ARTIFICIAL NEURAL NETWORK

Classifying Hand-written Roman numerals through MLP[*]

Muhammad Bilal

Kiran Shahid

Sunday 12th December, 2021

Acknowledgement

I would like to express my special thanks of gratitude to my honorable teacher Dr. Junaid Akhtar who provided us a golden opportunity to perform this wonderful project and gain the experience of what ANN is? This helped me in doing a lot of Research and i came to know about so many new things. I am really thankful to them.Secondly, i would also like to thanks the TA Ali Raza Khan also my friends who helped me a lot in finalizing this project within the limited time frame.

[*] AI [CS304]

Abstract

The study examined and explored that how the machine can learn the concepts as a human brain can learn it. Neural networks often termed as Artificial Neural Networks (ANNs) are a part of algorithms that are similar to the structure and working of a single neuron. As part of Artificial Intelligence, neural networks incorporate technologies such as deep learning and machine learning. The important thing in this world that led us to the innovations and new development of technologies is the desire to make the life easier. If we talk about AI it has the both impacts on our lives. It has made it a way easier and difficult also. In Artificial neurons we have edges and weights and as the connection moves forward the adjust accordingly.

Table of Contents

Roadmap

- Problem Statement.
- Natural Phenomenon of Learning
- Iqbal's Theory of Learning
- Modeling
- Application

1 Problem Statement

A simple data set is given and we don't have to force the machine to cram all the data set. We just have to distinguish among the data whether a given set is good or bad. A single power of perceptron. We can calculate the actual output of the data by the power of the perceptron.

2 Natural Phenomenon of Learning

Learning is a creative process in human behaviour. All living is learning. Assuming we analyze the basic, rough manners by which a youngster feels and acts within an environment. Whereas if we compare it with the perplexing modes of adult behaviour, abilities, habits, feelings and opinions, we will come to know what improvement learning has made to that person. The individual is continually associating with and impacted by the environment. This experience makes him to change or adjust his behaviour to manage himself accordingly in order to survive in that environment. In this way, learning is about the

Figure 1: Social Learning through Environment

adjustment of behaviour, influenced by previous behaviour. Therefore the cognitive elements of human behaviour which contains abilities, knowledge, habits, opinions, interests and other personal characteristics are generally the consequence of learning.

3 Iqbal's Theory of Learning using Trial & Error Method

The method of trial and error was originally known in the field of psychology. In later developments, this method has become popular in various scientific fields, including education. This method is actually a combination of two definitions, namely trial and error. Since the term has been used as a common language in science, it has a common meaning which is used as a way to find the truth. Therefore, method of trial and error is the method of passing through various trials or errors to achieve success or achievement which can be experienced.

According to Iqbal, the life of the mortal ego passes in an environment full of obstacles. In order to free the ego from these obstacles, one must constantly master the science based on real experience through the educational path. The mortal ego is open to all possibilities. It can only grow significantly if it is taught using trial and error methods. Avoiding mistakes altogether is a mistake in itself. Because in addition to being completely inevitable, errors serve to confirm or help us n seeking the truth. Then only those who dare to face the possibility of wrong will finally find the truth. The method of trial and error teaches students the spirit of intellectual wandering which can reinforce experiences and gain real experiences. Furthermore, Iqbal warned readers that not to be satisfied with the knowledge one has acquired. This will result in intellectual boredom. One should advocate promoting freedom of thought to gain new insights and knowledge even though one had to deal with established and strong institutions and methods.

4 Modeling

Being a computer scientist it is our duty to find solution to our problem by studying a natural reality and the theory and the theory associated to it. After reading and analyzing the both section we build a model to solve the problem which is purely inspired from the theory. Here the problem which we want to solve is Classifying Hand-written Roman numerals through MLP. In machine learning, classification refers to a predictive modeling problem where a class label is predicted for a given example of input data. In order to build up our model we first discuss some key concept which help us in understating this model.

