Course Name: Machine Learning and Deep learning - Fundamentals and Applications

Professor Name: Prof. M. K. Bhuyan

Department Name: Electronics and Electrical Engineering

Institute Name: Indian Institute of Technology, Guwahati

Week-11

Lecture-37

Welcome to NPTEL online course on machine learning and deep learning fundamentals and applications. In my last week, I explained the concept of the artificial neural networks. Also I explained the concept of the supervised and the unsupervised artificial neural networks. And also I explained the importance of the nonlinear activation function that is the sigmoid activation function. Also one important concept is the training of the feed forward network. So, we consider the back propagation training algorithm to adjust the weights of the artificial neural networks.

So, one major problem of the artificial neural network is the training of huge number of parameters that is the major problem of the artificial neural networks. So, how to train the artificial neural networks, because we have huge number of parameters for a complex classification problem. So, that problem can be addressed by the deep architectures. Today I will introduce the concept of deep learning.

And also I will be explaining the concept of the convolutional neural networks that is the CNN. So, first I will introduce what is the deep learning. And after this I will explain the concept of the CNN that is the convolutional neural network. So, let me begin this class. So, introduction to deep learning and the convolutional neural network.

So, the first question is what is deep learning. So, in this figure here you can see I am considering the problem of image recognition. So, you can see one input image here. So, this is the input image and a deep learning is nothing but the cascade of nonlinear transformations. So, it is important to extract important information from the input data.

In this case, we are considering the image. So, we can extract all the important information from the input image by considering this cascade of nonlinear transformations. So, it is nothing but a hierarchical model. So, if I consider the traditional pattern recognition, and the problem is the object classification or maybe the image recognition. In this case, I have shown the image recognition.

So, in case of the traditional pattern recognition, you can see here, I can extract the handcrafted features. So, maybe like shift and the one is the image feature and hog the histogram of oriented gradient or maybe the color feature texture feature. So, these image features I can extract from the input image. And after this, I can apply the supervised and unsupervised classification techniques for image recognition or image classification. This k-means clustering I can consider for clustering and we can consider the supervised classifiers like support vector machine for image classification or image recognition.

So, in this case, you can see the handcrafted features that is the shift the scale invariant feature transformation or the hog the histogram of oriented gradient. So, these type of features I can extract from the input image. So, in this case, I want to show the distinction between the machine learning and the deep learning. So, in the first figure you can see here. So, in case of the machine learning from the input image, I can extract features and based on these features, I can do the classification.

So, first I have to do the feature extraction. And after this, I have to do the classification in case of the traditional machine learning algorithms. In case of the deep learning, you can see here. So, simultaneously I can extract features and also do the classification. So, here you can see the feature extraction and classification I can do simultaneously and from the input image directly I can extract features and also I can do the classification in case of the deep learning.

So, that concept I am going to explain in my next slide. So, what is actually the deep learning in this case, you can see from the input image, I can extract information like the low level information, the mid level information, the high level information. So, all this information I can extract from the input image. So, I can consider some filters to extract this information, the low level information, the mid level information or the high frequency information. So, all this information I can extract from the input image.

So, in this example, I am considering the image but this is applicable for data also. So, if I want to compare the deep learning versus machine learning that is the comparison between the deep learning and the machine learning. So, first point is the data requirement. So, in case of the deep learning, I need the huge amount of training data. But in case of the machine learning, I can train on lesser data.

That means in case of the deep learning, I need the huge amount of training data. But in machine learning, I can train on lesser data. Accuracy, so if I consider the accuracy, so deep learning algorithms provides high accuracy as compared to machine learning algorithms. If I consider the training time, the deep learning techniques takes longer time for training. But in case of the machine learning takes less time to train.

And another parameter is the hardware dependency. So, in deep learning, generally we considered GPU. So, we need GPU. But in case of the machine learning, we can consider or we may use CPU because the computational complexity is less as compared to deep learning algorithms. And hyper parameter tuning.

So, in case of the artificial neural networks or in case of the machine learning techniques, we have limited tuning capabilities. But in the deep learning algorithms, these hyper parameters can be tuned in various ways. So, there are many techniques for tuning the hyper parameters in case of the deep learning algorithms. But that is not available in the case of the artificial neural network or the machine learning algorithms. So, the performance versus sample size.

