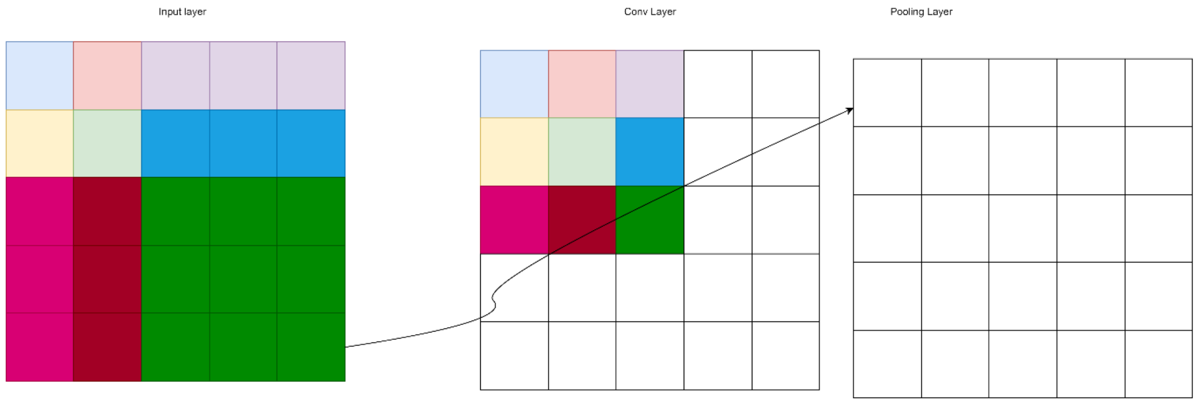
2. Neural Network Theory

Suppose we have convolutional network with architecture:

- Input layer RGB of image size 256×256
- Convolutional Layer with 32 feature maps and filter size 3×3
- Pooling layer with stride 2 and pooling group of size 3×3



To determine the size of input region that influences the activation of single unit, receptive field, in pooling layer, first let's see the first layer convolution layer.

On convolutional layer, input is process using filter of size 3×3 and resulting in output of size 254×254. Each output region is influenced by kernel of size 3×3 therefore receptive field in this layer is 3×3.

Next, we move into pooling layer. In this layer the single unit in pooling layer is obtained with pooling group of 3×3. This mean output of previous layer is the input on this layer.

Now with pooling group of size 3×3, one single unit is produced. But then input of this layer is no other than output of previous layer with original input of size 5×5. Thus, size of receptive field of pooling layer is 5×5.

3. Neural Network Theory

Given target function representation as

$$y_d = w_0 + w_1 x_1 + \cdots + w_n x_n$$

Suppose we have training example $d \in D$ with target output t_i we have cost function:

$$E = \sum_{d \in D} (t_d - y_d)^2$$

we express y_1 in form of target function

$$y_d = w_0 + w_1 x_1 + \cdots + w_n x_n$$

Gradient descent algorithm is minimizing cost function by adjusting weigh, so we have

$$\partial w_j = -\beta(\frac{\partial E}{\partial w_j})$$

$$\frac{\partial E}{\partial w_j} = \frac{(\partial \sum_{d \in D}(t_d - y_d)^2)}{\partial w_j}$$

$$= \sum_{d \in D} 2\,(t_d - y_d)\left(\frac{\partial(t_d - y_d)}{\partial w_j}\right)$$

$$= \sum_{d \in D} 2\,(t_d - y_d)\left(-\frac{\partial y_d}{\partial w_j}\right)$$

$$= -\sum_{d \in D} 2\,(t_d - y_d)\left(\frac{\partial(w_0 + w_1 x_{1d} + \cdots + w_d x_{jd} + \cdots + w_n x_{nd})}{\partial w_j}\right)$$

$$= -\sum_{d \in D} 2\,(t_d - y_d)\,x_{jd}$$

Thus $\partial w_j = -\beta\left(\frac{\partial E}{\partial w_j}\right) = -\beta\left(-\sum_{d \in D} 2\,(t_d - y_d)\,x_{jd}\right) = \beta \cdot -\sum_{d \in D} 2\,(t_d - y_d)\,x_{jd}$