the yield layer. ANNs are just layers of Perceptrons which take some inputs and produce related outputs.

Simple & Logistic regression

Simple regression is widely used for continuous valued output, where you need to predict values for some inputted data. Suppose you have to find out the price of your house so in that case you need to devise an algorithm that can predict the price of your house. Suppose we have one attribute that is size of house, depend on this attribute we can formulate a function to define the house price sale.

So in view of qualities we have devise a theory that can foresee the house estimation on the off chance that we effectively anticipate the estimations of steady and parameter utilized as a part of this forecast .[5] This speculation can be straight or polynomial in view of the traits of the set. This speculation encourages into decide, that anticipated cost is how much nearer to real cost.
Logistic regression deals with classification problem. Where we need to anticipate the class of information to which it has a place. Assume we have a product and we need to anticipate that, this product is kind or harmful. So we can make two class one with threatening programming and other one to generous. Logistic hypothesis is

$$f(x) = \frac{1}{1+e^{-h_\theta(x)}}$$

Sigmoid Function

Perceptrons are constrained in degree. It demonstrates an entire flip flounder in results if any progressions made to weight related with system edges and bias. Perceptrons tuning is truly a troublesome assignment.[6] In the event that some way or another any progressions made to bias related with the layer than yield changes from 0 to 1 and the other way around. So this feature of Perceptrons completely changes the behavior of calculation.
To stop such changes in conduct we present sigmoid neuron. Sigmoid function is genuine esteemed, monotonic and differentiable non negative first derivative. The yield band is lies between 0 and 1. So to take care of classification we utilize sigmoid function.

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

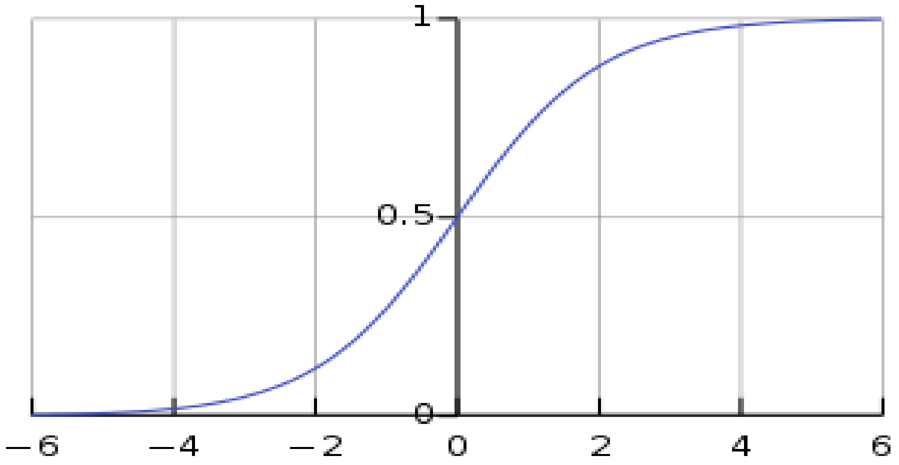


Fig 2: Sigmoid Curve

Hence there is only difference in activation function in perceptron and sigmoid neuron.

Loss Function

Loss function or cost function is a measuring apparatus, which demonstrates that how the anticipated yield is far from genuine yield. It demonstrates disparity in yield. So keeping in mind the end goal to compute the inconsistency we utilize MSE (mean squared error).

MSE can be characterized as:

$$|T(S,a) = \tfrac{1}{p}(\sum_{i=1}^{p} || \; h(x^i) - y^i \, ||^2)$$

Where :

- p is the total number of examples .
- x^i is the i^{th} training example .
- y^i is the actual output for specified set of values .
- $h(x^i)$ is the hypothesis value .

So our fundamental goal is to discover the disparity by limiting the squared mean squared error. We need to set the estimations of weights and inclination such that we can limit the MSE. To limit this we will use Gradient Descent algorithm.[7]

1.2 MOTIVATON FOR THESIS

Current ways to deal with protest acknowledgment make fundamental utilization of machine learning strategies. To move forward their execution, we can gather bigger datasets, take in more effective models, and utilize better methods for anticipating overfitting. Up to this point, datasets of named pictures were moderately little — on the request of a huge number of pictures (e.g., NORB , Caltech-101/256 [8, 9], andCIFAR-10/100 [10]). Basic acknowledgment undertakings can be fathomed great with datasets of this size, specifically on the off chance that they are increased with mark protecting changes. For instance, the currentbest mistake rate on the MNIST character-acknowledgment errand (<0.3%) approaches human execution .But protests in reasonable settings show extensive changeability, so to figure out how to remember them it is important to utilize significantly bigger preparing sets. Also, in reality, the deficiencies of little picture datasets have been broadly perceived (e.g., Pinto et al. [11]), yet it has just as of late turned out to be conceivable to gather marked datasets with a huge number of pictures. The new greater datasets consolidate LabelMe [12],whichcomprises of countless completely portioned pictures, and ImageNet [13], which comprises ofmore than 15 million named high-determination pictures in more than 22,000 classes.

To find out around a huge number of items from a huge number of pictures, we require a model with an extensive learninglimit. Be that as it may, the enormous unpredictability of the protest acknowledgment errand implies that this issuecan't be determined even by a dataset as huge as ImageNet, so our model ought to likewise have parcelsof earlier information to make up for every one of the information we don't have. Convolutional neural networks (CNNs) constitute one such class of models [16, 11, 13, 18, 15, 22, 26]. Their ability can be controlledby changing their profundity and expansiveness, and they additionally make solid and generally redress assumptionsabout the idea of pictures (in particular, stationarity of measurements and area of pixel dependencies).Thus, contrasted with standard feedforward neural systems with likewise estimated layers, CNNs have many less associations and parameters thus they are less demanding to prepare, while their hypothetically best execution is probably going to be just somewhat more awful.

1.2 THESIS OUTLINE

The main goal in the first part of the thesis is the purpose of which the proposed work is done according to purpose, which is done earlier in the field of multi-level sense analysis, thus the weakness is described which is present in the work presented earlier. . The next part includes the proposed method in the thesis, where we define the purpose of our system. We have a brief introduction about the neural network, while explaining the main concepts of neural networks and artificial neural networks. The next part gives a theoretical description of the phases of emotional analysis and then after the details of practical implementation. In the next part the results are obtained. It consists of three parts, the exactness received and the subsequent accuracy was obtained. The next part is compared with other algorithms already implemented. After all, we have concluded the thesis with the final conclusion and future part, which can be done in this.[6]

Chapter 2

Literature Survey

In 2016, (“Fang zhu”,lie ding,2016), another study was doneto better achieve character recognition,analyze the impact of noise character.BP neural networkapplication describes the process of character recognition,and the corresponding algorithm improvements.Createdwith MATLAB and training the neural network to identify the different samples, combined toolbox simulink simulation module, so that the character recognitionto get better recognition rate.In MATLAB, the ideal and noisy sample trainingnetwork is used to train the neural network with betterfaulttolerance.The correct recognition of the character recognition with certain noise ratio can be carried out.Through noise free and noisy input training network.Thefault tolerance performance of the two networks is compared by testing.At the same time, the Simulink module is composed of neural network, the same in the noise free samples as input to test, the correct recognitionrate of the test is 100%The comparison of the neuralnetwork and the neural network composed of the Simulink module through the MATLAB script program.Further understanding of the character recognition

In 2016, sumit verma and ritikaverma(NIT srinagar)workedon character recognition(character recognition),Their paper tells about character acknowledgment framework for disconnected transcribed character acknowledgment. The frameworks have can yield fantastic outcomes. In that there is the point by point discourse about manually written character perceive and incorporate different ideas included, and support additionally propels in the zone. The exact affirmation is particularly depending upon the possibility of the material to be scrutinized and by its quality. Pre - dealing with strategies used as a piece of report pictures as a basic walk in character affirmation structures were presented. The component extraction venture of optical character acknowledgment is the most imperative. It can be utilized with existing character acknowledgment techniques, particularly for English content. This framework offers an upper edge by having leverage i.e. its versatility, i.e. despite the fact that it is designed to peruse a predefined set of record positions, at present English archives, it can be arranged to perceive new sorts. Future research goes for new applications, for example, online character recognitionused in portable devices,extraction of content from video pictures, extraction of data from security reports and handling of verifiable records. Acknowledgment is regularly trailed by a post-handling stage. We trust and foresee that if post-get ready is done, the exactness will be extensively higher and a while later it could be particularly executed on mobile phones. Completing the gave structure post-planning on mobile phones waas also taken as a noteworthy part of thier future work.

In 2017(kianpeymani and mohsensoryani) ,worked on optical character recognition in parsi language disscussedin paper “From Machine Generated To Handwritten Character recognitionA Deep Learning Approach” of IEEE. While the task of Optical Character recognitionis deemed to be a solved problem in many languages, it still requires certain improvements in some languages with morecomplex script structures such as Farsi. Furthermore, Deep Convolution Neural Networks have reached excellent results in various computer vision tasks, including character recognition.Although, these networks require a great amount of data to be properly learned and (in some cases) lack generalization. Inorder to address this issue, in that work, they propose a tailored dataset and a delicately designed model that can be trained ononly machine-generated character images with various typefaces and not only achieve an excellent result on machine generated images, but also achieve a decent accuracy in detecting handwritten characters.Furthermore, it was seen that the model is adequatelycapable of correct detection in the domain of machine generated images and training it with handwritten images,which they have not done for the purpose of their main work,would indeed yield a decent accuracy way more than 68%.Another aspect is that that enhanced machine generated dataset owes its efficacy to the great capabilities ofconvolution layers. Neural networks are moving more andmore toward *End-to-End* approaches and one of the mainsteps toward this path is that feature extraction phase, whichusually limits the scope of the model and was a ponderous task, is now obviated using convolution neural networks.[16]

In 2011 (Sunil Kumar Khatri,Shivali Dutta and Prashant Johri) researched on recognizing images of Handwritten Characters using “Learning Vector Quantization Artificial Neural Network ”.The research is based on supervised learning vector quantization neural network categorized under artificial neural network. The images of characters are recognized, trained and tested. After the network is created characters are trained using training dataset vectors and testing is applied to the images of characters which are isolated to each other by segmenting the image and resizing the character image accordingly for better accuracy.LVQ is the feedforward network with connected weights to the inputs constituting two layers namely competitive and linear layer with competitive neurons connected to the linear neurons. Various challenges occur in recognizingthe handwritten characters due to different writing styles,shapes and size such problems are difficult to achieve butresearchers have applied different techniques to solvethese kind of issues and still cannot be overcomecompletely and further study has to be done in this field.

