**A REPORT**

**ON**

**ETHNICITY/NATIONALITY IDENTIFICATION**

**USING DEEP LEARNING TECHNIQUES**

Submitted to

Dr. Navneet Goyal

**In partial fulfillment of the course**

**BITS F464 Machine Learning**

Prepared by

|  |  |
| --- | --- |
| Amit Agrawal | 2015B3A40610P |
| Chaitanya Maheshwari | 2015B1A70424P |
| Shrey Shah | 2015B1A70854P |



BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE,

PILANI

27 November, 2018

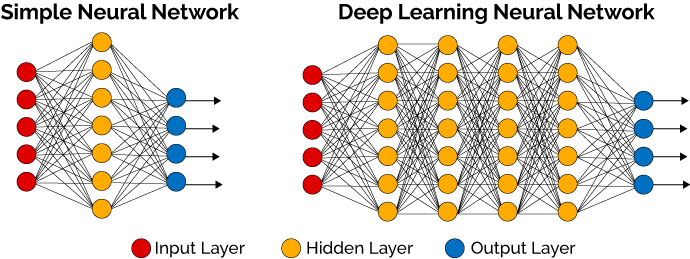
**Table of Contents**

1. Introduction
   1. Deep Learning
   2. Metric Learning
   3. LDA: A supervised metric learning algorithm
2. Problem Formulation
3. Data Preprocessing
4. Collection of data for training/testing
5. Model comparison and selection
6. Result and Analysis
7. Conclusions
8. Reference

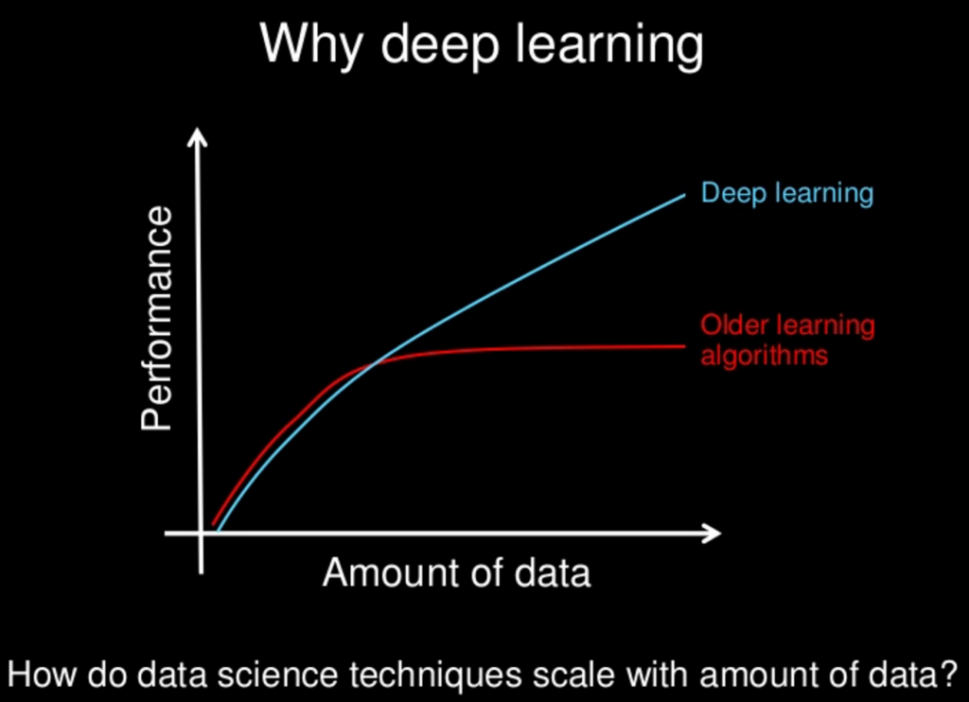
**Introduction**

**Deep Learning**

In layman terms, deep learning can be thought of as a neural network with more than one hidden layers.



Deep learning is also sometimes referred to as hierarchical learning or deep structured learning. It is a branch of machine learning that uses several layers of non-linear units such as deep neural networks, recurrent neural networks, convolutional neural networks, deep belief networks, etc. It can be used for all types learning, supervised and unsupervised and for all kinds of classification as well as pattern recognition tasks. It has found its use in a multitude of fields such as computer vision, audio and speech recognition, NLP, bioinformatics and drug discovery, etc., and have produced comparable or in some cases better results than humans.