So, here in the x axis, I am plotting the size of data that is the training samples, the size of the training samples or the training data. And in the y axis, I have shown the performance. So, the first the blue colored curve that is for the traditional machine learning algorithms. And if you see the deep learning algorithms, so, you can see the performance improves with the increase of the training data set. So, you can see the performance is improving with the help of the huge number of training data set.

If I have it use training data set, then the performance will improve in case of the deep learning algorithms. So, that is the comparison between the machine learning algorithms and the deep learning algorithms. So, this figure I have shown the concept of the traditional pattern recognition approach. So here also I am considering the input image and the problem is the object detection or the object classification. So, from the input image, I have to extract features that is nothing but the handcrafted features.

So, the low level computer vision features like the edges, shift, hog, color feature, texture feature. So, all these features I can extract from the input image. And based on these features, I can do the classification, I can do the object detection. This is the case of the traditional pattern recognition approach. So, maybe we can consider the classifier like support vector machine, k nearest neighbor, decision trees, these type of classifiers we can consider.

So, I have shown the input image, the problem is the classification or the object detection. And first we have to extract handcrafted features. After this we are considering tree enable classifier like support vector machine random forest. And after this, I am getting the output that is the classifier output I am getting. So, this is the supervised learning.

And in case of the deep learning, we considered the hierarchical learn representation. So, that is nothing but the cascade of nonlinear transformations we are considering. And we are extracting different information like the low level feature, the mid level feature, high level feature. And finally, we can consider the classifier that is the tree enable classifier for object detection or object classification. So, these type of features I can extract from the input image.

So, let us now discuss the concept of the convolutional neural network. So, before explaining the concept of the convolutional neural network, I will explain briefly the concept of the artificial neural network and what are the problems of the artificial neural network that I will highlight. So, already I told you the problem is the huge number of training parameters. So, let us consider a simple neural network, you can see here the input layer and I have the nodes x1, x2, x3, x4, these are the input nodes. And we have one output node that is the y and also we are considering hidden layers.

So, corresponding to these hidden layers, I have the node a11, a21, a31, a41. And in the second hidden layer, you can see the node a12. So, I have two hidden layers between the input layer and the output layer. And I have to consider the connections between these nodes in case of the simple artificial neural networks. So, all these connections I am showing here.

So, this is the simple artificial neural network that is the feed forward artificial neural network. So, this is the concept of the back propagation training. So, in case of the back propagation training, so I have to update the network weights. So, for this what we are considering, we are back propagating the error to adjust the weights of the artificial neural networks. So, the error can be computed that is nothing but the difference between the desired output and the actual output.

So, the objective is to reduce the error. So, the error is back propagated to the input to adjust the weights of the artificial neural network. And that is the concept of the training of the artificial neural network. So, again, I am showing the simple artificial neural network. And in this case, you can see, you can determine the neuron output that is the a11 that you can determine that is nothing but we are considering the activation function, maybe we can consider the sigmoid function or maybe another activation function is the ReLU function.

So, that concept I am going to explain later on. But we can consider the sigmoid activation function. And you can see here, the x1 is multiplied with w1 that is the weight is w1 connecting weight plus x2 is multiplied with w2 x3 is multiplied with w3 like this, all the multiplications I have to consider that is nothing but the weighted inputs I have to consider. So, w1, w2, w3, w4, these are the connecting weights. And after this, I am considering the activation function that is the sigmoid activation function.

So, we can determine a11 we can determine. So, ReLU function is nothing but the maximum of (0, x). So, what is the importance of the ReLU function that I will be explaining later on, but in this case, I am considering the sigmoid activation function. So, if I consider the ReLU activation function, it will be like this a11 and we are considering the maximum value maximum of this. So, and this is the concept of the ReLU function the maximum of (0, x).

And after this you can see I am showing another node that is the another node a12 in the second hidden layer. So, we have one input layer input layer is this the input layer and we have the output node. So, this output node and in between I am considering the hidden layers. So, we can compute the activation the neuron activation, the a11, a21, a31, a41. So, all these neuron activation I can determine and similarly a12 also I can determine the neuron activation I can determine.