In 2012, (BalakrishnanGokulakrishnan, 2012) proposed a model for emotional classification where he used a publication on the Elephant Dataset, which contained 53000 tweets, containing about 41,000 positive tweets and 8500 negative tweets. Tweets were made for testing and then pre-processed to get critical figures of raw data. Then the different classification algorithms were functional in addition to the sense analysis. They were Kannai Breeze, Random Forest, Sequential Mining Adaptation (SMO) and Support Vector Machine (SVM) and were set accurately. It was seen that compared to SMO, SVM and Random One, according to the DMNB show, Kannai twenty two classifier was better.[6]

In 2013 (Guan and Rang, 2013) had planned a method for Chinese examination on Chinese microblog. The reason was to classify opinions in positive or negative in Microblogs. The required preprocessing steps for the method such as word segmentation and noise removal classification require the removal of specifications for every communication and self-training, TTOcActualAnulebell data was used. A method for semi-supervision is self training, where jointly labeled and unlocked data is used as training corpus. First of all, this training label starts with data; When the iterations are introduced, it is able to organize the unenilabeled data with the labeled data, compared to the supervision education algorithm, the self-training for classification of classification is not the best performance.

In 2015, (RuiJiayet., 2015) proposed a novel data expansion approach, which is called Dual Sense Test (DSA), which talks about the problem of division for emotion classification. The main reason behind this step was to create back-up reviews which are contrary to the original reviews of emotions and then using the pair of reversed and unique to train the spirit classifier (dual training); In order to predict further emotions (double prediction), he further clarified that in the case of polarization classification compared to other classifiers, the DSA algorithm supports the preparation of effective DSA and a selective data expansion technique, With high degree of degree required to expand the data along with a much more efficient performance. He has also worked on the expansion of the DSA to the DSA, which can deal with the 3-class (positive, negative, neutral) emotion classification. Finally, to overcome the dependence of the DSA on the External Antonyms dictionary, they suggested a corpus-based technique for the creation of a pseudo-Antonim dictionary.[4]

Chapter 3

The Proposed Work

3.1 Objectives and Goals

- Overview of MNIST data and Multinomial regressions.
- To recognize character creation of a function , based on traversing eachand every pixel image.
- To train the system to accurately identify characters use of TensorFlow. Model come across to thousands of examples of images to train itself.
- Train and enhance the system using **convolutional neural network and fully connected layers**
- Fully connected layers is used rather than CNN to reduce the complexion ,getting high accuracy in less time.
- Check the accuracy of the system with our testing data.

3.1 Dataset

MNIST is a simple collection of Machineimage dataset. The images in it is of handwritten charactersidentical these:

These images includes labels for each and every image,for identifying what character it is . here in above fig, the labels for the a images are [5, 0, 4, and 1].

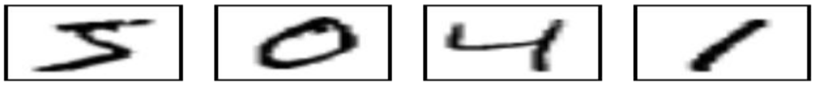


Fig 3: MNIST Image Representation

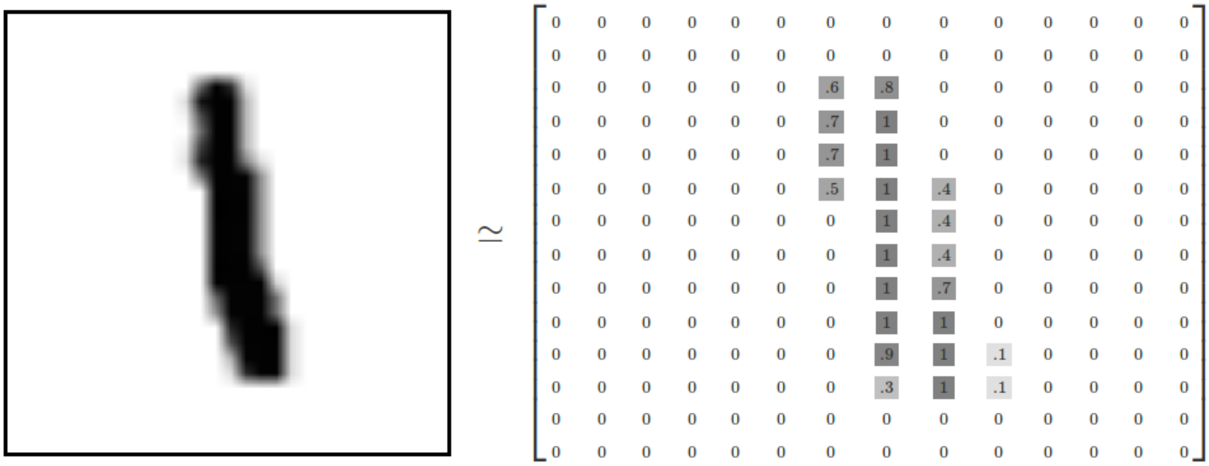
In this project, Model is being trained to look at images and guess what characters they are. Our purpose is to teach the model to achieve state-of-the-art performance -- I have made the program to do it later! –here , we will begin with an exceptionally basic model, called a Multinomial logisitic Regression

-

The MNIST information is part into three sections: 55,000 information purposes of preparing information (mnist.train), 10,000 purposes of test information (mnist.test), and 5,000 purposes of approval information (mnist.validation). This split is critical: it's fundamental in machine discovering that we have isolate information which we don't gain from with the goal that we can ensure that what we've realized really sums up!

As said before, each MNIST information point has two sections:a corresponding label and an handwritten character images and. We'll say the images "x" and the labels "y". Both the training set data and test set data contain images and their corresponding character labels; for instance the preparation pictures are train images if mnist that we have taken online and the preparation marks are trainlabels of train images character.

Each image is 28 pixels * 28 pixels. We can infer this as a large array of numerical:



$$\approx \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .6 & .8 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .7 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .7 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .5 & 1 & .4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & .4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & .4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & .7 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & .9 & 1 & .1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & .3 & 1 & .1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

Fig 4: Image Array Representation

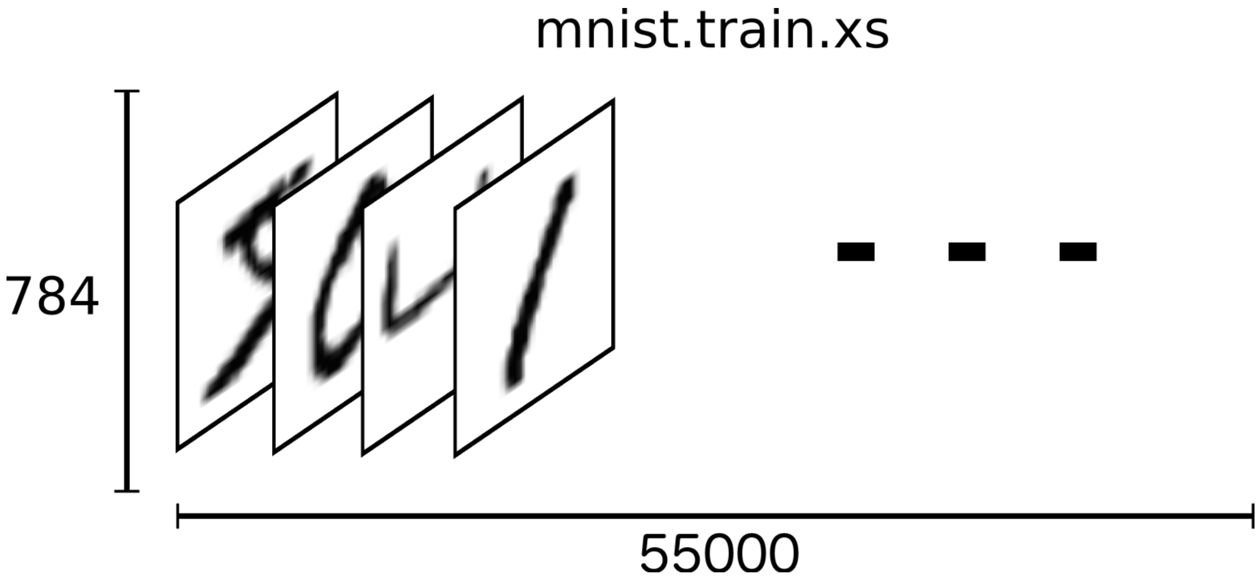


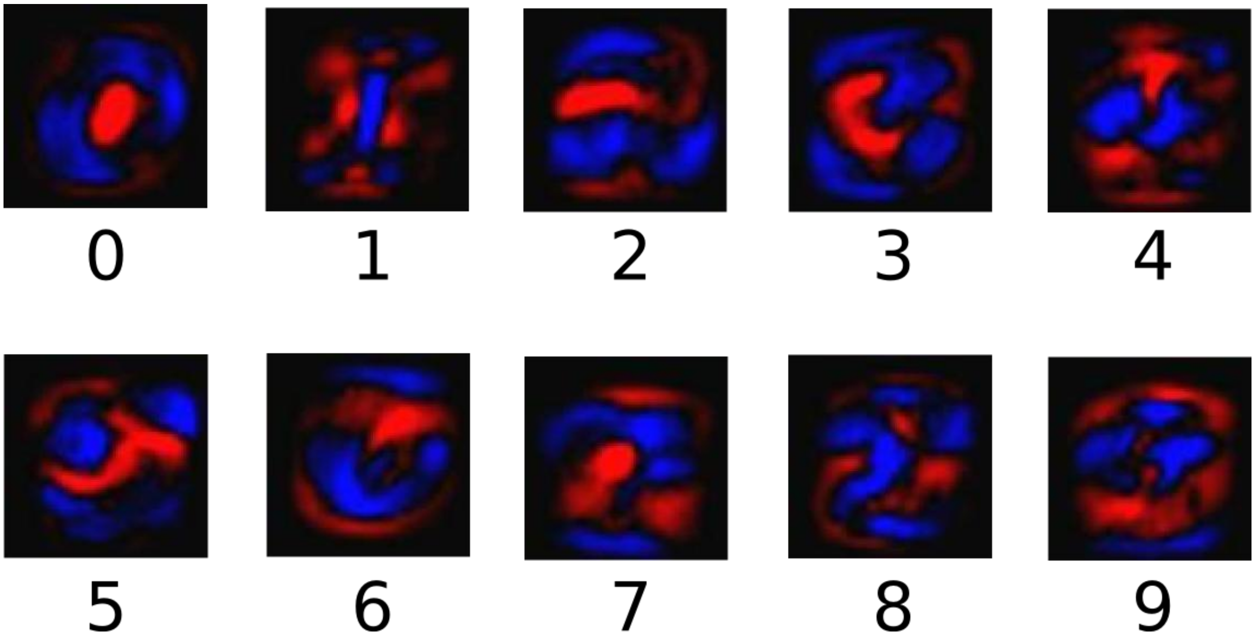
Fig 5: Vector Representation of image

3.2Multinomial Logistic Regression(Multinomial logistic)

Systemneed to be talented to look at an image and show the likelihoods for it being each character corresponding to the input image. For example, our System might look at a picture of a eight and be 85% sure it's a eight, but give a 10% chance to it being an nine (due to ambiguity of top loop) and it also matches with other character so we can never be 100% sure.