4.1 Human Brain Neuron & Perceptron

A human brain consist of billions of neuron in it. Each neuron has dendrites which takes input from other axons or neuron. Each neuron process the data received in the form of electrical signal and process on that data. After processing the result were then sent down to axon which in further will become the input of some other neuron. The network of neurons working in this way is known as Neural Network. A picture showing the structure of a neurons is as follow:

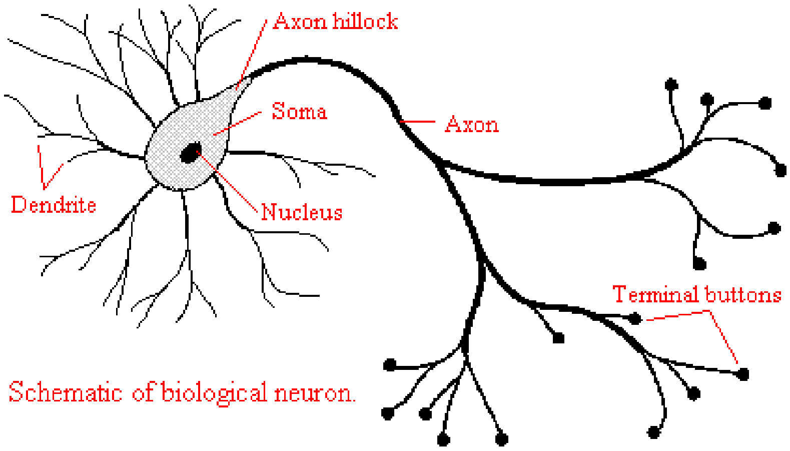

Figure 2: Graphical Model of Human Brain's Neuron

The idea behind the perceptron is that it is possible to replicate parts of neurons such as dendrites, cell bodies and axons using simple mathematical models to show their internal functions. With the limited knowledge we have, we can receive signals from dendrites, and after processing signals, the output will be sent down to axon which will be used as input to another perceptron or can also serve as an output.Artificial neural network could be constructed by repeating the process.

4.2 History of Neural Networks

The first Artificial Neural Network i.e. a single perceptron was invented by Frank Rosenblatt in the department of Naval research. After that many functional layers' neuron were being invented as the time proceeded. The famous names at that time were Kelley and Bryson, they used the concepts of Dynamic programming.

4.3 ANNs' Constituents

The important components that are in Artificial Neural Networks are as under:

4.3.1 Learning Capacity

The learning mechanism whether it is of Human or of artificial machine, it consists of two important parts:

- Hardware
- Software

In software we talk about the algorithms that we make during the process of the learning, while in hardware we talk about the neural networks. In humans we give the concept that how to differentiate between a cat a dog. Similar with the case of the Artificial Neural Networks we have to give this information to the perceptron. In case of perceptron there each neuron takes inputs and weighs them and computes them.