And finally, we are considering the sigmoid function you can see here the sigmoid function we are considering. So, sigmoid function already I told you it is nothing but $\frac{1}{1+e^{-x}}$. So, that is the sigmoid function. So, with the help of the sigmoid function, I can get the output.

So, output is the y, y is the output. So, if I consider this simple artificial neural network. So, how many parameters in this case you can see I have 4 nodes x1, x2, x3, x4. In the first hidden layer, I have 4 nodes a11, a21, a31, a41. And in the second hidden layer, I have only 1 node that is a12 and I have the output node that is the y. So, for this you can see the all the interconnections.

So, how many parameters $4 \times 4 + 4 + 1$. So, these number of parameters I have to train. Suppose if I consider the input image one input image that is the RGB image that is the color image and the size of the image is $400 \times 400 \times 3$ because I am considering RGB image. So, one is the R component, one is the G component and another one is the B component. So, I have 3 components and the size of the image is 400×400. Corresponding to this if I want to consider the problem of the classification or the object detection.

So, I have to consider input nodes. So, how many nodes you can see x1, x2, x3 up to x480000. So, these number of nodes I have to consider in case of the input layer. I have 1 node in the output layer and I am showing 2 hidden layers. So, you can see how many nodes in the first hidden layer. So, a11, a21, a480000 these number of nodes in the hidden layer and only 1 node in the second hidden layer that is the a 12.

So, corresponding to this problem that is the image classification problem, you can see the number of parameters I can determine $480000 \times 480000 + 480000 + 1$. So, approximately 230 billion parameters I have to train. So, in this simple network, I have 230 billions parameters. So, it is very difficult to train or it is it is very difficult to estimate the values of the parameters. So, the huge number of parameters in this simple artificial neural network.

So, that is the major problem of the artificial neural network that is the large number of parameters I have to train. So, if I consider the back propagation training, so, all these parameters all these weights, I have to update and I have to estimate the values of these weights the parameters during the training. So, if I only consider suppose the connection is only 1000 connections instead of 480000 connections if I consider only 1000 connections, then in this case also approximately 480 million parameters that is also a huge number the big number. So, all these parameters I have to train during the back propagation training. So, that is the major disadvantage of the artificial neural network.

So, the question is how to reduce the number of parameters of the artificial neural network. So, based on this, I have to discuss the concept of the convolutional neural network. So, the objective is to reduce the number of parameters. Again, I am showing another artificial neural network.

In this case, you can see I have one input layer. So, I have three input nodes, four nodes in the hidden layer and two output nodes. So, corresponding to this you can see the weights the weights are w1 w2 these are the weights and we are considering the sigmoid activation function. So, in this case, you can see we are not counting the inputs. So, how many neurons 4 + 2 4. So, 4 neurons in the hidden layer and 2 neurons in the output layer.

So, we are not considering the input neurons. So, corresponding to this if I consider this interconnections, I have to consider 20 weights and also 6 bias. So, 6 biases I have to consider that means I have 26 learnable parameters for this simple artificial neural network. If the problem is complicated, I have to consider a multi layer neural network and I have to train the large number of parameters. So, that is the problem of the artificial neural network. So, now let us see how the convolutional layers help in reducing the number of parameters.

So, objective is to reduce the number of parameters with the help of the convolutional neural network. So, in case of the convolutional neural network, we are considering number of filters. So, here I have shown one filter. This is actually the Laplacian filter. So, with the help of this filter, I can detect the high frequency information of the image.

So, high frequency information of the image means the edges and the boundaries I can determine with the help of this filter. So, this is a high frequency filter. So, if I consider this is the input image and these are the pixel values of the input image. So, what I need to consider, I have to place the musk or the filter over the image and I have to do the convolution. So, like this, the placing the musk over the image and after this I am doing the convolution for all the pixels of the image.

So, I am showing it again, the placing the max over the image and I am doing the convolution. Convolution is nothing but the multiplication and the sum up, the multiplication of the max value and the image values and after this, the summing of this. So, that is the concept of the convolution and corresponding to this, if I do the convolution with this filter, there is a high pass filter. So, I am getting the convolved image like this. So, you can see I can detect the edges and the boundaries of the input image that is nothing but the high frequency information.