This is a standard situation where a multinomial calculated relapse is a consistent, honest model. In the event that we need to dole out probabilities to a protest being one of a few unique things, multinomial strategic is the thing to do, in light of the fact that multinomial calculated gives us a rundown of qualities in the vicinity of 0 and 1 that mean 1. Indeed, even later on, when we prepare more advanced models, the last stride will be a layer of multinomial calculated.

A multinomial strategic relapse has two stages: first we aggregate up the confirmation of our contribution for being in one of the many classes, and after that we change that proof into probabilities. To tally up the confirmation that a given picture is in a particular class, we do a weighted whole of the pixel powers. The weight is certain on the off chance that it is prove in support and negative if that pixel having a high power is confirm against the picture being in that class, andThe accompanying outline demonstrates the weights one model scholarly for each of these classes. Red speaks to negative weights, while blue speaks to positive weights.



Bias is some additional evidence added. Basically,we need to have the capacity to state that a few things are more probable free of the input information. evidence for a class r for input x in the result is:

(“evidence”= ∑jWr, jxj+br)

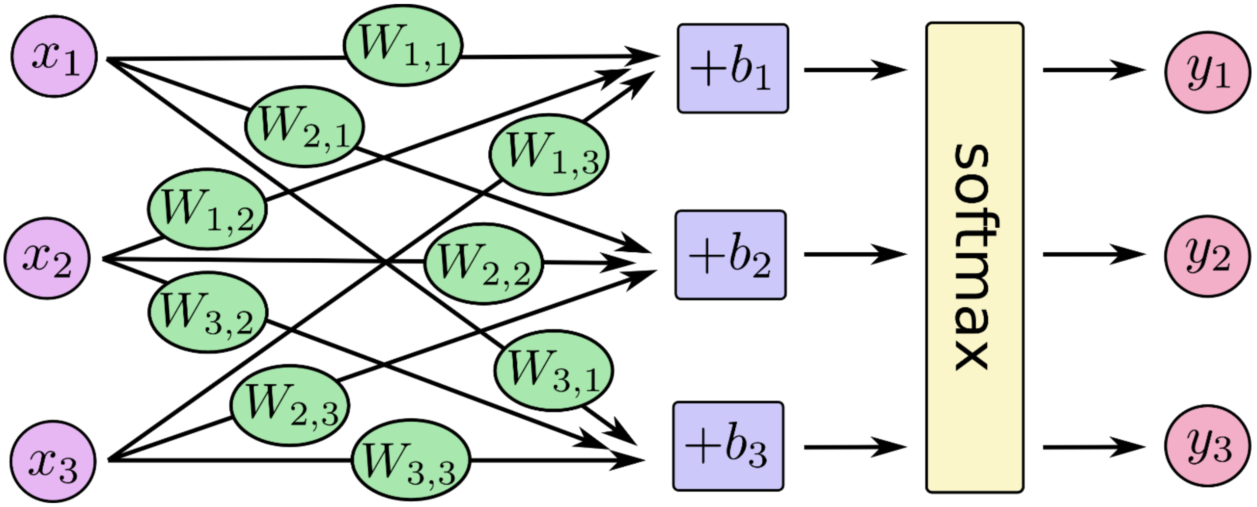


Fig 6: Softmax Regression

Corresponding equation is

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \text{softmax}\begin{pmatrix} W_{1,1}x_1 + W_{1,2}x_2 + W_{1,3}x_3 + b_1 \\ W_{2,1}x_1 + W_{2,2}x_2 + W_{2,3}x_3 + b_2 \\ W_{3,1}x_1 + W_{3,2}x_2 + W_{3,3}x_3 + b_3 \end{pmatrix}$$

Simplified expression is

y=multinomial logistic(Wx+b)

3.3 Training and Cross Entropy

With a specific end goal to prepare our model, we have to characterize what it implies for the model to be great. All things considered, really, in machine learning we regularly characterize what it implies for a model to be awful. We call this the cost, or the loss, and it speaks to how distant our model is from our coveted result.

We One extremely normal, exceptionally decent capacity to decide the departure of a model is called "cross-entropy." Cross-entropy emerges from contemplating data compacting codes in data hypothesis however it winds up being an imperative thought in loads of zones, from betting to machine learning. It's characterized as:

$Hy'(y)=-\sum iyi'\log(yi)$

Where y is our anticipated likelihood circulation, and y′ is the genuine conveyance (the one-hot vector with the character marks). In some unpleasant sense, the cross-entropy is measuring how wasteful our forecasts are for portraying reality. Comprehensively clarifying cross-entropy is past the degree of this wander, however it's well worthunderstanding.

3.4 Dropout

The task of prediction averaging is quite tedious for large neural networks. So to address this issue we use Dropout. In Dropout we usually off some of neurons from computation in hidden layer. With the end goal that these neurons can be kept from co-adjusting each other to much. On each feed-forward cycle, some of neurons present in hidden layer are separated from the system, Now they don't constitute as an individual from system. Separation of neurons happens just in hidden layer not in input and yield layers. So the sustain forward pass connected on changed system and after that data get back engendered on the same adjusted system. On every iteration weights and inclination are refreshed and dropout neurons are reestablished. For next iteration we need to discover the following concealed layer of neurons to be detached on.

As dropout technique in NN's expansion neuron independency on each other so it's a superior thing to learn powerful elements in seclusion rather in conjugation. It appears resembles that on every cycle we are managing diverse neural system. Now they don't constitute as an individual from system. Separation of neurons happens just in hidden layer not in input and yield layers. So the sustain forward pass connected on changed system and after that data get back engendered on the same adjusted system. On every iteration weights and inclination are refreshed and dropout neurons are reestablished So we need to average every single changed system and after that foresee the yield.

3.5 Model Architecture

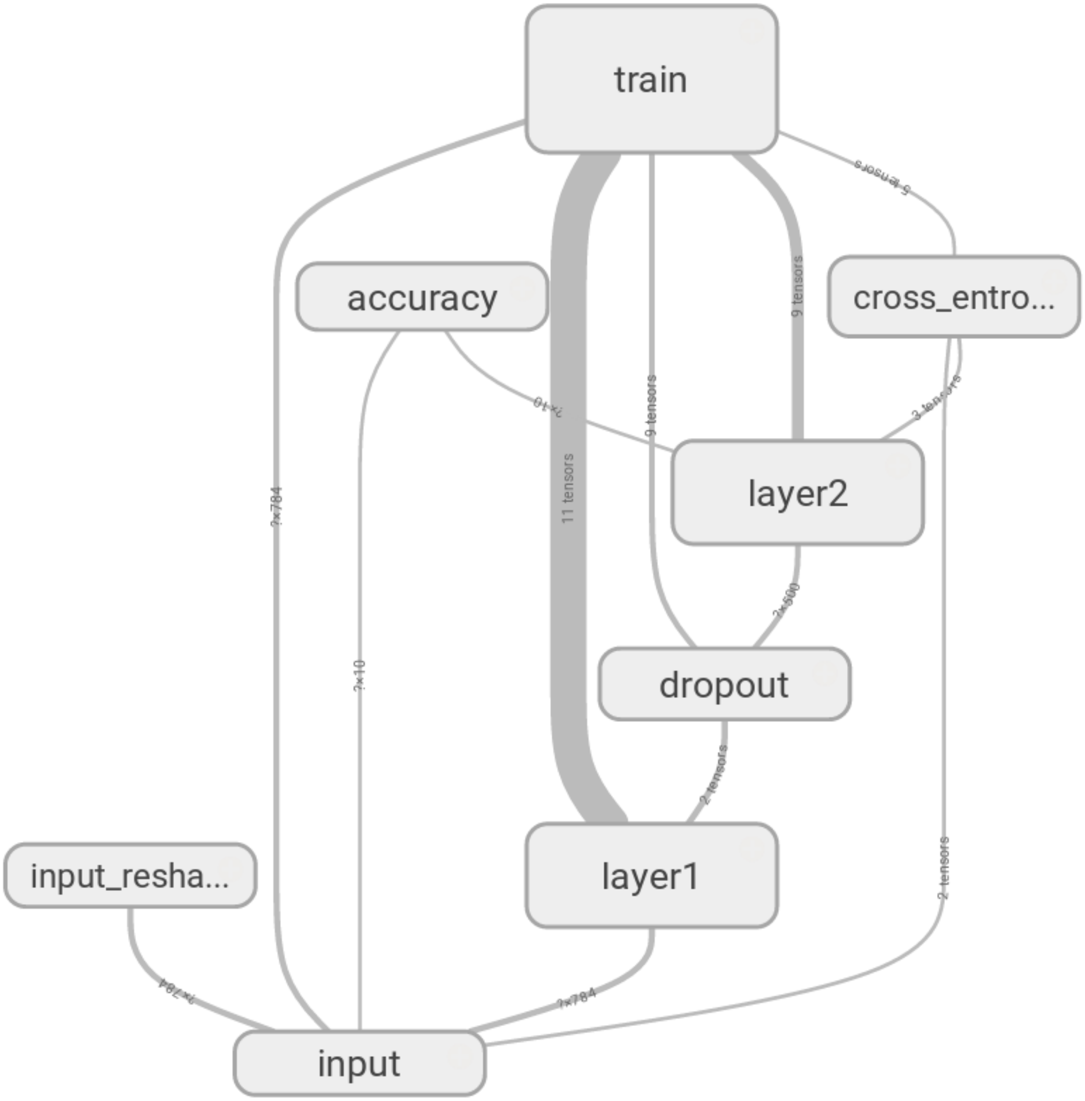


Fig 7: Model Architecture

I have implemented the model on python language

The main purpose of this method is to propose an effective way to filter out valuable information from images and to reduce the complexity of the system.First we have analyzed the system how accuracy varies with respect to iterations. For this purpose, In our system we have used cross entropy loss function and dropout regularization technique,in this technique we randomly dropout some weights, we have used dropout probability of .9 in training and testing for as expected there is as such no dropout. Figures in experminetal and result phase shows the graph which represent how loss function varied with respect to iterations as expected at the end of iteration count.