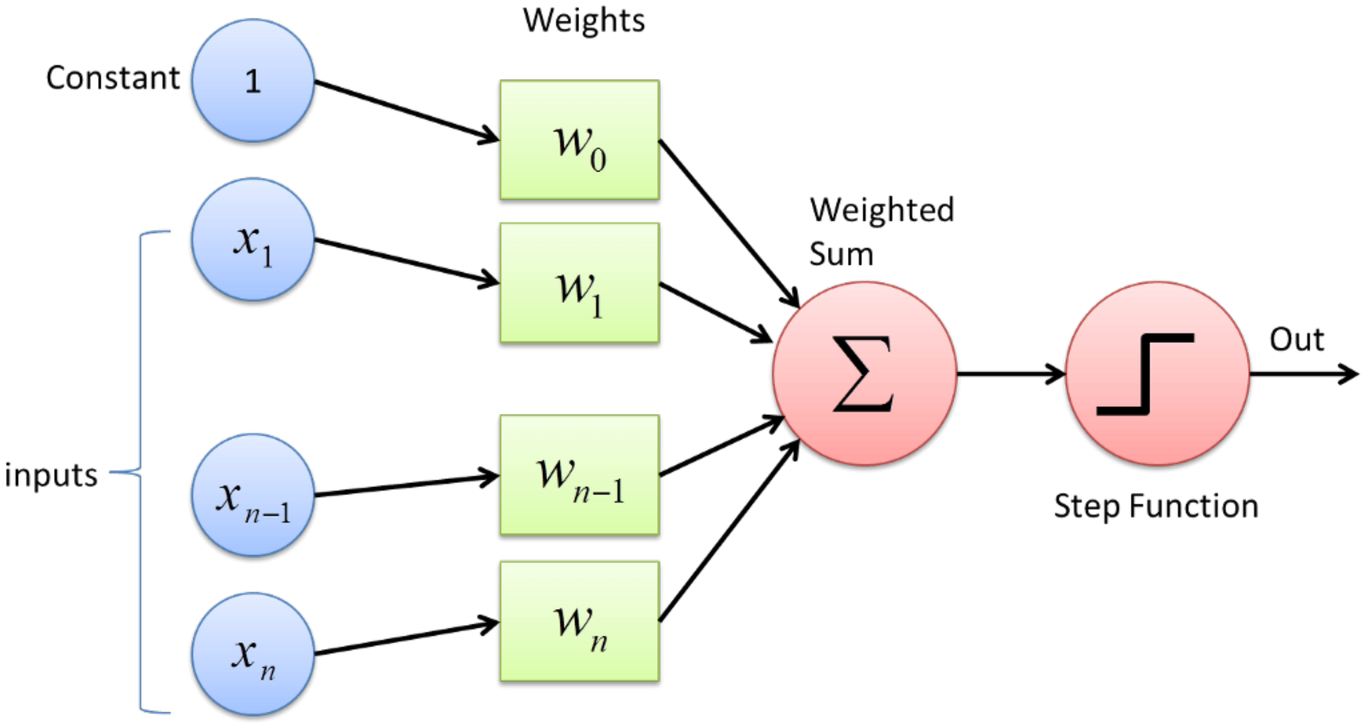

Figure 3: Activation Function using Mathematical Model of Perceptron

4.4 ANN Model

We use the idea of concept learning in terms of ANNs by learning from the information rather than cramming. In our model, there would be an input layer, one or more hidden layers, and an output layer make up the node layers of artificial neural networks (ANNs). Each artificial neuron, is connected to the others and has a weight and threshold assigned to it. If a node's output exceeds a certain threshold value, the node is activated and data is sent to the next tier of the network. Otherwise, no data is sent to the network's next tier.

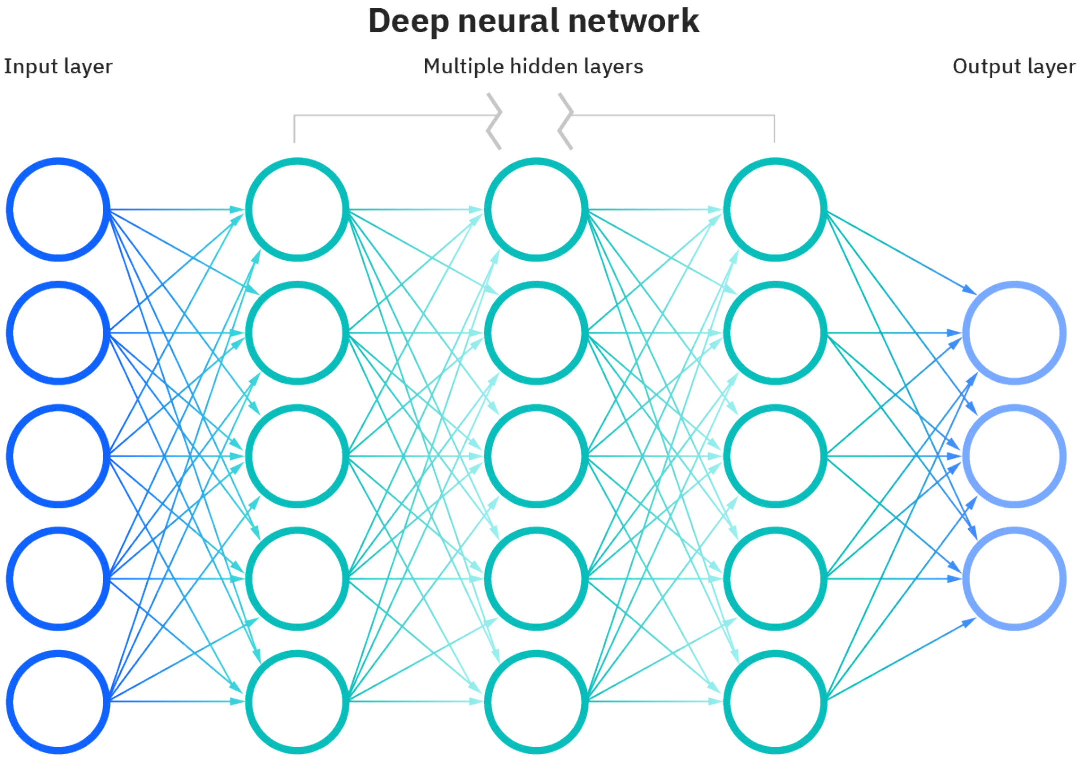

Figure 4: Artificial Neural Network Model

5 Application

In this section of report we will be discussing the actual problem which was given to us and describe the overall procedure how we tackle the overall problem and how we design the program using the model defined above which help us in solving the classification problem. We were giver a data set of Hand-Written Roman Numerals. We need to design a ANN which can classify the images to their label ranging from 'i' to 'x' in roman numerals. In order to implement this we are using a library named Sklearn.