So, I am getting the feature map by considering this filter, the filter is the high pass filter. So, what is the convolution? So, what is the meaning of the convolution? Suppose if I consider this is the input image, this A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P. So, these are the pixel values of the image and this is the filter. So, if you see it here, the filter weights are W1, W2, W3 and W4.

So, these are the weights of the filter. After this, I have to place the filter over the image and I have to do the convolution. So, if I do the convolution, I am getting the convolved image that is nothing but the feature map. So, the placing of the max or the placing of the filter over the image, and after this I am doing the convolution. So, how to do the convolution, the mask value is multiplied with the pixel value $A1 \times W1 + B \times W2 + E \times W3 + F \times W4$. And after this we are considering the activation function, F is the activation function.

Corresponding to this I am getting the output. So, output is H1. After this I have to move the filter to the next pixel and corresponding to this, I have to do the convolution. So, H2 I can determine like this. So, I am determining the H2 and after this I am moving the filter to the next position and like this I have to determine the convolved image that is nothing but the feature map. So, to get one convolved image, I need only four parameters. So, number of parameters for one feature map that is equal to four, that means, that is a W1, W2, W3, W4 these are the parameters.

So, you can see for getting one feature map, I need only four parameters. So, with the help of this technique, that is a technique is the convolution technique, I can reduce the number of parameters. So, number of parameters for 100 feature maps, it is only 4×100. So, that is 400 parameters. So, if I consider 100 feature maps, only I need 400 parameters.

So, significantly I can reduce the number of parameters. So, in this figure also, I have shown the concept of the convolution. So, input matrix I have shown and one is the convolutional filter. And you can see the filter is placed over the image and I am getting the convolved feature that is a convolved image I am getting. So, this is the fundamental concept of the convolutional neural networks, I can reduce the number of parameters with the help of this convolution.

So, again, I am showing how to do the convolution. So, just placing the mask over the image here you can see the placing the mask over the image and I am getting the feature map. So, this is the feature map. So, place the mask over the image and I have to determine the convolution, move the mask to the next pixel, I have to determine the convolution like this, I have to get the feature map. So, in case of the boundary pixels, I can apply zero padding.

So, if I want to consider the boundary pixels, I can consider zero padding. So, I can put zero padding like this zeros. So, so that I can place the mask over the image. So, zero padding can be done to consider the boundary pixels. So, like this, I am getting the convolved image that is nothing but the feature map with the help of this filter.

So, this is the objective of the convolution. So, here also I have shown the convolution layer with the help of this filter. So, if you see this filter, that filter can detect the vertical lines, because you can see the response is this response 000 along the vertical lines. So, these are some examples of the convolution filters. So, all these filters I can consider to extract important information from the input image.

So, we can consider multiple filters. So, here in this case, you can see in the convolution layer, I can consider multiple filters. That means I have to learn multiple filters. So, if I consider it 200×200 image, so maybe I can consider 100 filters, the filter size may be 10×10. And corresponding to this I have 10 k parameters.

So, that is not a big number as compared to the artificial neural networks. So, again, I am showing the concept of the convolution. So, I am showing one image. So, 6×6 image I am considering and the filter the number of filters I am considering the filter 1 filter 2. So, all these filters I am considering. So, the filter 1 can detect the diagonal lines, you can see the response along the diagonal lines, the filter 2 can determine or detect the vertical line.

So, this you can see the response of the filter. So, so all this information I can extract from the input image. So, I can consider number of filters, the filter 1, filter 2, filter 3. So, number of filters I can consider for extracting the information from the input image. So, here you can see the concept of the stride.

So, stride is equal to one that means by how many pixels I am moving the max over the image. If I move the mask over the image by only one pixel, that means the stride is equal to one. If I move the mask over the image by two pixels, then the stride will be two. So, if I compare stride one or stride two, corresponding to stride two, the number of parameters will be less because I am considering shifting by two pixels. So, that means I can reduce the size of the feature map. So, with the help of the stride the parameter, the stride is a parameter, I can reduce the number of parameters.