Chapter 4

Introduction to Deep Neural Networks

Deep learning (also known as deep structured learning o) is to present the knowledge of artificial neural network (ANN) to the counterparties, which keeps more than one hidden layer. [5]Unlike the deep learning task data algorithms, knowledge Based on data representations, the machine model is part of a enhance understanding of learning methods. Learning should be possible by directed learning or in part administered or Some portrayals depend on the translation of data handling and correspondence designs in an organic sensory system, for example, nerve coding, which endeavors to characterize the connection between different boosts and related neuronic responses in the cerebrum. Attempts to make skilled systems to learn, on a broad scale, without stamped information(label). The profound instruction design, for example, profound neural systems, profound confidence systems and intermittent neural systems have been executed in zones including PC vision, discourse acknowledgment, regular dialect handling, sound acknowledgment, interpersonal organization separating, machine interpretation and bioinformatics. Make similar examinations and at times superior to human specialists

Deep Education is a class of machine learning algorithms that:

• Use a cascade of multiple layers of non-liner processing units for feature extraction and change. Each layer uses the output from the previous layer as an input. Algorithms can be monitored or may be inappropriate and applications include pattern analysis (inappropriate) and classification (supervision).

• The properties or representations of different levels of data (uneducated) are based on learning. High level characteristics are derived from low level characteristics so that there is a hierarchical representation.

• There are some parts of the field of comprehensive machine learning for presenting data learning.

• Representing many levels according to different levels of separation; Level build hierarchies of concepts.[12]

In these definitions, there are several layers of general (1) non-liner processing units and (2) the information of the requested or unincorporated learning in each layer, with the layers, creates a hierarchy of low-level to high-level facilities.) [15]The composition depends on the problem of solving a layer of non-liner processing units used in a deep learning algorithm. The layers used in deep learning are Ritrim set of sources of hidden layers and complex presentation of neural network ... they deeply generating models such as endangered variable levels can also include such deep faith networks and nodes in Deep Boltzmann machines [17].

Profound Learning was first outlined and executed by World School Council London, which utilizes calculations to change its contribution through layers more than the shallow educating calculation. On each layer, the flag is supplanted by a handling unit Is, for example, manufactured neurons, whose parameters are balanced through preparing.

• Credit Assignment Path (CAP) - A progression of changes in yield from inputs portray the conceivable association between the CAP input and the yield association.

• Depth of the cap - For a feed forward neural network, the depth of the CAP (this type of network) is the number of hidden layers and one (also parameterized as the output layer), but for the recurrent neural network, in which one Indications can be promoted through one layer more than once, the CAP depth is probably unlimited[19].

• Deep / shallow-deep learning is not universally agreed on in-depth learning, but most of the field researchers agree that there are many nonlinear layers in deep education (CAP> 2). Schmidburg understands CAP> 10 to learn "too deep

4.1 Fundamental Concept

Distributed representations have the underlying assumption that observed data is generated by interaction between layered factors, deep education is the assumption that these layers of factors combine consistent levels of abstract or composition, variations of the layers and layer shapes differ in abstract It is this idea of a deep-education-oriented level, where lower level people should get higher levels and more deep academic architecture is often a greedy With a layer-by-layer method, you can take advantage of this, deep learning helps in solving these absorption and facilitating improvements in improving performance, whether giving birth or learning. Useful, deep learning methods feature Use engineering, [6] the data is compact like a compact intermediate main component. The level of structures that translate and receive while reducing the redundancy representation learning how to learn algorithms can be applied to learning unchanged learning tasks. This is an important advantage. Ecosis label data is more abundant than the examples of deep structures that can be trained in an unexpected way. Unlock from, these are the nerve Ihas compressors and the trust network.

4.2 Artificial neural network

Artificial neural networks or connection systems are systems induced by the biological nervous system that constitute the animal brain. Such systems (in order to improve successfully), with examples of working, usually learn without work-specific programming. Image recognition, they can learn to identify pictures in which the pictures are analyzed by the bill, allegedly, "cat" or "no cat" is said to have been used and analytical results are used to identify cats in other images. The use of traditional computer algorithm has been used more and more. ANN, the collection of connected units Between half neurons, each connection (synapse) can transmit a signal to the neuron, which is called artificial neurons (in the biological brain, the axis The receptor (postsynaptic) can process the neuron signal (PRO) and then point to the downstream neurons associated with it. The state may be in neurons, usually indicating the actual number, usually 0 Between 1 and 1 Neurons and surgery may also have a weight that differs in learning form, which can increase or decrease the signal strength, which sends it down. In addition, they may have a limit, only when the total signals are below (or above), the level downward signal is sent. In general, neurons are organized in layers, different layers can make different types of changes on their input. To take the signals from the first (input) to the last (output) layer, possibly after taking the layers several times, the basic goal of the approach of neural networks was to solve problems in the same way that a human brain is focused on time, specific Corresponds to mental abilities, thereby distracts from biology in biology, or goes backwards in the information direction and adjusts that information Ratibimbit to the network [5]

4.3 Deep Neural Networks

A deep neural network (DNN) is an ANN in which there are many hidden layers between input and output layers. Like the shallow ANN, DNN can model complex non-linear relationships. The DNN architecture prepares structural models where the object is expressed as a layered structure of preferences. Additional layers enable the combination of features from the following layers, potentially fewer units than those similar to shallow networks modeling complex data. Select multiple architectures, some basic approaches each Rkiteccr succeeded in specific domains. It is not always possible to compare the performance of much architecture, unless they are evaluated on a single data set. DN usually feeds the networks ahead, in which the data flow from the input layer to the output layer is looping back without looping back. However, the recurrent neural network, in which the data can move in any direction, is also used, particularly for long-term applications such as [language modeling]memory. [Convolution deep neural networks (CNN) are used in PC vision. Deep learning method of Convolution neural networks has also been used on acoustic modeling for automatic speech recognition .

4.4Convolutional Neural Network

In machine learning in, a convolutional neural system (CNN) is a class of profound, bolster forwardfeed simulated neural system that have effectively been connected to dissecting visual symbolism.

CNNs utilize a variety of multilayer perceptrons intended to require insignificant preprocessing.[1] They are otherwise called move invariant or space invariant manufactured neural systems (SIANN), in light of their mutual weights design and interpretation invariance characteristics.[2][3]

Convolutional systems were propelled by natural processes[4] in which the network design between neurons is roused by the association of the creature visual cortex. Individual cortical neurons react to jolts just in a limited area of the visual field known as the responsive field. The open fields of various neurons halfway cover to such an extent that they cover the whole visual field.

CNNs utilize moderately little pre-preparing contrasted with other picture characterization calculations. This implies the system takes in the channels that in conventional calculations were hand-designed. This autonomy from earlier information and human exertion in include configuration is a noteworthy favorable position.

4.5Gradient Descent optimization Algorithms

There are three variants of gradient descent algorithm which can be used to optimize the basic gradient descent algorithm.

- **Batch Gradient Descent Algorithm*:*** This algorithm calculate the gradient over the loss function T(S,a) for the entire training set in one go. It gives ensured convergence to convex surfaces over global minimum and local minimum for non raised surfaces.

- **Stochastic Gradient descent*:*** it is an incremental gradient descent which refreshes its parameters iteration by iteration .it is utilized to limit the target work that composed as total of differentiable capacities. This calculation tries to discover maxima or minima by iteration.This algorithm can be explained as: suppose we have an gradient to minimize w=w-Q_i(w) . So to minimize iterate it through a loop up to the function is not minimized.

- **Mini-Batch Gradient Descent Algorithm:** this algorithm appears to be same as batch gradient descent as the name suggest. Here we are trying to build up a batch of n (smaller set of total samples) training samples to learn the model. So the gradient will be calculated over the loss function of n training samples. This algorithm have better convergence rate than stochastic and batch gradient descent algorithm.

Next it is displayed a framework of a few calculations utilized as a part of profound figuring out how to upgrade the gradient descent algorithm.

4.6 Regularization

Regularization is the strategy to punish vast weights in systems and like to take in things from smaller weights. To actualize regularization we simply include regularization term in loss function. Regularization does not influence bias added to the system since huge bias make neurons less demanding to saturate which is alluring in specific conditions. Large bias doesn't make neurons more sensitive to data as by large weights.

Regularization has simply two levels:

- **L2 regularization**

$$L(S,a) = L(S,a)_0 + \frac{\lambda}{2n}\sum_{s} s^2$$

- **L1 regularization**

$$L(S,a) = L(S,a)_0 + \frac{\lambda}{n}\sum_{s} |s|$$

Here $L(S,a)_0$ is the original unregularized loss function.

4.7 ReLU Units

To solve the vanishing Gradient problems we are using rectified linear units (ReLU) introduced in 2010. Until then this problem is not solved. The activation function for ReLU units is defined as:

I

$$f(x) = \sum_{i=1}^{\infty} \sigma(x - i + 0.5) \approx log(1 + e^{x})$$

Where

- $\sum_{i=1}^{\infty} \sigma(x + i - 0.5)$ is **stepped sigmoid** function .
- $log(1 + e^{x})$ is **softplus** function

We can approximate this soft plus function soft plus function to **max function or hard-max** function as max(0, x+N(0,1)). It is also called as rectified linear function (**REL**).

So to stop the vanishing gradient problem we are using REL. REL is differ from sigmoid function in many aspects. Some of aspects are defined below.

1. Sigmoid capacity is utilized to display the likelihood as its range lies between [0,1] .though REL is intended to anticipate or display genuine esteemed numbers as its extents lies in range [0,] .
2. Angle of sigmoid capacity vanishes as we increment or lessening estimation of x though slope of REL doesn't vanish on expanding or diminishing the x esteem.
3. In the event that we utilize hard-max work as enactment capacity then it will actuate sparsely in shrouded layers.
4. REL doesn't require pre-preparing to prepare the neural systems.
5. ReLU can be utilized as a part of Limited Boltzmann machine to model genuine/whole number esteemed information sources.

