NEURAL NETWORKS AND DEEP LEARNING CST 395 CS 5TH SEMESTER HONORS COURSE- Dr Binu V P, 9847390760

CS 5th Semester Honors course for the Computer Science at KTU- Dr Binu V P

Techniques to Prevent Overfitting in Neural Networks

September 08, 2022

One of the most common problems that is encountered while training deep neural networks is **overfitting**.

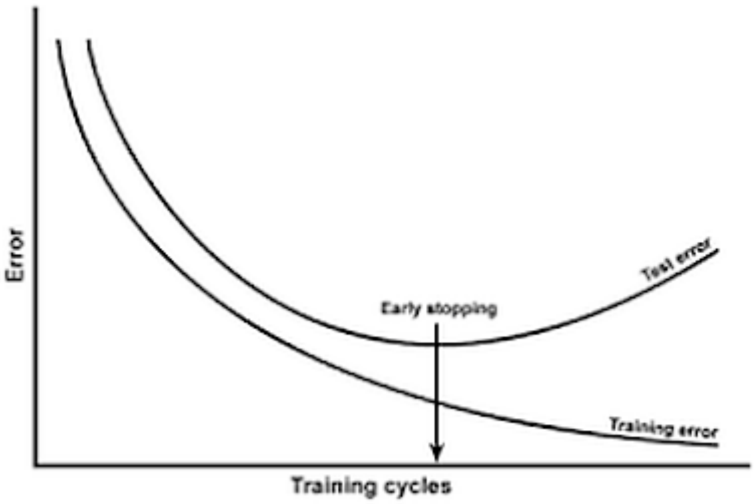
Overfitting occurs when a model tries to predict a trend in data that is too noisy. This is the caused due to an overly complex model with too many parameters. A model that is overfitted is inaccurate because the trend does not reflect the reality present in the data. This can be judged if the model produces good results on the seen data(training set) but performs poorly on the unseen data(test set). The goal of a machine learning model is to generalize well from the training data to any data from the problem domain. This is very important as we want our model to make predictions in the future on data that it has never seen before.

1. Simplifying The Model

The first step when dealing with overfitting is to decrease the complexity of the model. To decrease the complexity, we can simply remove layers or reduce the number of neurons to make the network smaller. While doing this, it is important to calculate the input and output dimensions of the various layers involved in the neural network. There is no general rule on how much to remove or how large your network should be. But, if your neural network is overfitting, try making it smaller.

2. Early Stopping

Early stopping is a form of regularization while training a model with an iterative method, such as gradient descent. Since all the neural networks learn exclusively by using gradient descent, early stopping is a technique applicable to all the problems. This method update the model so as to make it better fit the training data with each iteration. Up to a point, this improves the model’s performance on data on the test set. Past that point however, improving the model's fit to the training data leads to increased generalization error. Early stopping rules provide guidance as to how many iterations can be run before the model begins to overfit.



This technique is shown in the above diagram. As we can see, after some iterations, test error has started to increase while the training error is still decreasing. Hence the model is overfitting. So to combat this, we stop the model at the point when this starts to happen.

3. Use Data Augmentation

In the case of neural networks, data augmentation simply means increasing size of the data which helps in better generalization. Some of the popular image augmentation techniques are flipping, translation, rotation, scaling, changing brightness, adding noise etcetera.Using data augmentation a lot of similar images can be generated. This helps in increasing the dataset size and thus reduce overfitting. The reason is that, as we add more data, the model is unable to overfit all the samples, and is forced to generalize.

4.Use Regularization

Regularization is a technique to reduce the complexity of the model. It does so by adding a penalty term to the loss function. The most common techniques are known as $L1$ and $L2$ regularization:

The $L1$ penalty aims to minimize the absolute value of the weights. This is mathematically shown in the below formula.

$$L(x,y) \equiv \sum_{i=1}^{n} (y_i - h_\theta(x_i))^2 + \boxed{\lambda \sum_{i=1}^{n} |\theta_i|}$$

The $L2$ penalty aims to minimize the squared magnitude of the weights. This is mathematically shown in the below formula.

$$L(x,y) \equiv \sum_{i=1}^{n} (y_i - h_\theta(x_i))^2 + \boxed{\lambda \sum_{i=1}^{n} \theta_i^2}$$