The term was coined by Rina Dechter in 1986. Architectures of deep learning are constructed using layer-by-layer greedy approach, with each layer transforming the data into more abstract representation. It can be compared to the network of neurons in human brain.

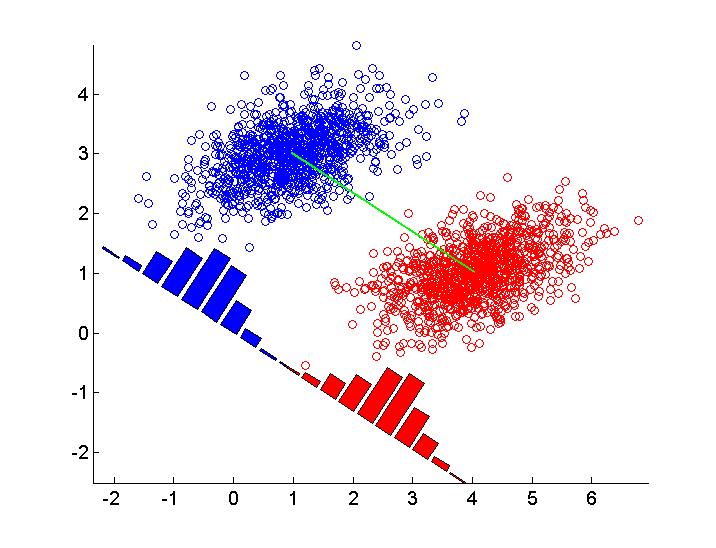
One of the striking features of deep learning is its increasing performance with the amount of data contrary to other machine learning algorithms which tend to stop learning after a certain amount of data is fed to them.

**Metric Learning**

In metric learning, our task is to learn the distance function over objects. These approaches heavily relies on the distances or similarity metric between the two samples. However, the metric used has to satisfy the four axioms or conditions which are non-negativity, identity, symmetry and triangle inequality. Supervised metric learning can be further bifurcated into global and local learning while unsupervised learning include linear methods like PCA as well as non-linear methods such as LLE and Isomap.

**LDA: A supervised metric learning algorithm**

Linear Discriminant Analysis is a method that is commonly used for dimensionality reduction and to find out a linear combination of the features that can characterize and separate classes or objects. It is based on preserving as much class discriminatory information as is possible and maximizing the ability to separate classes.



The approach was developed by R A Fischer in 1936. It is used before applying machine learning algorithms as a preprocessing step. LDA can be used only when the independent variables are measured as continuous quantities. LDA can also be used to predict bankruptcy, face recognition, product management, marketing, etc.

**Problem Formulation**

The human face is a profoundly rich upgrade that gives different data to versatile social association with individuals. People can process a face in an assortment of approaches to arrange it by its personality, alongside various other statistic attributes, including ethnicity or race, sex, and age. In the course of the last few decades,a parcel of eﬀort has been dedicated in the natural, mental, and intellectual sciences regions, to find how the human mind sees, speaks to, and recollects faces. Computational models have likewise been created to increase some understanding into this problem.The statistic highlights, for example, race and sexual orientation, are associated with human face personality acknowledgment. People are better at perceiving appearances of their own ethnicity/race than countenances of other race.

While confront acknowledgment has been around in some shape since the 1960s, later mechanical improvements have prompted a wide expansion of this innovation. This innovation is never again observed as something out of sci-fi motion pictures. Ethnic character can be utilized to astutely improve this world a place to live. A portion of the employments of Ethnicity Identification are:

* Counteract Retail Crime
* More brilliant Advertising
* Help Forensic enforcements
* Analyze certain ailments
* Encourage Secure exchanges

**Data Preprocessing**

As mentioned in the problem statement we had to compare the performance of a conventional neural network build from the basic perceptron model to that of a deep learning network.

