AE401A REPORT (20/10/20)

Aditya Raghuwanshi
170052

Abstract

My presentation topic was "Deep Learning," and the aim of this presentation was to layout all essential tools & concepts to build a Deep Learning Model. It has a unique property that its model performs "feature engineering," i.e., the model itself learns crucial input data features. Most traditional machine learning algorithms exhaust with an increase in data, but deep learning algorithms do not. With increased data assembly and increased computation power, deep learning has become one of the hottest topics among machine learning techniques.

Neural Network Architecture

A single neuron (or a node) is a fundamental component of a Neural Network. It performs linear regression on input data on the node. Stacking multiple nodes forms a "layer," and stacking multiple layers forms a "deep" neural network. The number of layers and the number of nodes (depth) in each layer is a hyperparameter.
Activation Functions are used to map the input to the desired output range to avoid overflow and get more useful values. Nonlinear Functions are used as activation function as it enables the model to learn more complex functions. Some widely used activation functions are sigmoid, hyperbolic tan, and ReLU (Rectified Linear Unit).

Training a Deep Neural Network

We use "Stochastic Gradient Descent," which is an iterative algorithm to minimize the loss. It performs three operations in each iteration: (a) Make a prediction on a "batch" of data (forward propagation), (b) Evaluate loss on the prediction, (c) Update parameters to decrease the loss (backward propagation).

Overfitting

One of the main issues of a deep learning model is that it performs "too well" on not too complex data, therefore, overfitting. The solution to this problem is to use "Regularisation." The main aim of regularisation is to help the model not to trust the data fully. It not only helps in avoiding overfitting but also makes the model more "robust" from outliers. Some popular methods to implement regularisation are the Weight Decay Method, Dropout Method, and Early Stopping. In the weight decay method, we add an extra term

$$L = \Sigma_i L_i + \lambda . \Sigma_i ||w||_F^2$$

Adding this term moves all nodes' activation towards the origin (in linear range) and, thus, not allowing the model to learn complex nonlinear models.

Inside a Working Deep NN

During training, each node learns some features. Nodes at shallow layers learn simple features, and the ones at deeper layers learn complex features. While making predictions, the deep neural network tries to learn about the input data with its knowledge, starting from simple features to more complex features.

Disadvantages of using Deep NN

(a) To train a "good" Deep Learning Model, we require a large amount of data which is hard to compile, (b) Use of "sigmoid" activation function should be avoided in hidden layers as it slows down training because of small slopes at the extremes of the activation function. Use "ReLU" activation function instead, (c) have to write large complex code for a good model, use some frameworks like TensorFlow, (d) Training Process requires high computational power and lot of time, the solution is to use "Transfer Learning," which uses a pre-trained deep learning model to decrease the number of parameters, (e) We require to tune a lot of hyperparameters like learning rate, batch size, number of iterations, etc.

References

Kaggle DL Course	DL Explained (Melanie Swan)
DL, Andrew NG	3Blue1Brown
Hyperparameters in ML	Disadvantages od DL
CNN Basics	For Icons in ppt.

CS771A: Intro to Machine Learning Slides, IIT Kanpur by Prof Purushottam Kar