There are various variants of ReLU namely noisy, leaky and ELUs. There are many problems associated with ReLU. Some of them are listed below:

1. This activation function is non differentiable at zero. It is differentiable anyplace, and furthermore differentiable near zero however not at equivalent to zero.
2. It is non-zero focused.
3. It could possibly explode in light of the fact that it is unbounded.
4. ***Dying ReLU problem*** : ReLU neurons here and there winds up noticeably dormant with the end goal that they don't pass data and reach in idle state to all sources of info and in reverse yields . These neurons are called as biting the dust or dead neurons. We have to

separate these neurons. This framework diminishes the model limit and this impact can be relieved by utilizing leaky ReLUs.

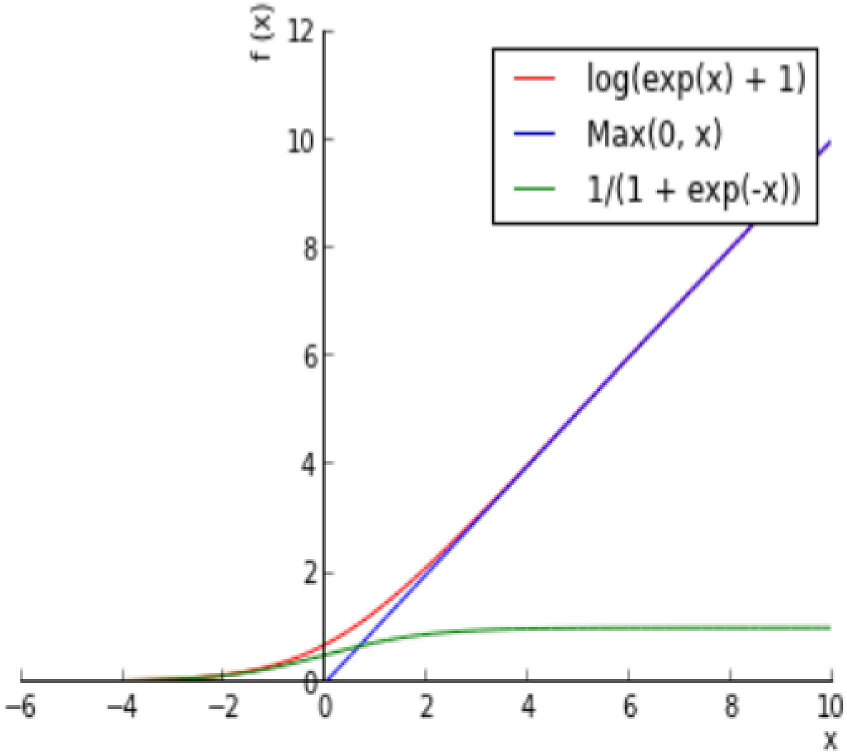


Fig 8 :REL and Sigmoid function Comparison

Here fig 2.4 shows the graphical comparison between ReLU and sigmoid function.

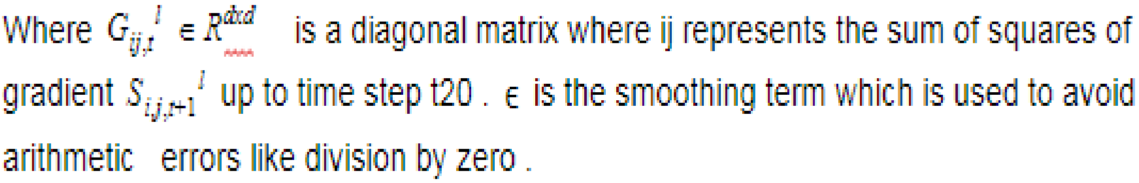
With the help of ReLU we have a lot of algorithm that provides optimum utilization of gradient descent algorithm. ReLU provides faster and effective learning for deep neural networks in comparison to sigmoid function on asymmetric and complex dataset.

4.8 ADAGRAD

This calculation used to improve the learning rate of slope plummet calculation. Its a versatile approach and used to refresh the learning rate on the premise that how this calculation carries on past cycles. So on premise of past emphasis information learning rate of angle plunge calculation can be refreshed.

At each time step t this calculation refreshes and change the weights of systems or you may state it will modify parameters Si,j,t l. it will refresh learning rate on every parameters and it will perform littler updates for visit parameters while rare parameters have bigger updates in light of the past slope esteems

$$\left[S_{i,j,t+1}{}^{l} = S_{i,j,t}{}^{l} - \frac{\alpha}{G_{t,ij}{}^{l} + \epsilon} * \frac{\partial}{\partial S_{i,j,t}{}^{l}} T(S,a) \right]$$



Where $G_{ij,t}{}^{l} \in R^{dxd}$ is a diagonal matrix where ij represents the sum of squares of gradient $S_{i,j,t+1}{}^{l}$ up to time step t20 . ϵ is the smoothing term which is used to avoid arithmetic errors like division by zero .

4.9 ADAM

It is also one of the adaptive techniques to optimize the gradient descent algorithm. The strategy is clear to actualize, is computationally effective, has little memory prerequisites, is invariant to corner to corner rescaling of the angles, and is appropriate for issues that are extensive as far as information or potentially parameters. a calculation for first-arrange inclination based advancement of stochastic target capacities, in view of versatile assessments of lower-request minutes. The technique is likewise fitting for non-stationary destinations and issues with exceptionally uproarious as well as inadequate angles.

The hyper-parameters have instinctive elucidations and commonly require small tuning. A few associations with related calculations, on which Adam was propelled, are talked about. We likewise break down the hypothetical merging properties of the calculation and give a lament bound on the joining rate that is equivalent to the best known outcomes under the online irably by and by and looks at positively to other stochastic advancement strategies. At last, we examine AdaMax, a variation of Adam in view of the limitlessness standard. Arched streamlining system. Observational outcomes show that Adam functions. It stores an exponentially decaying average of the past squared gradients that we will denote v_t and similar to momentum, it keeps an exponentially decaying average of past gradients $m_{t:}$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) * g_t v_t = \beta_2 v_{t-1} + (1 - \beta_2) * g_t^2$$

What’s more, g_t is the angle of the target capacity and β1 and β2 are the rot rates. M_t and v_t are the evaluations of the main minute or mean and the second minute or un centered difference of the inclinations, separately. M_t and v_t are instated as vectors of 0's and in consequence, during beginning time steps and when the rot rates are little they have a tendency to be one-sided towards 0. To take care of this issue, they figured the predisposition remedied first and second minute appraisals:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

As a result, the Adam update rule is defined as:

$$W_{t+1} = W_t - \frac{\alpha}{\sqrt{\hat{v}_t + \epsilon}} * \hat{m}_t$$

So these are certain tricks to optimize the basic gradient descent algorithm

Chapter 5

Results and Evaluation

Keeping in mind the end goal to prepare our model, we have to characterize what it implies for the model to be great. All things considered, really, in machine learning we normally characterize what it implies for a model to be awful. We call this the cost, or the misfortune, and it speaks to how distant our model is from our coveted result. We attempt to limit that blunder, and the littler the mistake edge, the better our model is.

5.1 Iteration Analysis and Accuracy Results

In our model we have found better accuracy as the number of iteration and less false positive results.

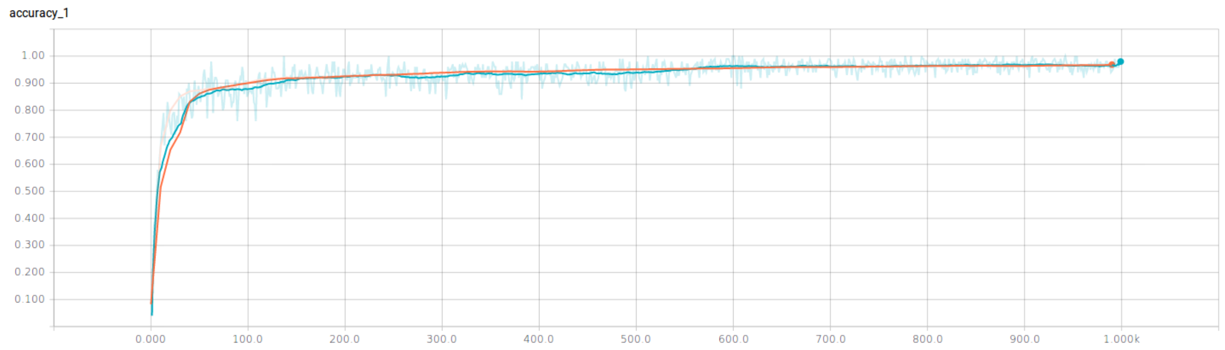


Fig 9: Accuracy VS Iterations

This figure captures how accuracy varies with respect to iterations, red line represent accuracy numbers on test dataset and blue represents accuracy on train datasets.

It can be seen that during training, there were fluctualtions as gradient descent was trying to locate local minima for the initial iterations but as iterations count increased
traing accuracy curve got stabilized.

130th iteration was the point where saturation occured from learning perspective after that there were improvemnts but quite small

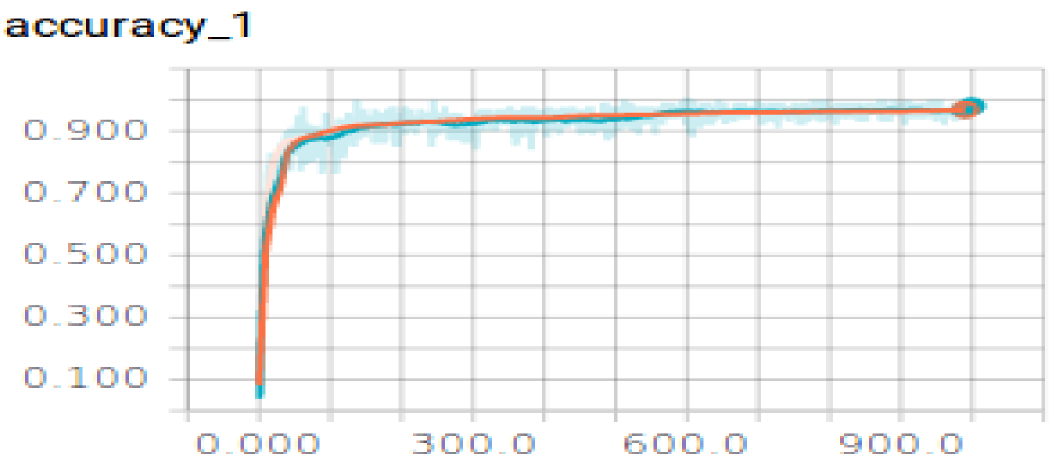


Fig 10 :Accuracy vs Iteration(2)