5.1 Sklearn Library

Scikit-learn(Sklearn) is probably the most useful library for machine learning in Python. The sklearn library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction. In our project the tool we are going to use is MLPClassifier where MLP stands for Multi Layer Perceptron which is essentially used to build nueral networks. As our problem contain dataset which maps on ten classes ranging from 'i' to 'x' so we this classifier to classify objects of each class. Here is the brief description of MLP classifier defined by Sklearn.

```
class sklearn.neural_network.MLPClassifier(hidden_layer_sizes=(100),
activation='relu', *, solver='adam', alpha=0.0001, batch_size='auto',
learning_rate='constant', learning_rate_init=0.001, power_t=0.5,
max_iter=200, shuffle=True, random_state=None, tol=0.0001, verbose=False,
warm_start=False, momentum=0.9, nesterovs_momentum=True, early_stopping=False,
validation_fraction=0.1, beta_1=0.9, beta_2=0.999, epsilon=1e-08,
n_iter_no_change=10, max_fun=15000)
```

5.2 Major Functions of MLPClassfier

5.2.1 Fit(X,Y)

Fit the model to data matrix X and target(s) y. Here X is the input data and Y is the labels for classification. Whenever this function is called a trained MLP model will be returned.

5.2.2 Predict(X)

Predict using the multi-layer perceptron classifier. Where X is the input data on which the dataset is to be predicted. When this is called it will returns a matrix of label predicted against each input value.

5.3 Experimentation

We have a dataset of hundred of images of Roman numeral. So the first thing is to read all the images from all folder and store them in numpy arrays. For this purpose the OS library is used for traversing through all files and folders. Then after reading all the images and storing them in arrays, we try to figured out several feature from the images so that we can classify our image according to that features. Feature for training dataset are stored in train_x and features for validation dataset are stored in text_x while label for training dataset is in train_y and label for validation dataset is in test_y. And after extracting those features we perform several experiments.

5.3.1 Extracted Features

- **Dark Pixel:** In this feature we extracted the count of black pixels from each image.
- **Centroid:** In this feature we extracted the centroid of the image by cropping the original image using the center pixel. The size of the cropped image is 50*50.
- **Centroid with Slits**: In this feature we extracted some of the row of pixel from the centroid not the whole centroid.

- **Edge detection** In this feature we captures the edges of the original images and store it in another array and then used that image for classifying.

5.3.2 Experiment Type 1:

Feature used:

- Dark Pixel

5.3.2.1 Experiment 1 In our first experiment, the parametric setting is set to default and we try to trained our model using the count of dark pixels. Because each roman numeral is written in black pixel so it's count could be a feature to classify the dataset.

Hypothesis:

As this is our first experiment with new feature, so we could not judge what kind of output would be. But we are guessing that the dataset could be somehow classified using this feature.

Result:

Accuracy result using train_x in predict= 12.67%

Accuracy result using test_x in predict= 10.08%

Discussion:

The result are not that much good which means that either the feature we have extracted is not good enough to classify the dataset or the parametric setting is not correct.

5.3.2.2 Experiment 2 In our second experiment, we changed the parametric setting, we added the 3 hidden layers each having 50 perceptrons in it. And the rest of the setting is same.

Hypothesis:

By introducing extra hidden layer the accuracy must increase and the loss rate should decrease because each addition of perceptron means extra unit of classification.

Result:

Accuracy result using train_x in predict= 9.54%

Accuracy result using test_x in predict= 10.20%

Discussion:

The result are even worst than the previous ones which means the changing neurons doesn't make any effect on the extracted feature.

5.3.2.3 Experiment 3 In our third experiment, we again changed the parametric setting, we changed the activation function from 'relu' which is a unit-step function with 'logistic' which is sigmoid function.

Hypothesis:

By introducing this activation function there is a chance that accuracy of the model increases because sigmoid activation function is considered good for multi class problems.