For the conventional neural network we had to resize the RGB image from 200\*200 pixels to 64\*64 pixels which drastically reduced the number of trainable parameters, thus helping us to train the neural network in spite of our computational power bottleneck. We flattened the image giving us a total of 64\*64\*3=12,288 attributes. We normalized them, dividing by 255 (because the RGB color space uses 8 bit Integer values to represent a color ranging from 0 to 255) and they served as the input to the neural network.

Since deep learning networks require even more computation power then the simple neural network owing to the higher number of trainable parameters,we used a pre-trained network and fine tuned it for our particular dataset. For this part of the problem we did not downsize the image and hence the whole image 200\*200\*3 with 120,000 attributes was used as input to the deep learning model.

**Dataset Information**

We decided to use a dataset called UTKFace. UTKFace dataset is a large-scale face dataset with long age span (range from 0 to 116 years old). The dataset consists of over 20,000 face images with annotations of age, gender, and ethnicity. The images cover large variation in pose, facial expression, illumination, occlusion, resolution, etc. This dataset could be used on a variety of tasks, e.g., face detection, age estimation, age progression/regression, landmark localization, etc.

To obtain the information about age,gender,race and other attributes we looked up the labels of each face image as the information was embedded in the file name, formatted as “[age]\_[gender]\_[race]\_[date&time].jpg”

Here in the filename:-

[age] is an integer from 0 to 116, indicating the age

[gender] is either 0 (male) or 1 (female)

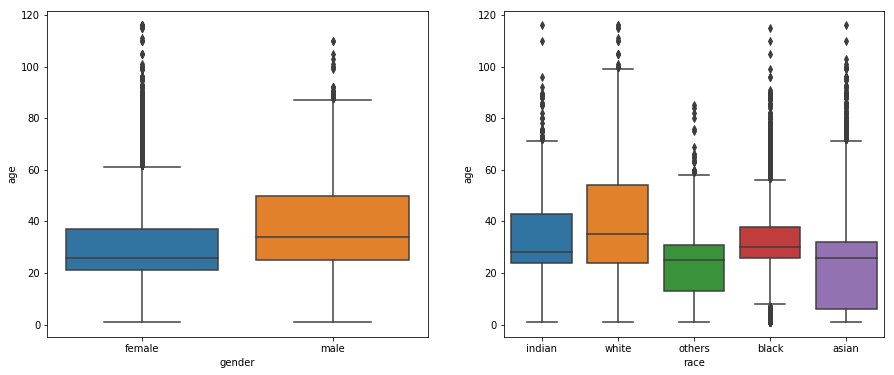
[race] is an integer from 0 to 4, denoting White, Black, Asian, Indian, and Others (like Hispanic, Latino, Middle Eastern) respectively.

[date&time] is in the format of yyyymmddHHMMSSFFF, showing the date and time an image was collected to UTKFace.

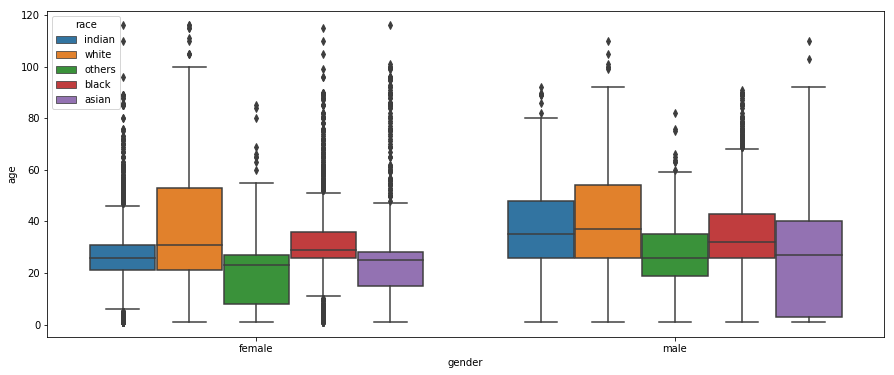
**Highlights**

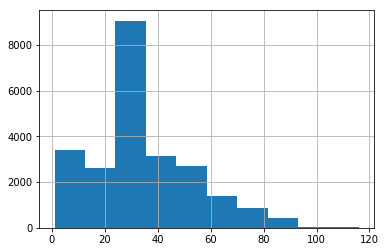
* consists of 20k+ face images in the wild (only single face in one image)
* provides the correspondingly aligned and cropped faces
* images are labelled by age, gender, and ethnicity
* The aligned and cropped images, as well as landmarks, are obtained by Dlib.