So, already I told you as compared to stride one, the stride two gives less number of parameters because I am moving the mask over the image by two pixels. And we are considering the filter, filter 1 I am considering and you can see the response of the filter because this filter can detect the diagonal lines. So, you can see the three I am getting first if I do the convolution, I am getting the three after this move to the next position, if I consider stride is equal to two, so like this we can consider. So, if I consider stride is equal to one, so you can see the response of the filter. So, I am getting three because I have the diagonal line in the input image and that can be detected by the filter 1.

So, I am getting the feature map. So, this is the feature map I am getting corresponding to the filter 1. So, this is the feature map. So, I am doing the convolution and in this case we are considering the stride one. So, in the previous case, I am considering both cases stride two and also the stride one. So, stride two means I am moving the mask over the image by two pixels and corresponding to the stride one and corresponding to the filter 1, I am getting the feature map.

So, this is the feature map. So, similarly I can consider the filter number two and we can do the convolution. In this case, I am considering the stride one, so I am getting the feature map corresponding to the filter 2. So, like this, I can get all the feature maps corresponding to all the filters. So, I can consider number of filters and I can determine the feature maps.

So, here you can see in the feature map two 4×4 images forming $2 \times 4 \times 4$ matrix. So, that is the size of the feature map. So, here you can see I am getting 4×4 images in the feature map, the input image is 6×6. So, the stride already I have explained. So, that is the stride means the number of steps the kernel is moved during the convolution.

So, corresponding to this input image 7×7 input image and if I consider the 3×3 kernel. So, corresponding to the stride one I am getting the feature map, this feature map and corresponding to the stride two I am getting this feature map. So, with the help of the stride I can reduce the size of the feature map. That means I can reduce the number of parameters. So, one case how to reduce the parameters by considering the convolutional neural network that is the by considering the concept of the convolution, I can reduce the number of parameters and also by considering the stride I can reduce the number of parameters.

Suppose if I want to apply this principle in case of the color image. So, how to apply this principle in the color image. So, we are considering the RGB image. So, I have three channels R channel, G channel and the blue channel. So, red, green and the blue.

So, corresponding to this color image, I want to apply the filters the filter 1 filter 2 I can apply. So, for all the channels I have to apply the filter 1 and similarly, the filter 2 can be applied for all the channels of the color image and corresponding to this I am getting the feature map. So, you can see I am getting the feature map. So, this is the feature map for the color image. This is a feature map for the color image.

So, I am showing the concept of the dense neural network and the convolutional neural network. The first figure I have shown the conventional multi layer perceptron that is the feed forward network. I have one input layer and one output layer and in between I have two hidden layers in case of the artificial neural network that is shown in the first figure. So, this is nothing but the ANN artificial neural network and that is the feed forward network. In the convolutional neural network that is the CNN. So, here you can see corresponding to this input image and this is the input image suppose input data and I am applying the filters the multiple filters.

So, this is the filters. So, number of filters I am applying and I am getting the feature map. So, this feature map is represented by these parameters, the width and the height that is nothing but the size of the feature map. The depth means the how many feature maps because I am considering number of filters and corresponding to each and every filter I am getting one feature map and that is represented by the depth. So, how many feature maps I am getting that corresponds to the depth and width and the height that is nothing but the size of the feature map. And finally, you can see this and the feature map I am applying to the classifier.

So, this is the final classifier. So, the input to the classifier is nothing but the feature maps. So, this is the structure of the convolutional neural network. So, you can see the difference between the dense neural network and the convolutional neural network. So, now, I will discuss another concept of the pooling. So, first one is the max pooling and another one is the average pooling.

In a rectangular neighborhood, I have to find a maximum output in case of the max pooling. And in case of the average pooling, so average output of a rectangular neighborhood I have to consider that is the average pooling. So, this concept I am explaining now. So, suppose if I consider this is the feature map. So, input matrix is this and we are considering the 2×2 neighborhood.