Result:

Accuracy result using train_x in predict= 13.56%

Accuracy result using test_x in predict= 10.33%

Discussion:

The result are now somehow better than the previous one. So, our hypothesis is true. The result from he 'logistic' activation is better than 'relu' so we choose 'logistic' function for our next experiment. So, our hypothesis is wrong here.So lets change some other parameters and look whether we can classify our dataset with this feature or not.

5.3.2.4 Experiment 4 In our fourth experiment, we again changed the parametric setting, we changed the learning_rate_init from 0.001 to 0.0001.

Hypothesis:

The change in learning rate ensure the integrity of the model. If the jumps of learning is short and consistent then we can the output generated by this model also good enough as compared to model with high learning rate.

Result:

Accuracy result using train_x in predict= 13.56%

Accuracy result using test_x in predict= 10.33%

Discussion:

The result are the same as the previous one. So, we could not say anything about our hypothesis. But these experiments shows that the extracted feature is not good enough to make sure the classification of the dataset.

5.3.3 Experiment Type 2:

Feature used:

- Centroid

5.3.3.1 Experiment 1 In our first experiment, the parametric setting is set to default and we try to trained our model using centroid. Because each roman numeral is written in almost the middle of the image so it could be a feature to classify the dataset.

Hypothesis:

As this is our first experiment with new feature, so we could not judge what kind of output would be. But we are guessing that the dataset could be somehow classified using this feature.

Result:

Accuracy result using train_x in predict= 42.47%

Accuracy result using test_x in predict= 16.48%

Discussion:

The result are much better than the result in our previous experiment type where we used only dark pixels of each image.

5.3.3.2 Experiment 2 In our second experiment, we changed the parametric setting, we added the 3 hidden layers each having 50 perceptrons in it. And the rest of the setting is same.

Hypothesis:

By introducing extra hidden layer the accuracy must increase and the loss rate should decrease because each addition of perceptron means extra unit of classification.

Result:

Accuracy result using train_x in predict= 59.94%

Accuracy result using test_x in predict= 20.91%

Discussion:

The result are much better than the result in our previous experiment. Which means that our hypothesis is true that by increasing the number of perceptrons a better model can be generated.

5.3.3.3 Experiment 3 In our third experiment, we again changed the parametric setting, we changed the activation function from 'relu' which is a unit-step function with 'logistic' which is sigmoid function.

Hypothesis:

By introducing this activation function there is a chance that accuracy of the model increases because sigmoid activation function is considered good for multi class problems.

Result:

Accuracy result using train_x in predict= 46.97%

Accuracy result using test_x in predict= 16.23%

Discussion:

The result are not that much better than the previous one even they are little bit less. The result from he 'relu' activation is better than that so we choose that function for our next experiment. So, our hypothesis is wrong here. So lets change some other parameters and look whether we can classify our dataset with this feature or not.

5.3.3.4 Experiment 4 In our fourth experiment, we again changed the parametric setting, we changed the learning_rate_init from 0.001 to 0.0001.

Hypothesis:

The change in learning rate ensure the integrity of the model. If the jumps of learning is short and consistent then we can the output generated by this model also good enough as compared to model with high learning rate.

Result:

Accuracy result using train_x in predict= 61.78%

Accuracy result using test_x in predict= 21.64%

Discussion: The result are much much better the previous one. So, our hypothesis is correct. We will further explore and test some feature which can be used to make sure the classification of the dataset.

5.3.4 Experiment Type 3:

Feature used:

- Centroid with Slit

5.3.4.1 Experiment 1 In our first experiment, the parametric setting is set to default and we try to trained our model using centroid. Because each roman numeral is written in almost the middle of the image so instead of using the whole image we used 5 slits from the 50*50 matrix image and now this could be a feature to classify the dataset.

Hypothesis:

As this is our first experiment with new feature, so we could not judge what kind of output would be. But we are guessing that the dataset could be somehow classified using this feature.

Result:

Accuracy result using train_x in predict= 74.84%

Accuracy result using test_x in predict= 23.98%

Discussion:

The result are much better than the result in our previous experiment type where we used the whole centroid of each image.

5.3.4.2 Experiment 2 In our second experiment, we changed the parametric setting, we added the 3 hidden layers each having 50 perceptrons in it. And the rest of the setting is same.

Hypothesis:

By introducing extra hidden layer the accuracy must increase and the loss rate should decrease because each addition of perceptron means extra unit of classification.

