Unit-3 **Backpropagation**

"Backpropagation is the essence of neural network training"It is the method of fine-tuning the weights of a neural network based on the error rate obtained in the previous epoch (i.e., iteration).

Proper tuning of the weights allows you to reduce error rates and make the model reliable by increasing its generalization.

Backpropagation in neural network is a short form for "backward propagation of errors."

It is a standard method of training artificial neural networks.

This method helps calculate the gradient of a loss function with respect to all the weights in the network

"Backpropagation algorithm calculates the gradient of the error function"

Backpropagation can be written as a function of the neural network.

Backpropagation algorithms are a set of methods used to efficiently train artificial neural networks following a gradient descent approach which exploits the chain rule.

The main features of Backpropagation are the iterative, recursive and efficient method through which it calculates the updated weight to improve the network until it is not able to perform the task for which it is being trained.

Differentiation

Differential calculus is an important tool in machine learning algorithms.

Neural networks in particular, "the gradient descent algorithm depends on the gradient, which is a quantity computed by differentiation"

Differential calculus is an important tool in machine learning algorithms

Neural networks in particular, the gradient descent algorithm depends on the gradient, which is a quantity computed by differentiation

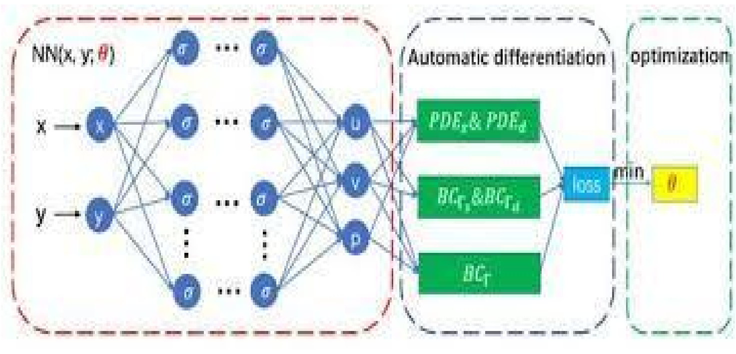


Fig: Differentiation

Hessian Matrix

Hessian matrices belong to a class of mathematical structures that involve second order derivatives They are often used in machine learning and data science algorithms for optimizing a function of interest.

Role of Hessian in Neural Computing :-

Several nonlinear optimization algorithms for neural networks are based on second order properties of error surface Basis for fast procedure for retraining with small change of training data

Identifying least significant weights

Evaluating the Hessian Matrix :-

Full Hessian matrix can be difficult to compute in practice

quasi-Newton algorithms have been developed that use approximations to the Hessian

Various approximation techniques have been used to evaluate the Hessian for a neural network

calculated exactly using an extension of backpropagation

Important consideration is efficiency

With W parameters (weights and biases) matrix has dimension W x W

Efficient methods have O(W2)

<u>Methods for evaluating the Hessian Matrix :-</u>

Diagonal Approximation

Outer Product Approximation

Inverse Hessian

Finite Differences

Exact Evaluation using Backpropagation

Fast multiplication by the Hessian

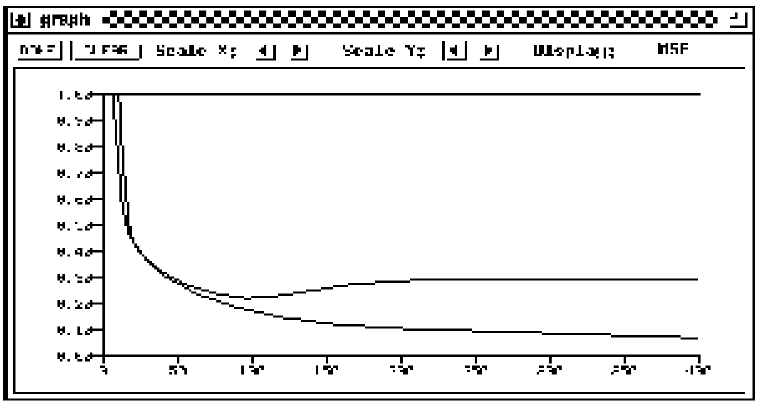
Generalization

one of the major advantages of neural nets is their ability to generalize.

This means that a trained net could classify data from the same class as the learning data that it has never seen before.

To reach the best generalization, the dataset should be split into three parts:

- The ***<u>training set</u>*** is used to train a neural net. The error of this dataset is minimized during training.
- The ***<u>validation set</u>*** is used to determine the performance of a neural network on patterns that are not trained during learning.
- A ***<u>test set</u>*** for finally checking the over all performance of a neural net.



Generalization of the ANN is ability to handle unseen data.

The generalization capability of the network is mostly determined by system complexity and training of the network.

Poor generalization is observed when the network is over-trained or system complexity (or degree of freedom) is relatively more than the training data.

A smaller network which can fit the data will have the k good generalization ability.