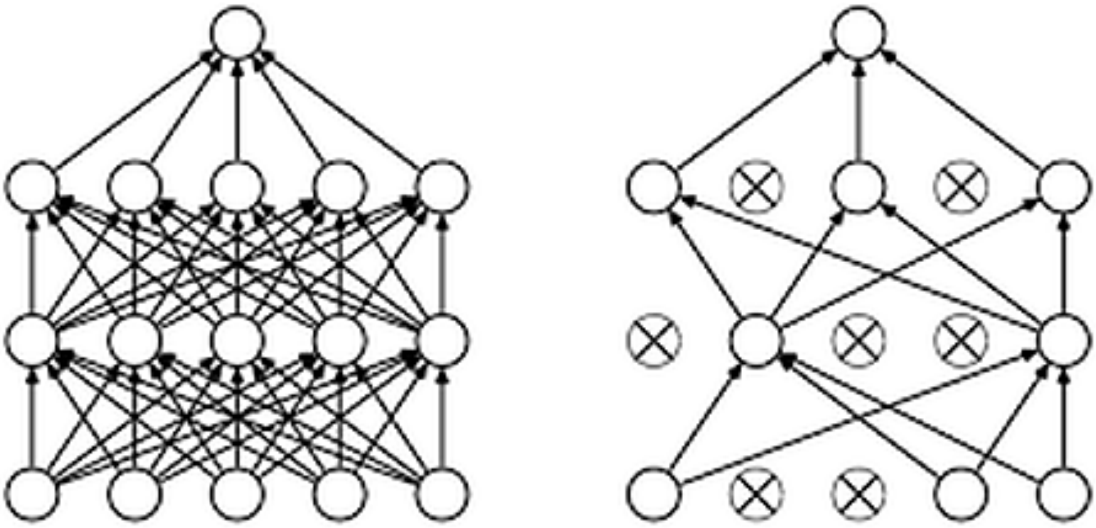
The below table compares both the regularization techniques.

L1 Regularization	L2 Regularization
1. L1 penalizes sum of absolute values of weights.	1. L2 penalizes sum of square values of weights.
2. L1 generates model that is simple and interpretable.	2. L2 regularization is able to learn complex data patterns.
3. L1 is robust to outliers.	3. L2 is not robust to outliers.

So which technique is better at avoiding overfitting? The answer is — it depends. If the data is too complex to be modelled accurately then $L2$ is a better choice as it is able to learn inherent patterns present in the data. While $L1$ is better if the data is simple enough to be modelled accurately. For most of the computer vision problems, $L2$ regularization almost always gives better results. However, $L1$ has an added advantage of being robust to outliers. So the correct choice of regularization depends on the problem that we are trying to solve.

5. Use Dropouts

Dropout is a regularization technique that prevents neural networks from overfitting. Regularization methods like $L1$ and $L2$ reduce overfitting by modifying the cost function. Dropout on the other hand, modify the network itself. It randomly drops neurons from the neural network during training in each iteration. When we drop different sets of neurons, it's equivalent to training different neural networks. The different networks will overfit in different ways, so the net effect of dropout will be to reduce overfitting.



(a) Standard Neural Net (b) After applying dropout.

This technique is shown in the above diagram. As we can see, dropouts are used to randomly remove neurons while training of the neural network. This technique has proven to reduce overfitting to a variety of problems involving image classification, image segmentation, word embeddings, semantic matching etcetera.

READ MORE

Machine Learning Algorithm

October 01, 2022

A machine learning algorithm is an algorithm that is able to learn from data. But what do we mean by learning? Mitchell (1997) provides the definition "A computer program is said to learn from experience E with respect to some class of tasks T and …

READ MORE

Syllabus CST 395 Neural Network and Deep Learning

October 01, 2022

Syllabus Module - 1 (Basics of Machine Learning) Machine Learning basics - Learning algorithms - Supervised, Unsupervised, Reinforcement, overfitting, Underfitting, Hyper parameters and Validation sets, Estimators -Bias and Variance. Challenges …

READ MORE

DR.BINU V P-9847390760

Associate Professor and Head Dept of Computer Science-IHRD Karunagappally.Love to share the thoughts and views .I play lot of Games and basically an Athlet in my school and college days.I never keep my studies as a second option.Stood first In MTech and BTech.I did my resea …