So, in the 2×2 neighborhood if I apply the max pooling, so the size of the filter is 2×2. So, that means the neighborhood size is 2×2 with a stride 2. So, corresponding to this, so output matrix will be like this. So, corresponding to this 2×2 neighborhood, this is the maximum value, 4 is the maximum value, the 5 is the maximum value in the 2×2 neighborhood, the 3 is the maximum value and the 4 is the maximum value. So, only these values we are considering.

That means I am getting the information of the local statistics of the input data. And with the help of this pooling, I can reduce the size of the feature map. So, here in the input, you can see the size is 4×4. But if I see the output matrix, the size is 2×2. So, that means you can reduce the size of the feature map with the help of the pooling.

And also you can extract information of the local statistics of the input data. So, this is the concept of the max pooling again I am showing. So, you can see the feature map, input feature map. And we are considering 2×2 filters for the max pooling and the stride 2 is considered and the maximum value we are getting the first value is 6, another value is 8, another value is 9 and the final value is 7 corresponding to the 2×2 neighborhood. So, in summary, you can see I am doing this one and that you can see the first I am doing the convolution. So, in the figure you can see the first figure I am doing the convolution and we are applying the ReLU activation function that is nothing but the maximum of (0, Z).

So, in the convolutional neural network, you can see the first part is the feature extraction and second part is the classification. So, you can see I am considering the first convolution, the ReLU function, the next one is the pooling again the convolution again the nonlinear activation function that is the ReLU function and after it is the pooling. So, number of the convolution and the pooling layer we can consider to extract the features from the input image and finally, I can consider a fully connected layer for classification. So, how to determine the size of the feature map W is the input volume size.

So, that is the size of the input image or the input volume size. After this F is the size of the filter that is the filter size P is the zero padding used on the border. So, if I consider the boundary pixels I have to consider zero padding S is nothing but the stride. So, with the help of this formula and $\frac{W-F+2P}{S}+1$ I can determine the output size that is the size of the feature map I can determine. So, corresponding to this example, so you can see the filter sizes 3×3 size of the input images 5×5 and we are not considering the zero padding.

So, that is why it is zero stride is considered as one. So, what is the size of the feature map the size of the feature map is 3×3 that is the size of the feature map. So, with the help of this formula and this formula I can determine the size of the feature map that is the output size I can determine. So, the pooling layer already I told you so it can extract local statistics of the input data. So, corresponding to this you can see we are considering all these input nodes and corresponding to this I am extracting the maximum value or the average value of this that means I am extracting the information of local statistics of the input data and also I can reduce the size of the feature map with the help of this pooling either I can use the max pooling or the average pooling.

So, max pooling already I have explained. So, these filter 1 and filter 2 we are considering. So, 3 × 3 filters filter 1 and filter 2 and we are considering the 2 × 2 neighborhood. So, in the 2 × 2 neighborhood I have to consider the max pooling. So, in the first figure you can see the maximum value here is 3. So, this 3 is considered similarly in the next one is the 0 0 is the maximum value.

So, we are considering 0 there is a maximum value we are considering in this case the maximum value is 3. So, that is considered in this case the maximum value is 1. So, that is considered after the pooling similarly in case of the filter 2 this we are considering the maximum value we are considering minus 1 in the feature map 1 is considered the maximum value 0 is considered and 3 is considered. So, max pooling is nothing but the subsampling. So, if I consider this problem the identification of the bird.

So, if I do the subsampling the subsampling will not change the object. So, with the help of this pooling I can reduce the number of parameters. So, fewer parameters to characterize the image. So, with the help of few number of parameters I can characterize the input image.

So, the complete CNN is like this. So, suppose the input image is this is the input image. So, I can apply convolutional layer the convolution layer and the max pooling layer or the average pooling layer convolution pooling. So, repeatedly I can do the convolution and the pooling and I am getting the new image you can see I am getting the new image. So, which is smaller than the original image.

So, number of channels is the number of filters. So, if I consider 3 filters, so that means 3 channels I am getting. So, this convolutional operation and the pooling operation I can do repeatedly to extract important information from the input image. So, here I am showing this concept, the convolution pooling convolution pooling I am doing repeatedly to extract the important information from the input image. So, that is why we are considering the filter banks here. So, you can see we are considering the filter banks the number of filters to extract important information from the input image.