Result:

Accuracy result using train_x in predict= 79.24%

Accuracy result using test_x in predict= 22.26%

Discussion:

The result are much better than the result in our previous experiment. Which means that our hypothesis is true that by increasing the number of perceptrons a better model can be generated. and now this time the centriod are also about the much less than the previous centroid.

5.3.4.3 Experiment 3 In our third experiment, we again changed the parametric setting, we changed the activation function from 'relu' which is a unit-step function with 'logistic' which is sigmoid function.

Hypothesis:

By introducing this activation function there is a chance that accuracy of the model increases because sigmoid activation function is considered good for multi class problems.

Result:

Accuracy result using train_x in predict= 34.01%

Accuracy result using test_x in predict= 16.97%

Discussion:

The result are not that much better than the previous one even they are little bit less. The result from he 'relu' activation is better than that so we choose that function for our next experiment. So, our hypothesis is wrong here. So lets change some other parameters and look whether we can classify our dataset with this feature or not.

5.3.4.4 Experiment 4 In our fourth experiment, we again changed the parametric setting, we changed the learning_rate_init from 0.001 to 0.0001.

Hypothesis:

The change in learning rate ensure the integrity of the model. If the jumps of learning is short and consistent then we can the output generated by this model also good enough as compared to model with high learning rate.

Result:

Accuracy result using train_x in predict= 51.57%

Accuracy result using test_x in predict= 23.12%

Discussion: The result are much much better the previous one. So, our hypothesis is correct. We will further explore and test some feature which can be used to make sure the classification of the dataset.

5.3.5 Experiment Type 4:

Feature used:

- Centroid of Edge Detected Image with Slit

5.3.5.1 Experiment 1 In our first experiment, the parametric setting is set to default and we try to trained our model using centroid. Because each roman numeral is written in almost the middle of the image so instead of using the whole image we used 5 slits from the 50*50 matrix image and this time the image only contains the edges of the roman numerals now this could be a feature to classify the dataset.

Hypothesis:

As this is our first experiment with new feature, so we could not judge what kind of output would be. But we are guessing that the dataset could be somehow classified using this feature.

Result:

Accuracy result using train_x in predict= 93.90%

Accuracy result using test_x in predict= 17.09%

Discussion:

The result are much better than the result in our all previous experiment types where we used the whole centroid of each image or even with slits on centroid.

5.3.5.2 Experiment 2 In our second experiment, we changed the parametric setting, we added the 3 hidden layers each having 50 perceptrons in it. And the rest of the setting is same.

Hypothesis:

By introducing extra hidden layer the accuracy must increase and the loss rate should decrease because each addition of perceptron means extra unit of classification.

Result:

Accuracy result using train_x in predict= 93.90%

Accuracy result using test_x in predict= 18.57%

Discussion:

The result are somehow better than the result in our previous experiment. Which means that our hypothesis is true that by increasing the number of perceptrons a better model can be generated. But the change is only in validation images predication the training image prediction is still the same as previous one.

5.3.5.3 Experiment 3 In our third experiment, we again changed the parametric setting, we changed the activation function from 'relu' which is a unit-step function with 'logistic' which is sigmoid function.

Hypothesis:

By introducing this activation function there is a chance that accuracy of the model increases because sigmoid activation function is considered good for multi class problems.

Result:

Accuracy result using train_x in predict= 52.29%

Accuracy result using test_x in predict= 15.00%

Discussion:

The result are not that much better than the previous one even they are little bit less. The result from he 'relu' activation is better than that so we choose that function for our next experiment. So, our hypothesis is wrong here. So lets change some other parameters and look whether we can classify our dataset with this feature or not.

5.3.5.4 Experiment 4 In our fourth experiment, we again changed the parametric setting, we changed the learning_rate_init from 0.001 to 0.0001.

Hypothesis:

The change in learning rate ensure the integrity of the model. If the jumps of learning is short and consistent then we can the output generated by this model also good enough as compared to model with high learning rate.

Result:

Accuracy result using train_x in predict= 66.03%

Accuracy result using test_x in predict= 16.48%

Discussion: The result is somehow better then the previous one. So, our hypothesis is correct. We will further explore and test some feature which can be used to make sure the classification of the dataset.

5.3.6 Experiment Type 5:

Feature used:

- Centroid of Original Image with Slit
- Centroid of Edge Detected Image with Slit

5.3.6.1 Experiment 1 In our first experiment, the parametric setting is set to default and we try to trained our model using centroid. Because each roman numeral is written in almost the middle of the image so instead of using the whole image we used 5 slits from the 50*50 matrix image and this time we used both images i.e the original one and image which only contains the edges of the roman numerals now this could be a good feature to classify the dataset.

