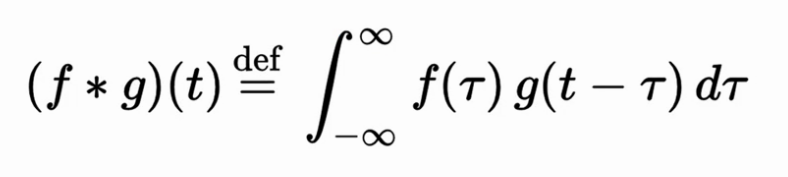
The convolutional neural networks have 4 stages, convolution, max pooling, flattening and full connection.

They work by representing images as arrays of pixels and putting these through a CNN, to match to output classes.

**Convolution:**



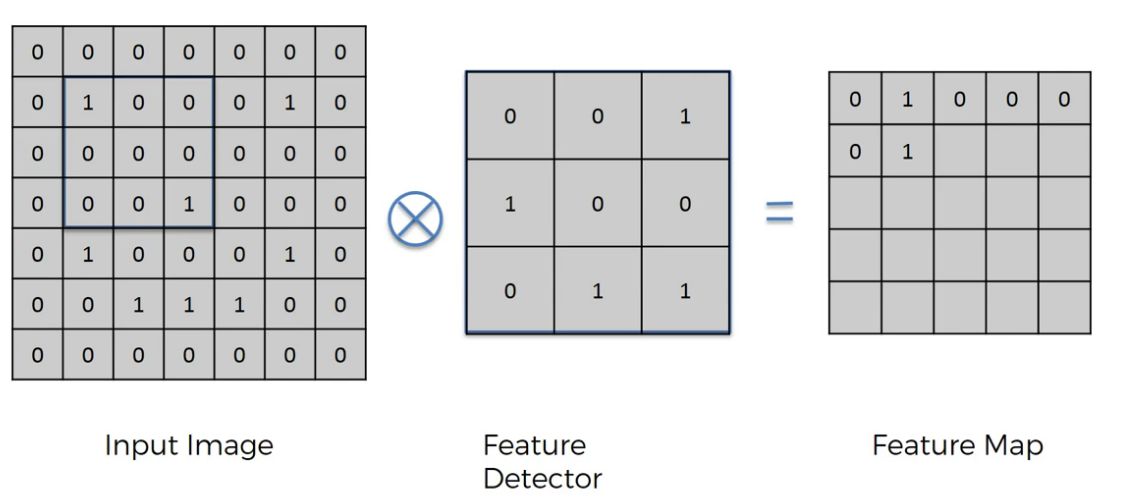
**Input image**, matrix representing the image

**Feature detector:**

3x3 matrix, (can be other dimensions).

\*also known kernel, or filter

The way to operate the feature detector is to place it at the top left corner of the image and run an “AND” (multiplication) function on the numbers the feature detector lines up with on the input image:



In the end we SUM all the line ups.

So in the above case we multiply the top left corner box of both the feature detector and the selected box in the input image.

1x0 = 0. Then we do the next one and so on and sum al of them. The end result is only one of them actually match up and the result of the AND is 1+(0+0+0+0+0+0+0+0).

The feature map is the result.

We start the feature map at the top left corner, and move slide it one ‘box’ to the right until the end, Then move to the next row down. The amount you move Is called the ‘stride’. Common strides = 2.

Feature maps can also be called activation maps or ‘convolved’ maps; ie; subject to a convolution operation.

One of the advantages is that it works like a compression. Algorithm that makes the image smaller. We do lose some value.

The way the feature map works as a filter that looks at each part of the image, and matches it against the feature detector and gives a score for whether the feature is present.

Several of thses feature maps are used to identify different features, each generating a score for whether they exist in the input image.

**Relu Layer**

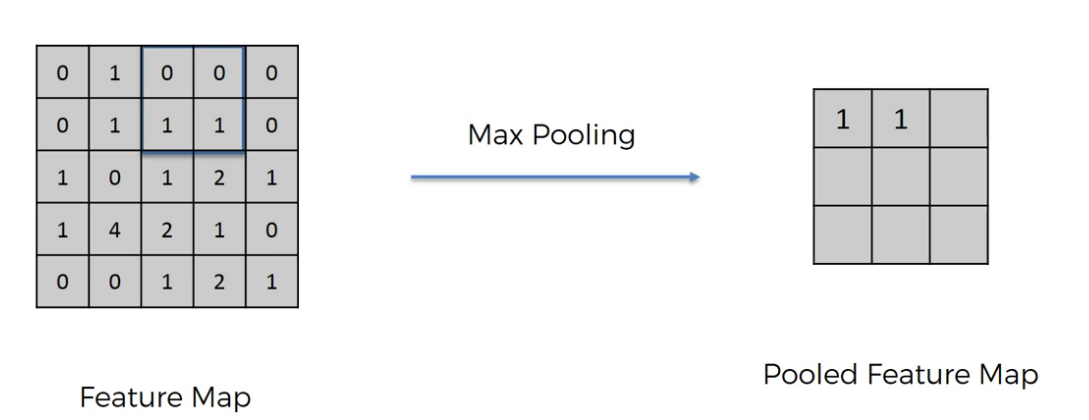
A rectifier activation function is placed here in order to promote nonlinearity of the system. The rectifier functions off then on behaviour turns parts of the image into almost binary states 0-1 with no fading in or out, making it easier to recognise images.