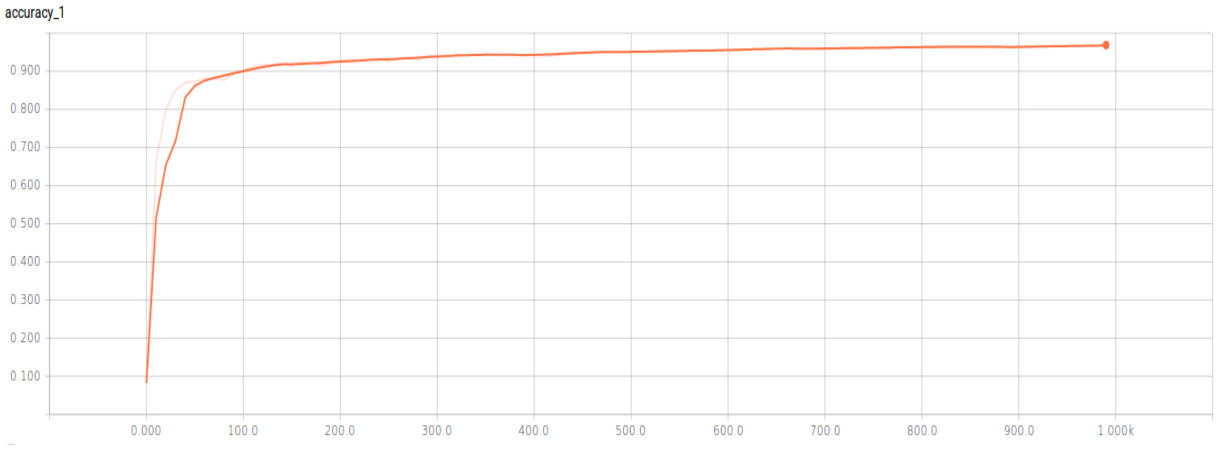


Fig 11 : Test data accuracy VS Iteration

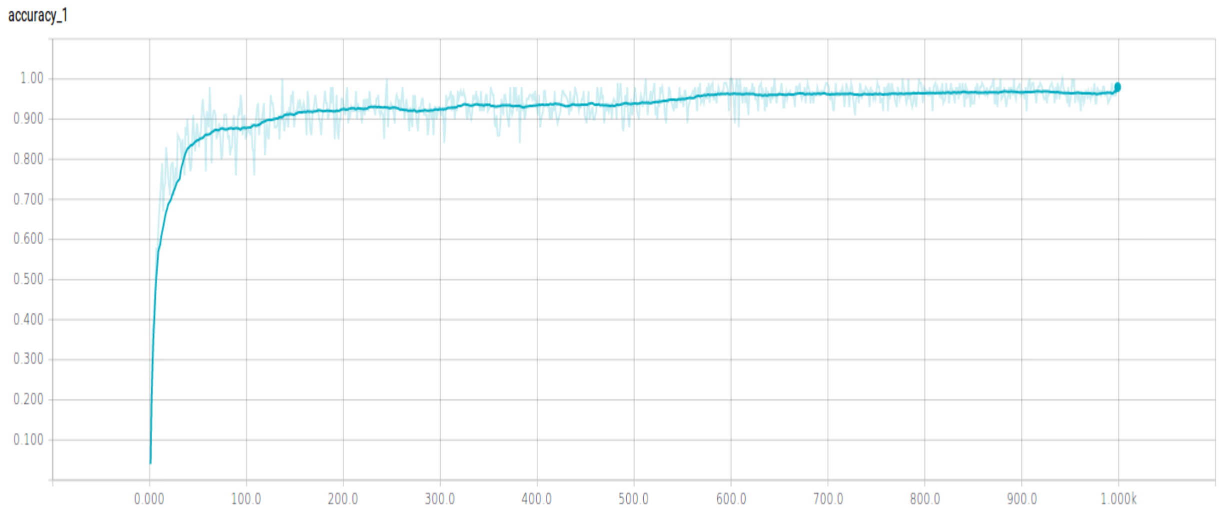


Fig 12:Train data Accuracy vs Iteration

5.2 Cross Entropy And dropout Results

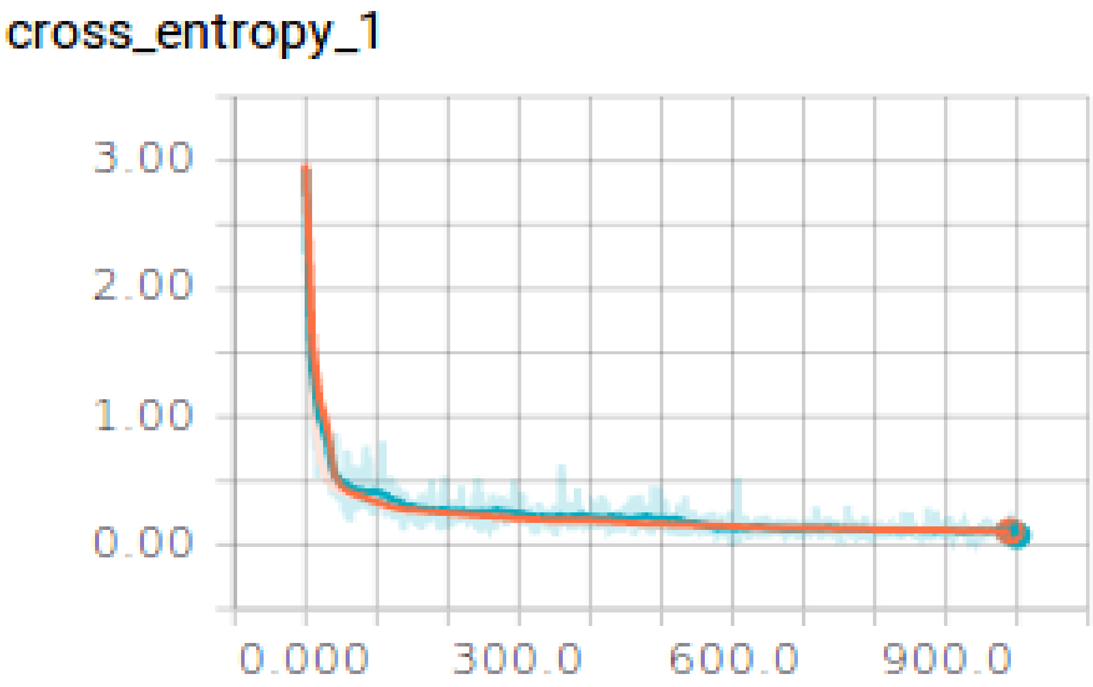


Fig 13 : Cross Entrophy

We have used cross entropy loss function , this graph represent how loss fuction varied with respect to iterations as expected at the end of the iteration count

error got minimized

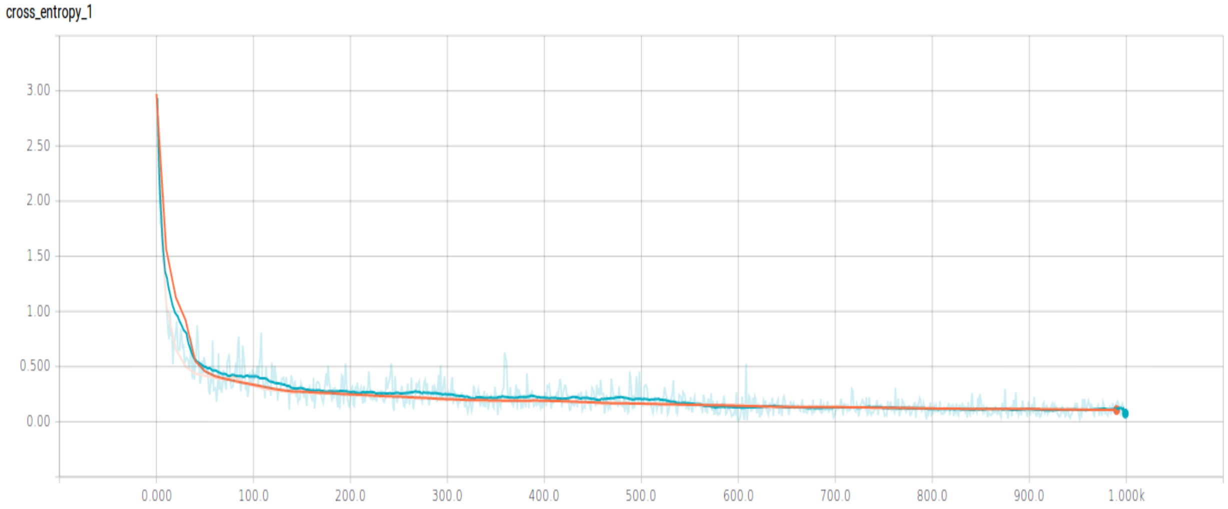


fig 14 : Cross Entrophy(1)