And we are applying the rectification that is nothing but the ReLU activation function. And after this, we are applying the pooling operation also. So, you here you can see the this is the output of the pooling. So, we can apply this principle convolution pooling convolution pooling like this. So, here you can see this is the typical architecture of the convolutional neural network.

So, one stage is the convolution and the pooling. So, this I can repeat. So, like this the in the first stage I am applying the convolution and the pooling second stage also I am applying the convolution and the pooling. So, like this I can apply repeatedly the convolution and the pooling. So, that I can extract important information from the input image. And finally, I am considering fully connected layers for classification object classification or object recognition. So, here you can see this is the input image and we are applying the these filters the filters are represented like this and max pooling is represented like this.

So, it is 64 filters after this we are applying the max pooling again we are applying the filters. Again I we are applying the max pooling we are applying the filters. Again, we are applying the max pooling again we are applying the filters again we are applying the max pooling again we are applying the filters there is a convolution like this we can do and finally, we are getting the max pool output vector. So, we are getting the output vector like this the max pool output vector and finally, we are considering the fully connected layers. So, all the interconnections I have shown for the classifications.

So, feature extraction is done by these steps, the filters that is the convolution and the pooling. And finally, you can see I am doing the classification and the classes are living room, bedroom, kitchen, these are the classes, all these classes. So, here you can see I want to show the difference between the convolution and the fully connected. So, in the first case, you can see I am showing the concept of the convolution and I am getting the pacer map with the help of this convolution.

So, this is the pacer map. So, we are applying the filters. So, this filter also we are applying and this filter is applying and I am getting the convolution. So, I can reduce the number of parameters with the help of the convolution. But if I consider suppose corresponding to this input image, this input same input image, if I consider the fully connected neural networks that is the feed forward network, I have to consider 36 nodes, 36 node, you can see all the pixels we are considering 36 pixels, because size of the image is 6×6. So, that means I have to consider 36 input nodes and I have to consider all the interconnections.

So, that means I have to consider huge number of parameters in case of the fully connected system. But in case of the convolutional neural network, I can reduce the number of parameters. So, you can see here we are considering this concept of the convolution and we are considering the filter, the filter one is it is considered and we are getting the feature map. So, this is the feature map we are getting.

So, in this case, just only connect to 9 inputs not fully connected. So, we are only connecting the 9 inputs in case of the convolution. So, that means we have few parameters and we are connecting only 9 inputs. But in case of the feed forward network, I have to connect all the all the points all the nodes or all the pixels. But in this case, we are connecting only the 9 inputs. If I consider sharing of the weights, I can even reduce the number of parameters.

That means if I consider the shared weights, I can reduce the number of parameters. So, in the figure you can see I am just sharing the weights. So, this is the complete CNN. So, input image and we are doing repeatedly the convolution and the pooling convolution and the pooling. And after this the output of this max pooling that is flattened that is flattened means it is converted into a vector and that is inputted to a fully connected feed forward network. And with the help of this feed forward network, I can do the classification.

So, that is the concept of the flattening that means the array is converted into vector. So, you can see with the help of this convolution, I am getting the new image, convolution max pooling convolution max pooling. So, I am getting the new images like this. And I am flattening and it is connected to the fully connected feed forward network.

So, this is the concept of the flattening. So, this array is converted into 1D vector and that is the input to the fully connected feed forward network. So, you can see the concept of the flattening. So, the final architecture of the CNN is we are considering the input image or input data convolution nonlinearities obtained by considering the ReLU function. Again the pooling convolution nonlinearities obtained by the ReLU function, pooling operation and finally the fully connected layer we are considering.

So, this is the structure of the CNN. So, the activation function sigmoid that is actually already I have discussed. So, takes a real valued number and squashes into a range between 0 and 1. So, 0 means not firing and 1 is the fully firing. So, the problem of the sigmoid function is the the vanishing gradient problem. So, this problem I will be explaining in my next class what is the problem of the saturation or the vanishing gradient.