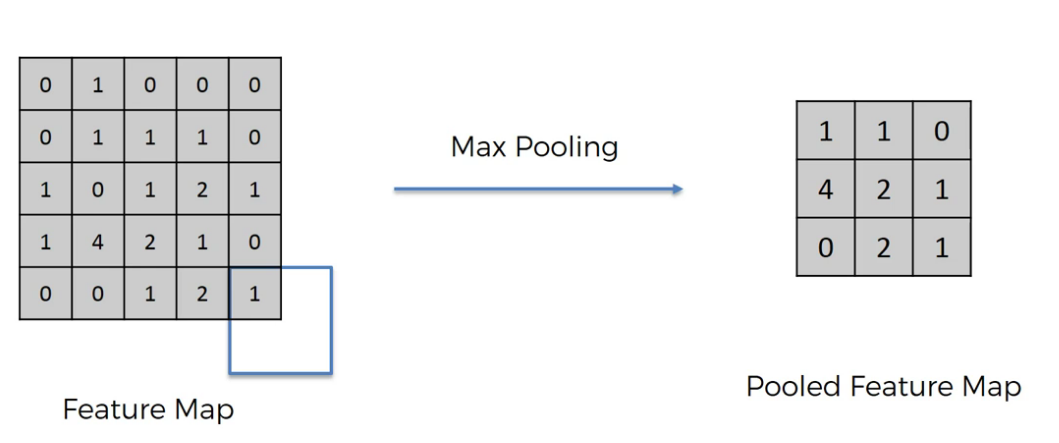
**Max Pooling**

To do with dealing with rotated images, or features which may be in-exact.

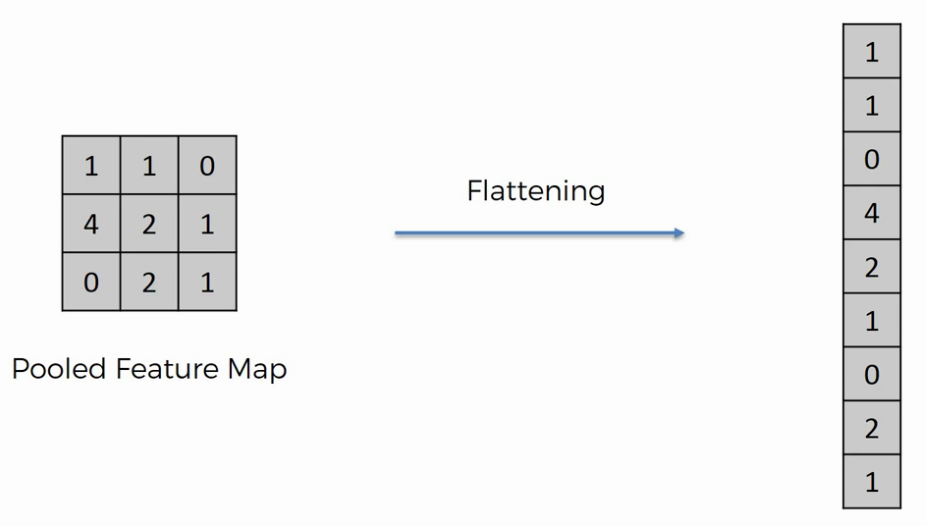
It starts with the feature map, (which is the output of step 1, including the relu layer). That is; the input and feature detector convolution. On this feature map we select a 2x2 ‘cursor’ of 4 boxes, and slide in by <stride> units left to right then down. (remember the strides ‘stride’ downwards aswell as across.) For every cursor resting spot we take the highest value number and put it into a ‘pooled’ feature map.

By doing this we eliminate 75% of the data. What actually happens is by taking the highest number in every 4 square, the ‘detection’ is the same wherever in those 4 squares the feature appears. This means that features can be slightly distorted, rotated etc and still be detected.





**Flattening:**



**Full connection:**

The flattened layer is inserted as inputs into a neural network. The neural network trains itself to recognise these features against a taught output of what it’s looking at. The back propogatation not only mutates the weights, it also mutates the feature detector so that what it’s looking for in each image slowly changes to what differentiates the two images.