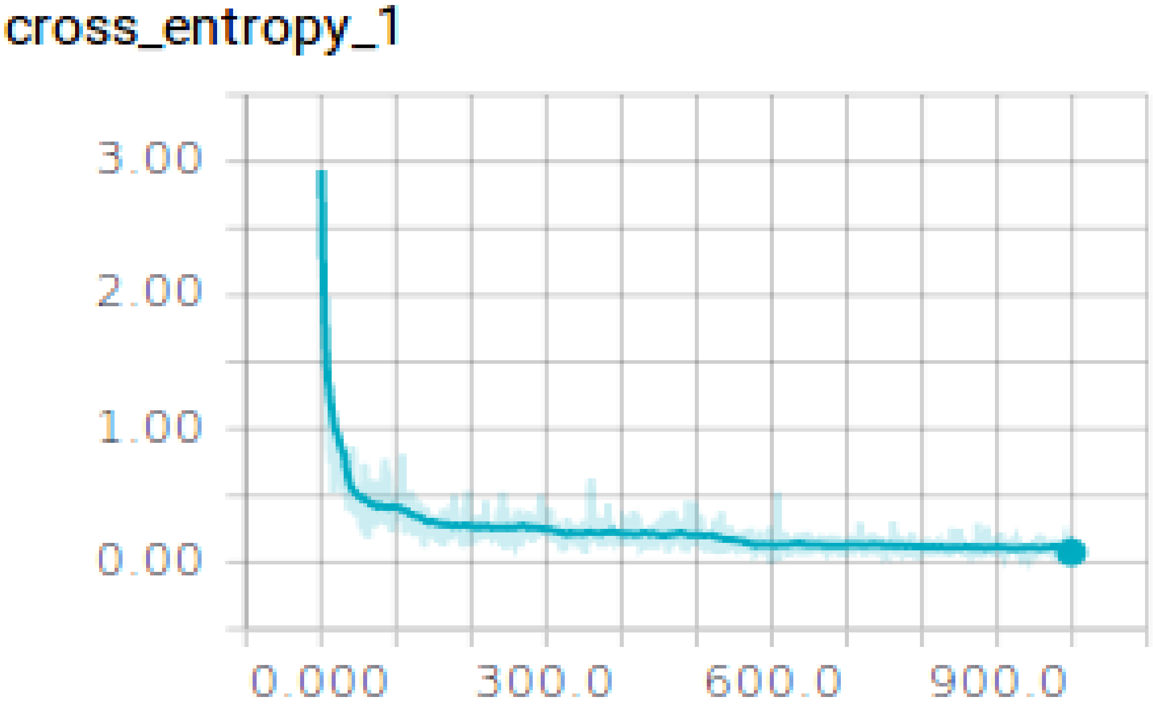


fig 15 : Variation of Crossentropy loss on training data

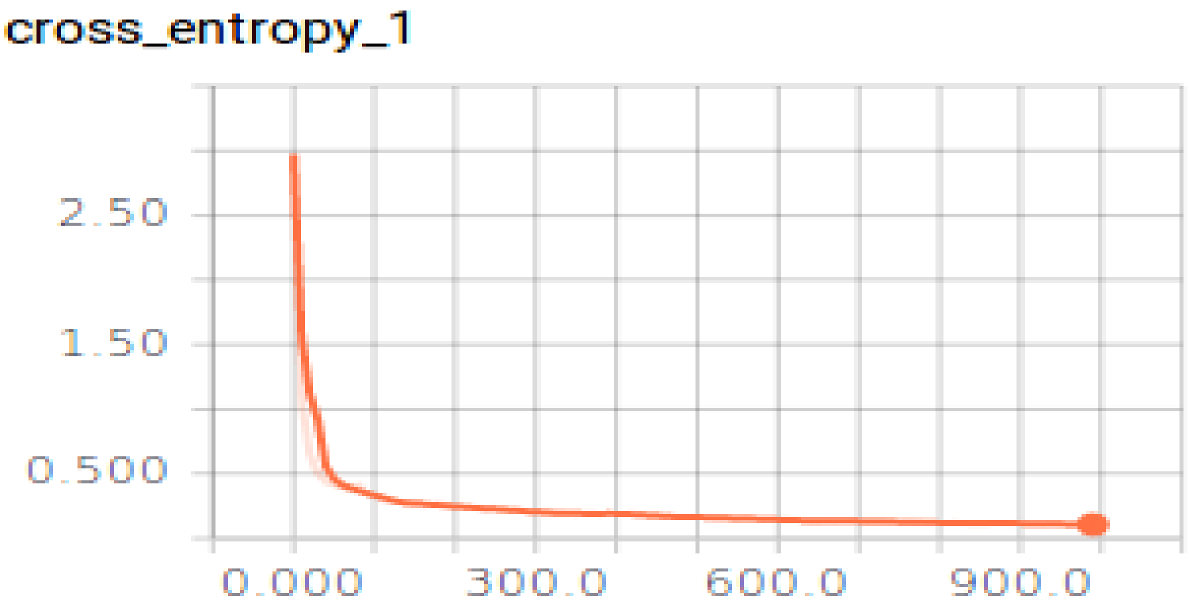


fig 16 : variation of crossentropy loss on test data

Dropout Deactivates each neuron probabilistically in both forwardand backward pass. This method is substantially important inincreasing generalization of the model

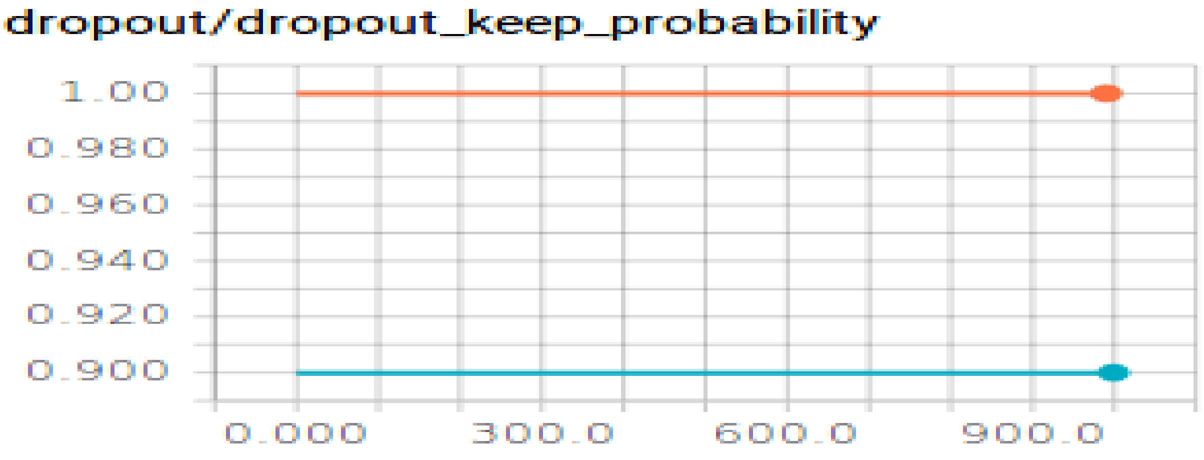


Fig17: Dropout/Dropout keep probability

dropout regularization technique , we have used dropout regularization technique in this we randomly drop out some weights , we have used dropout probability of .9 in training and for testing asexpected there is as such no dropout.

5.3 Statistics Parameter Results

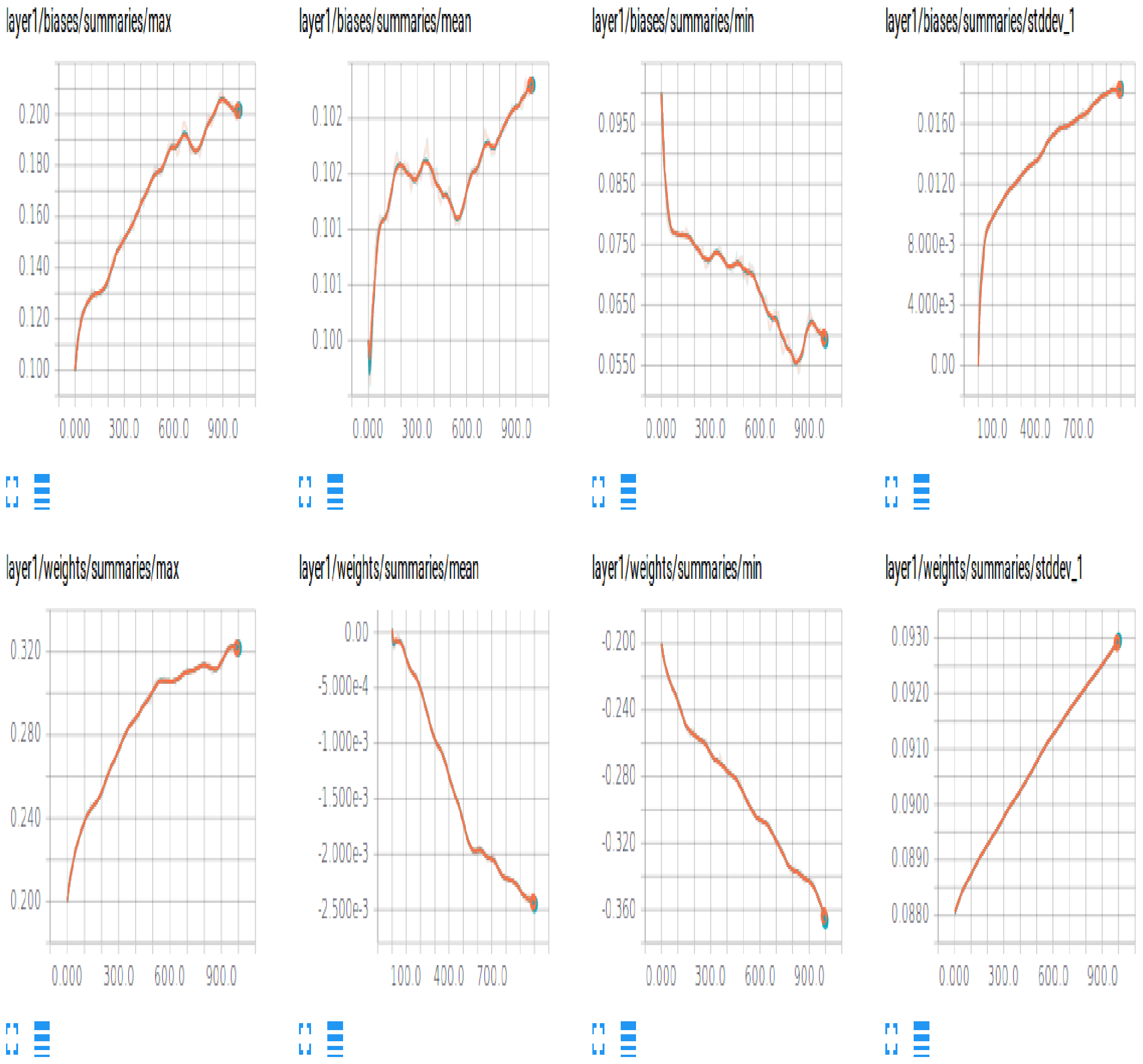


Fig 18 : statistics of parameters , which were fine tuned over iterations of layer 1.