So, that concept I will be explaining in my next class. So, that is why the sigmoid function is not used in case of the CNN. In case of the CNN, we are not using the sigmoid function, the problem is the vanishing gradient problem and the saturation. The another function is the tan hyperbolic function. So, ranges from - 1 to 1. So, this is the tan hyperbolic function tanh function, like the sigmoid function, the problem is the saturation and unlike sigmoid, the output is 0 centered, you can see the output is 0 centered, it is centered at 0 and the tan hyperbolic is a scale sigmoid.

So, tanh x is nothing but it is equal to 2sigmoid(2x) - 1. So, it is a scale sigmoid, the tan hyperbolic is nothing but the scale sigmoid. But in this case, the problem is again the saturation problem. So, that is why it is also not popular. And the popular activation function is the ReLU activation function that is generally used in the CNN.

So, fx is equal to maximum of (0, x) and corresponding to this you can see. So, all the values, negative values will be 0. So, if you see these values, all these values are 0 and corresponding to these values, it is nothing but fx is equal to x. So, we are considering the maximum of (0, x). So, this ReLU function is popular in case of the the CNN the convolution neural network, because it prevents the gradient vanishing problem.

The vanishing gradient problem is a major problem. And that is why we are considering the ReLU function. So, it is nothing but the maximum of (0, x). So, all the negative inputs, you can see that is mapped to 0. So, that means we are not considering all the activations. Only we are considering some of the activations, the negative activations we are not considering 0. So, this is the complete structure of the convolutional neural network, you can see the input image is there and the convolution plus ReLU activation function.

So, we are having. So, we are getting the feature map. After this we are applying the pooling layer. So, we can reduce the size of the feature map with the help of the pooling. And this process the convolution and the pooling we can do repeatedly to extract important information from the input data or input image. And after this we are considering the fully connected layer. So, we have to do the flattening, flattening of the feature map that is the array is converted into 1D vector. And that is the input to the fully connected layer. And we are applying this fully connected layer for classification. And in this case, we are considering the softmax activation function.

So, these values we are getting the probabilities. So, number is converted into probabilities, the probability is 0.94, 0.03. So, that means the numbers are converted into probabilities with the help of the softmax function. So, this concept already I have explained. So, what is the summary of this discussion. So, a CNN compresses a fully connected network in two ways.

The first one is the reducing the number of connections that already we have discussed and the sharing the weights on the edges. So, we can share the weights and also the max pooling reduces the complexity. So, that means in case of the CNN, we are reducing the number of parameters by considering the convolution operation, the max pooling operation, and also by considering the concept of the stride and also by considering the concept of the sharing of the weights. That is the advantage of the convolutional neural network. So, in case of the MLP versus CNN, you can see the first one is the CNN and the second one is the feed forward network.

So, you can see we have the input layer, one is the output layer and in between I have the hidden layers. And you can see the interconnections between these neurons. So, here you can see the comparison between the MLP versus CNN. So, in the first you can see I am showing the concept of the CNN and how to extract the feature map with the help of the convolution. And in the second case, I am showing the concept of the feed forward artificial neural network.

So, first point is the sparse connectivity. Every node in the convolution layer receives input from a small number of nodes in the previous layer. So, you can see the first concept, the every node. So, if I consider this node in the convolution layer receives input from a small number of nodes in the previous layer, that is nothing but the receptive field. So, that means, we are needing smaller number of parameters.

That means we are getting the smaller number of parameters because of this convolution. And also we are considering the sharing of the parameters. Each member of the convolution kernel is used at every position of the input and dramatically reduce the number of parameters. So, this makes CNN much more efficient than the MLP. So, this is the comparison between the MLP versus CNN.

In this class, I explained the concept of the deep learning and also explained the concept of the convolutional neural network. The objective is to reduce the number of parameters. So, for this we considered the concept of the convolution, the pooling and the concept of the sharing of the weights. So, with the help of the convolution, I can reduce the number of parameters. And also I can extract the statistics of the input data with the help of the pooling.

And also I can reduce the number of parameters or I can reduce the size of the feature map with the help of pooling. I can consider max pooling or average pooling. And finally, I can consider the sharing of the weights so that I can reduce the number of parameters. So, that is the advantage of the CNN as compared to the fully connected artificial neural networks. So, let me stop here today. Thank you.