**Useful plots:**



From the above plots, we can observe that most of the females are between 20 and 40 years old whereas males are between ~25 and ~50 years old. When grouped by race, we see that the age groups vary quite significantly.

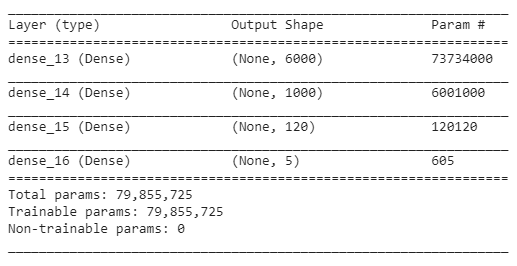




The above plot shows the histogram of age distribution in the dataset. We did not find it essential to take a subset of the data. Although if we take the dataset of people between 10-70 years of age, it might increase the accuracy as there is not much training data for ages 70-116.

**Model comparison and selection**

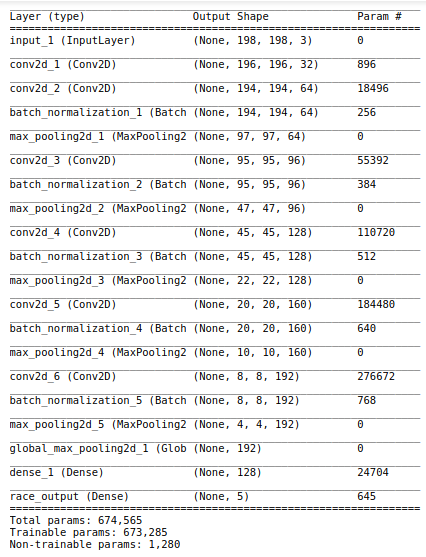
**Conventional Neural Network:**



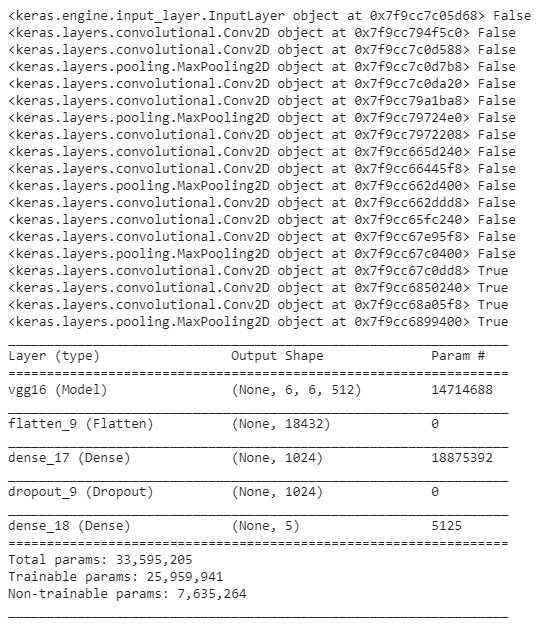
The normal neural network model had lot more attributes to be trained even though the network was not so deep owing to the fact that it was a fully connected architecture.

**Deep Learning:**

We first made and trained a deep learning model from scratch consisting of 6 convolution layers and 2 final fully connected layers which classifies the image into 5 different race classes. The above image represents a model summary of the constructed model.