Network parameter pruning is one of the promising methods to reduce the degree of freedom of a network and hence improve its generalization.

Cross Validation

Cross-validation is a technique for evaluating a machine learning model and testing its performance.

CV is commonly used in applied ML tasks. It helps to compare and select an appropriate model for the specific predictive modeling problem.

There are a lot of different techniques that may be used to cross-validate a model. Still, all of them have a similar algorithm:

Divide the dataset into two parts: one for training, other for testing

Train the model on the training set

Validate the model on the test set

Repeat 1-3 steps a couple of times. This number depends on the CV method that you are using

CV is easy to understand, easy to implement, and it tends to have a lower bias than other methods used to count the model's efficiency scores.

All this makes cross-validation a powerful tool for selecting the best model for the specific task.

There are a lot of different techniques that may be used to cross-validate a model. Still, all of them have a similar algorithm:

Divide the dataset into two parts: one for training, other for testing

Train the model on the training set

Validate the model on the test set

Repeat 1-3 steps a couple of times. This number depends on the CV method that you are using

Network Pruning Techniques

Neural network pruning is a method of compression that involves removing weights from a trained model.

In agriculture, pruning is cutting off unnecessary branches or stems of a plant.

In machine learning, pruning is removing unnecessary neurons or weights.

There are different ways to prune a neural network.

(1) You can prune weights.

This is done by setting individual parameters to zero and making the network sparse.

This would lower the number of parameters in the model while keeping the architecture the same.

(2) You can remove entire nodes from the network.

This would make the network architecture itself smaller, while aiming to keep the accuracy of the initial larger network.

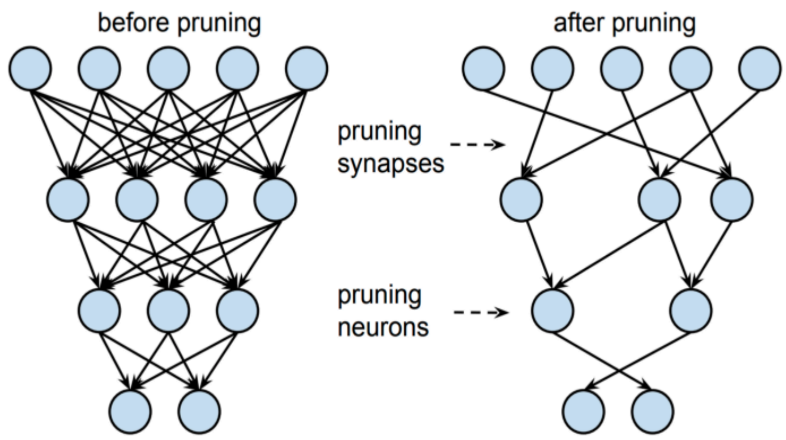


Fig: Prune weights

Limitations of Back Propagation Learning

It is slow, all previous layers are locked until gradients for the current layer is calculated

It suffers from vanishing or exploding gradients problem

It suffers from overfitting & underfitting problem

It considers predicted value & actual value only to calculate error and to calculate gradients, related to the objective function, partially related to the Backpropagation algorithm

It doesn’t consider the spatial, associative and dis-associative relationship between classes while calculating errors, related to the objective function, partially related to the Backpropagation algorithm

Accelerated Convergence

- Backpropagation is an optimization algorithm used in artificial neural networks to adjust the weights of the network during training.
- The algorithm works by propagating the error backwards from the output layer to the input layer, and adjusting the weights to minimize this error.
- There are several techniques that can be used to accelerate the convergence of backpropagation learning. Here are some of them:

  Use of Momentum: The use of momentum can help accelerate the convergence of backpropagation learning.

  Learning Rate Scheduling: Learning rate scheduling is the process of reducing the learning rate of the optimizer as training progresses.

  Weight Initialization: Proper weight initialization can also help accelerate the convergence of backpropagation learning.

Batch Normalization: Batch normalization is a technique that can help accelerate the convergence of backpropagation learning by normalizing the inputs to each layer.

Adaptive Learning Rate: Adaptive learning rate is the process of dynamically adjusting the learning rate of the optimizer during training

Supervised learning

Supervised learning is a type of machine learning in which an algorithm is trained on a labeled dataset.

- *Labeled data* is data that has already been tagged with the correct output.
- The *goal of supervised learning is to "learn a mapping function from input variables (features) to output variables (labels) based on the training data*"
- During training, the algorithm is presented with input-output pairs and it adjusts its parameters to minimize the difference between the predicted output and the actual output.
- Once the model is trained, it can be used to make predictions on new, unseen data.

Supervised learning algorithms can be used for a variety of tasks, including classification and regression.

"Classification involves predicting a discrete output"

output, such as a category or a label, while regression involves predicting a continuous output, such as a numerical value.

Some examples of applications of supervised learning include image recognition, speech recognition, spam filtering, and fraud detection

***