Convolutional Neural network (ConvNet)

Page No.
Date

Dropout is a hyperparameter tuning done in convolutional neural networks

→ What is feed-forward Network

Remember that feed-forward network are also called multilayer perceptrons (MLPs), which are the quintessential deep learning models. The models are called "feed-forward" because information flows right through the model. There are no feedback connections in which outputs of the model are fed back into itself.

→ These models are called feedforward because information flows through the function being evaluated from x, through intermediate computations used to define f, and finally to the output y. There are no feedback connections in which the output of the model are fed back into itself.

When feedforward neural networks are extended to include feedback connections, they are called recurrent neural networks

What does Convolution layer does?

Imagine you have an image represented as 5×5 matrix of values, and you take a 3×3 matrix and slide that 3×3 window or kernel around the image. At each position of that matrix, you multiply the values of your 3×3 matrix window by the values in the image that are currently being covered by the window. As a result, you'll get a single number that represents all the values in that window of images. You use this layer to

to filtering : as the window moves over the image, you check for patterns in that section of the image.

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Input

1	0	1
0	1	0
1	0	1

filter / kernel

1×1	1×0	1×1	0	0
0×0	1×1	1×0	1	0
0×1	0×0	1×1	1	1
0	0	1	1	0
0	1	1	0	0

Input x filter

4		

feature map of above case

Note that convolutional filter or kernel must be (odd x odd) for example 3×3
5×5
7×7 etc.

(2) The objective of Subsampling is to get an input representation by reducing its dimensions, which helps in reducing overfitting. One of the techniques of subsampling is max pooling. With this technique, you select the highest pixel value from a region depending of its size.

For example in this case our max pooling filter is 2×2 in size.

1	0	2	3
4	6	6	8
3	1	1	0
1	2	2	4

→

6	8
3	4

Result after max pooling

Now time to coding more stuffs we learn while doing coding.

0 = np.unique(train_Y)

↳ this gives the list of unique elements present in train_Y

Simple way to produce one_hot encoding

```
from keras.utils import to_categorical
one_hot = to_categorial(Y_train).
```

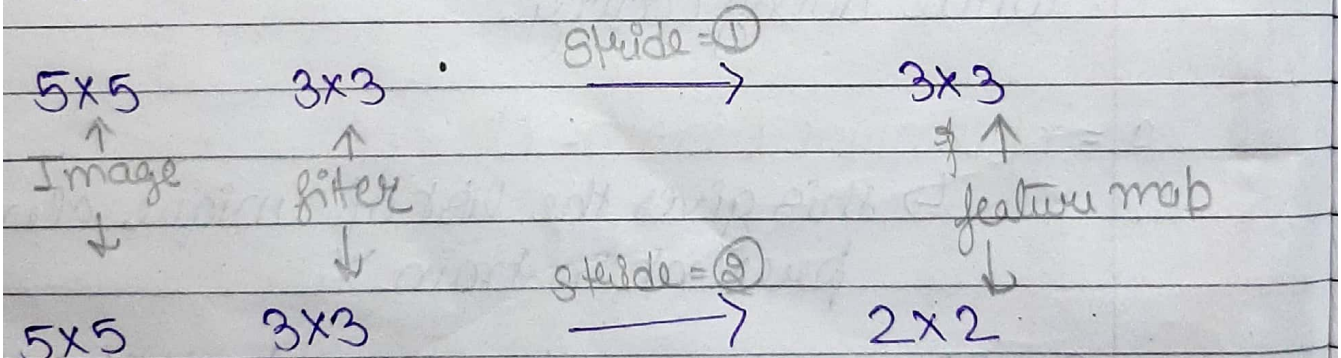
Concepts of Non-linearity

Page No.
Date

For any kind of neural network to be powerful, it needs to contain non-linearity. Both the ANN and autoencoder achieved this by passing the weighted sum of the inputs through activation function, ↳ in blogs point of view CNN is no different. we again pass the result of the convolution operation through relu activation function. So the values in the final feature maps are not actually the sums, but the relu function applied to them. Keep in mind that any type of convolution involves a relu operation, without that the network won't achieve its true potential.

Stride

Stride specifies how much we move the convolution filter at each step. By default the value is 1



Padding

we see that the size of the feature map is smaller than the input, because convolution filter needs to be contained in the input. If we want to maintain the same dimensionality, we can use padding to surround the input with zeros

we either pad with zeros or the values on the edge. Now the dimensionality of the feature map matches the input. Padding is commonly used in CNN to preserve the size of the feature maps, otherwise they would shrink at each layer., which is not desirable.

In CNN architectures, pooling is typically performed with 2×2 windows, stride 2 and no padding. While convolution is done with 3×3 windows, Stride 1 and with padding

→ In keras, there are 2 options for padding

① padding = 'same' → means we pad with the numbers on the edges.

② padding = 'valid' → means no padding

Dropout

Dropout is by far the most popular regularization technique for deep neural networks. Even the state-of-the-art models which have 95% accuracy can get a 2% accuracy boost just by adding Dropout.

Dropout is used to prevent overfitting and the idea is very simple. During training time, at each iteration, a neuron is temporarily "dropped" or disabled with probability P. This means all the inputs and outputs to this neuron will be disabled at the current iteration. The dropped-out neurons are resampled with probability p at every training step, so a dropped out neuron at one step can be

Page No. _____
Date | | | |

active at the next one. The hyperparameter p is called the dropout rate and it's typically a number around 0.5, corresponding to 50% of the neurons being dropped out.

It's surprising that dropout works at all. We are disabling neurons on purpose and the network actually performs better. The reason is that dropout prevents the network to be dependent on a small numbers of neurons, and forces every neuron to be too dependent on a small number of neurons, and forces every neuron to be able to operate independently.