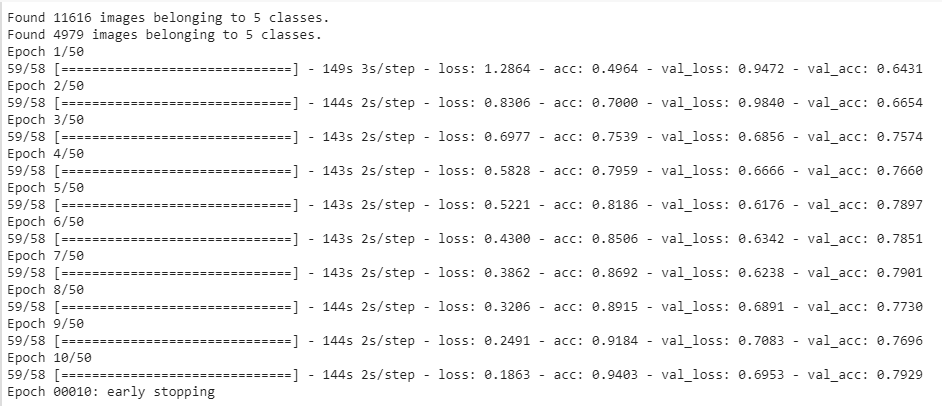
**Transfer Learning:**



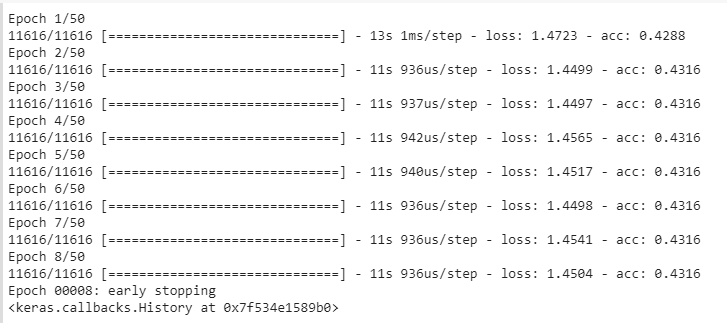
After that, we took a pretrained vgg-16 model trained on imagenet and froze all the layers except the last 3 convolutional layers. We re-trained the model on our UTKFace dataset which only modified the weights of the last 3 Convolution layers and the final fully connected layers and predicted the race from facial images in the dataset.

**Result and Analysis**

After training both the conventional neural network and the deep learning network we came up with the following results:



The image above shows the results after running the deep learning model. We can clearly see that the accuracy increases from 64% to almost 80% over a period of 10 epochs after which it stopped early because the model was moving towards over-fitting the data. The training accuracy climbed up from nearly 50% to almost 94%

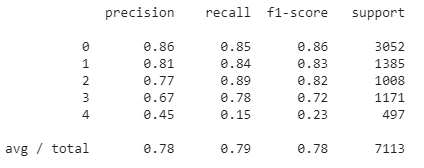


This image above are a result of the conventional neural network. As expected the above model performed poorly in comparison to the much better deep learning network. The accuracy obtained was just 43%. We certainly could improve a bit but due to limited computation power and time this is the model we came up with.

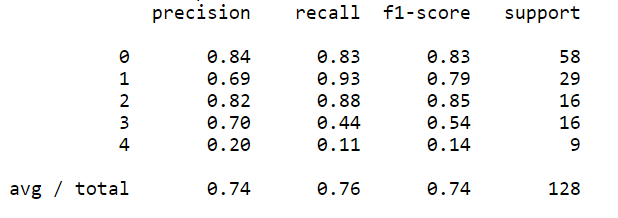
**Conclusions**



The fine tuned Deep learning model wrongly classified ~1400 images from a total of ~7100 images which can be quantified to roughly 80% accuracy as depicted in the classification report below.