Hypothesis:

As this is our first experiment with new feature, so we could not judge what kind of output would be. But we are guessing that the dataset could be somehow classified using this feature.

Result:

Accuracy result using train_x in predict= 93.71%

Accuracy result using test_x in predict= 22.26%

Discussion:

The result are much better than the result in our all previous experiment types where we used the whole centroid of each image or even with slits on centroid of edge detected image..

5.3.6.2 Experiment 2 In our second experiment, we changed the parametric setting, we added the 3 hidden layers each having 50 perceptrons in it. And the rest of the setting is same.

Hypothesis:

By introducing extra hidden layer the accuracy must increase and the loss rate should decrease because each addition of perceptron means extra unit of classification.

Result:

Accuracy result using train_x in predict= 93.56%

Accuracy result using test_x in predict= 20.17%

Discussion:

The result is not better than the result in our previous experiment. Which means that our hypothesis is not perfectly true that by increasing the number of perceptrons a better model can be generated.

5.3.6.3 Experiment 3 In our third experiment, we again changed the parametric setting, we changed the activation function from 'relu' which is a unit-step function with 'logistic' which is sigmoid function.

Hypothesis:

By introducing this activation function there is a chance that accuracy of the model increases because sigmoid activation function is considered good for multi class problems.

Result:

Accuracy result using train_x in predict= 78.37%

Accuracy result using test_x in predict= 17.58%

Discussion:

The result are not that much better than the previous one even they are little bit less. The result from he 'relu' activation is better than that so we choose that function for our next experiment. So, our hypothesis is wrong here. So lets change some other parameters and look whether we can classify our dataset with this feature or not.

5.3.6.4 Experiment 4 In our fourth experiment, we again changed the parametric setting, we changed the learning_rate_init from 0.001 to 0.0001.

Hypothesis:

The change in learning rate ensure the integrity of the model. If the jumps of learning is short and consistent then we can the output generated by this model also good enough as compared to model with high learning rate.

Result:

Accuracy result using train_x in predict= 69.90%

Accuracy result using test_x in predict= 18.20%

Discussion: The result is somehow better then the previous one. So, our hypothesis is correct. We will further explore and test these features to try to figure out the best setting of parameter for our model.

5.4 Analysis of Neural Network Setting

During the experiments, we have come to know about several things. One thing is that, sometime our accuracy rate for training data is more than even 90% whereas the accuracy of the validation data is much much less. So, this concept is known as over-learning where when we are training our model by cramming, instead of learning those concepts. The other thing we observed during the experimentation is that when we are using 'relu' as our activation function the chances of over-learning increases as compared to 'logistic'.

5.5 Recommended Experiment to be Done

Whle training our model the accuracy we got is almost near near 94%. So enhance this accuracy we could do something. When we predict our the training dataset on the trained model we should somehow pick up the images which are the real obstacle in the lour learning process. So if we somehow get rid of those image we can achieve the accuracy of 100%.

6 Conclusion

ANN is a vast field which compromises problems from different domains i.e classification, regression, clustering and dimensional reduction. Our effort in this project is to understand the what is classification? What are the concept behind it? How to build our own classifier? How to use MLPClassifier to classify a massive dataset of Roman Numerals? Throughout the project we have learned about these thing and also used these concept to implement an Artificial Neural Network for Classification problem.

7 References

1. https://towardsdatascience.com/the-differences-between-artificial-and-biological

2. https://ejournal.staim-tulungagung.ac.id/index.php/edukasi/article/download/533/513/2232

3. https://www.ibm.com/cloud/learn/neural-networks

4. https://www.psychologydiscussion.net/learning/learning-meaning-nature-types-and-652

5. https://1.cms.s81c.com/sites/default/files/2021-01-06/ICLH_Diagram_Batch_01_03-DeepNeuralNetwork-WHITEBG.png

6. https://lms.namal.edu.pk/pluginfile.php/10214/course/section/4572/Neuron.gif

7. https://res.cloudinary.com/edapp/images/w_300,h_272,c_scale/v1615779339/wordpress/production/social-learning-tagoras-1/social-learning-tagoras-1-300x272png?_i=AA?_i=AA

8. https://www.tonybates.ca/wp-content/uploads/Cognitive-elements-2.jpg