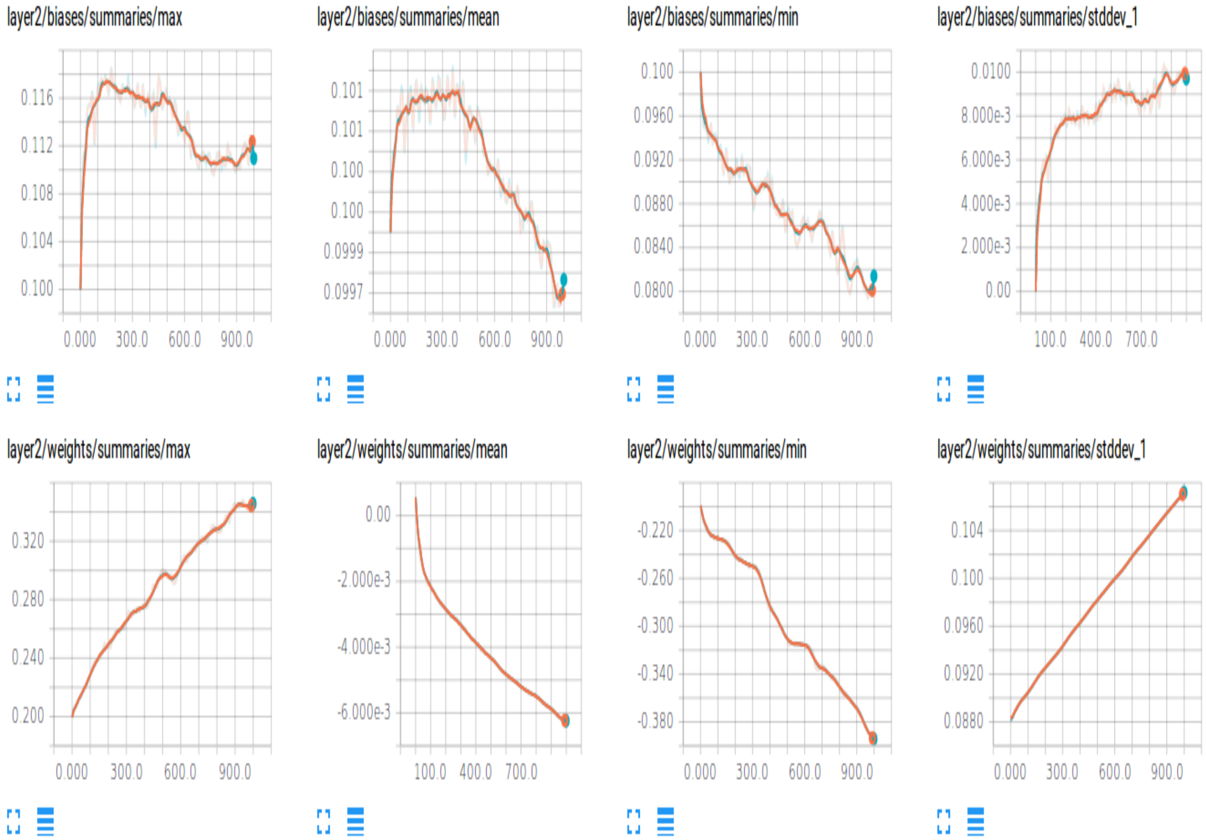


Fig 19 : statistics of parameters , which were fine tuned over iterations of layer 2.

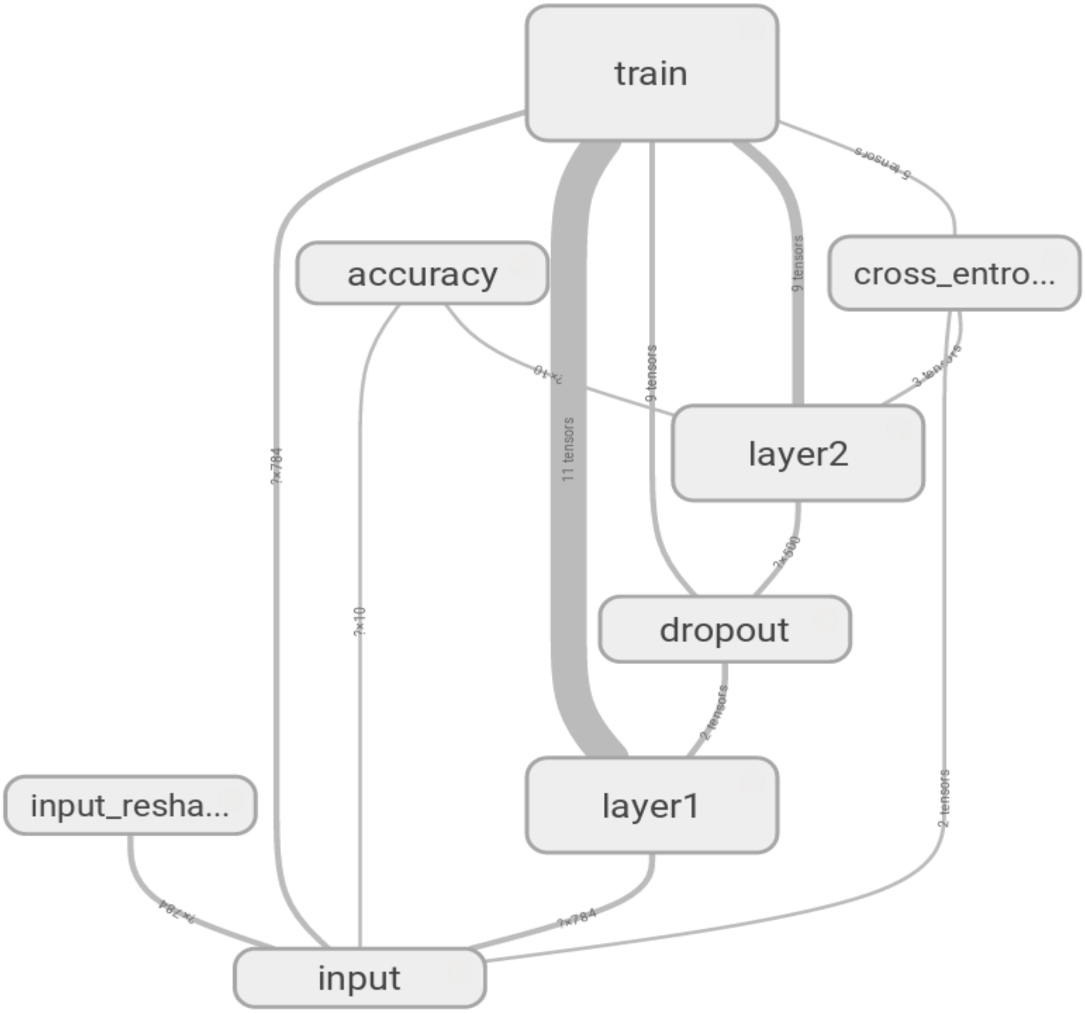


Fig 20 : Model Architecture

5.4 System Specification

Processor : Core *i*5 @2.27 GHz CPU

RAM : 4GB

Operating System : Windows 10

Python Version 2

Chapter 6

Conclusion And Future Scope

In this we have seen that the model is adequately capable of correct detection in the domain of machine generated images and also training it with handwritten images,which we have not done for the purpose of our main work,would indeed yield a decent accuracy way more than 70%. The data gathered from MNIST for training our model contains all the varieties of images that we may encounter in real life scenario our model attained the accuracy of >92% which can increase with high system capabilities.

Another viewpoint is this upgraded machine produced dataset owes its viability to the immense abilities of convolution layers. Neural systems are moving progressively and more toward End-to-End approaches and one of the primary ventures toward this way is highlight extraction stage, which normally restricts the extent of the model and was an unwieldy errand, is presently hindered utilizing convolution neural systems.

In our research, we have successfully implemented neural network algorithms to classify large amounts of datasets to address the topic. There are various challenges for image analysis first. In an attempt to combat this, we have used neural network algorithm to classify the image made by slangs, misspellings and achieved accuracy of 93.145%.

This paper tells about character recognition system for offline handwritten character recognition. The systems have the ability to yield excellent results. In this there is the detailed discussion about handwritten character recognize and include various concepts involved, and boost further advances in the area. The accuraterecognition is directly depending on the nature of the material to be read and by its quality. Pre-processing techniques used in document images as an initial step in character recognitionsystemswere presented. The feature extraction step of optical character recognitionis the most important. Itcan be used with existing CHARACTER RECOGNITIONmethods, especially for English text. This system offers an upper edgeby having an advantage i.e. its scalability, i.e. although it is configured to read a predefined set ofdocument formats, currently English documents, it can be configured to recognize new types. Futureresearch aims at new applications such as online character recognitionused in mobile devices,extraction of text from video images, extraction of information from security documents andprocessing of historical documents. Recognition is often followed by a post-processing stage. We

hope and foresee that if post-processing is done, the accuracy will be even higher and then it could bedirectly implemented on mobile devices. Implementing the presented system with post-processing onmobile devices is also taken as part of our future work.

References

[1] flying Si technology product research and development center. Neural network theory and [M]. to achieve MATLAB7 Electronics Industry Press, 2005:5-68

[2] Mao bin Tang, Xie Yuping, Li Qing. Based on neural network algorithm of character recognition method [J]. Microelectronics and computer 2009 (8)

[3] Ceng Zhijun, Sun Guoqiang, digital character recognition based on improved BP network [J]. Journal of University of Shanghai for Science and Technology, 2008,30 (2): 201-204

[4] Zhou Kaili, Kang Yaohong. The neural network model and MATLAB simulation program design of [M]. Tsinghua University press, 2004:4-25

[5] Chen Lei, Chen, Xing Rong Zhong, Wang Jiajun. Based on Improved BP algorithm for the digital character recognition [J]. Microelectronics and computer 2004 (12):(12):127-130

[6] Liu Hui, Yu Yanmei, Luo Daisheng. Momentum BP neural network character recognition based on English [J]. Journal of Sichuanb University, 2011 (6): 1324-05

[7] Chen Aibin, Lu Lina. Digital recognition of printed numerals based on multi features [J]. computing technology and automation, 2011 (3): 105-108

[8] Jia Shaorui, Li Lihong, Anqing bin.BP neural network algorithm in character recognition application [J]. science and technology information development and economy, 2007 (2): 167-170

[9] Peng Shumin, Wang Junning. An image recognition method based on neural network [J]. Electronic Science and technology, 2005 (01): 38-41

[10] Hou Shengwei, Teng Qizhi, Gao Mingliang, He Xiaohai. Automatic detection and recognition in the [J]. Journal of Sichuan University, 2013 (3): 522-528

[11]H Demuth,M Beal Neural network toolbox-for use with MATLAB,users guide,Version 5,The Math works 1998:24-51

[12] Yao Huijuan Luan Xiaoming Huijuan.Luan.Yao Xiaoming improved BP network algorithm in image recognition application [J]- Electronic Science and technology 2010,23 (9): 86-88

[13]Du Xian Nie, Zeng Wenqu. Random fractal dimension and wavelet transform of the license plate character recognition [J]. Journal of Guangdong University of technology, 2005,22 (4): 58-61

[14] Zheng Shenglin, Peng Mingming, Pan Baochang. A neural network character recognition method based on Hough transform [J]. Journal of Guangdong University of Technology, 2003, V20 (4): 73-77

[15] Chen Rui, Tang Yan, based on key words extraction of handwritten Chinese characters text dependent handwriting recognition technology [J]. Journal of Sichuan University, 2013 (4): 719-727