Classification report for Fine tuned Deep Learning model

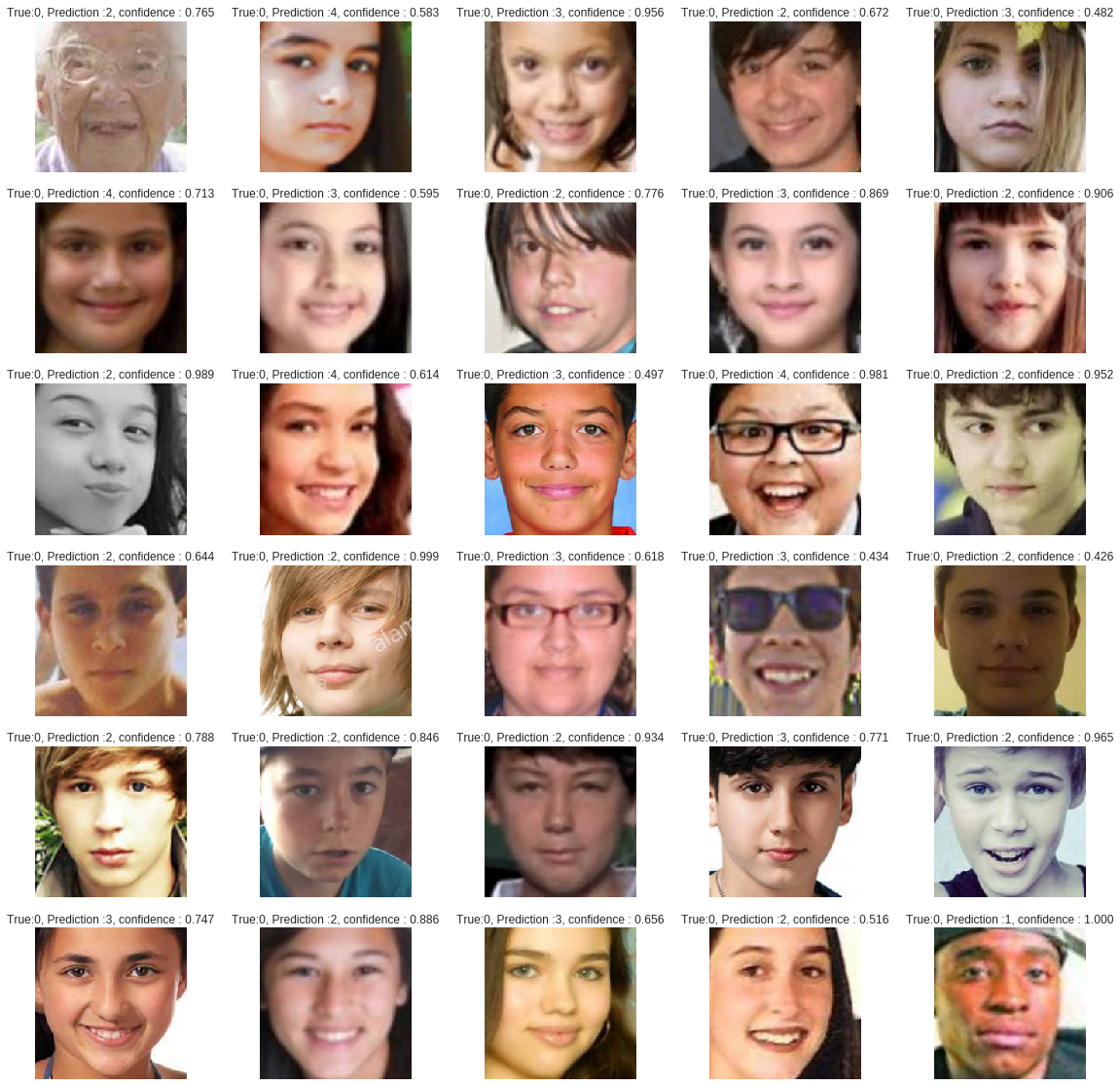


Classification report for Deep Learning model trained from scratch

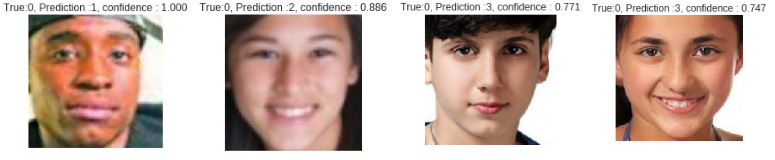
From the above classification report we can see a minor improvement in terms of the final results but the computation power and the amount of time required were significantly less for the fine tuned network in comparison to the network built from scratch.



The above figures represent the training and validation accuracy and loss of the fine tuned network while it was learning. It was early stopped automatically as the model was over-fitting. The validation accuracy was not increasing but the training accuracy was continuously increasing.



The above image shows the wrong predictions by our model. Some of the Images are deceiving to the human brain as well. Given below are some examples that the we as humans can also classify the same as the model did:



From the images above if one of us tries to predict the class of the individuals, we are likely to end up predicting the same result as the model did. This suggests that maybe these images were wrongly labelled/misclassifies in the UTKFace Datset.

**References**