CNNs

- Images generally have a high dimensionality.
- Think of an image - say 1080p image.

  e.g.- 1080p = $1920 \times 1080 \times 3$

  $\approx$ 6 million features

- If you treat each pixel as an independent input variable, the input dimensionality will blow up. (weights)

Equivariance

- A fn. $f(\cdot)$ is equivariant if for a transformation T :

  $$f(T(x)) = T(f(x))$$

- In CNNs, we see this principle.

Invariance

- A fn. $f(\cdot)$ is invariant to a transformation T if :

  $$f(T(x)) = f(x)$$

We have the output to a particular receptive field as :

$$y = \mathrm{Relu}(w^T x + w_0)$$

Lagranian

$$f(x) = w^T x$$
$$g(x) = x^T x - c = 0$$
$$L(x, \lambda) = w^T x - \lambda(x^T x - c)$$
$$\nabla_x L = w - 2\lambda x = 0$$
$$\Rightarrow x = \tfrac{1}{2\lambda} w$$
$$x = \alpha w \quad (\alpha = \tfrac{1}{2\lambda})$$

we define Convolution as :

- $I(j, k) \rightarrow$ Image pixels
- $K(l, m) \rightarrow$ Filter pixels

$$C(j, k) = \sum_l \sum_m I(j+l, k+m)\, K(l, m)$$

Also $C = I \circledast K$ (notation)

Padding

- The convolution map is smaller than the orginal image.

  Image $\rightarrow J \times K$

  Kernel $\rightarrow M \times M$

  $\Rightarrow (J-M+1) \times (K-M+1)$

  Feature-Map

$$\cancel{J} - M + 1 + 2P = \cancel{J}$$
$$2P + 1 = M$$
$$\boxed{P = (M-1)/2}$$

$$K - M + 1 + 2P = K$$
$$2P + 1 = M$$
$$\boxed{P = (M-1)/2}$$

- $P = 0$ is no-padding. called *valid* padding.
- As seen above to achieve the same dimension of the feature map as of the image the we need $\left[P = \frac{M-1}{2}\right]$. This is called <u>same</u> padding.

Strides

- Input : J
- Kernel : M
- Padding : P (total len $= J + 2P$)

Index : 0 to $J + 2P - 1$
- First valid start index $= 0$
- Last valid start index $= J + 2P - M$

with stride S, kernel start positions are :

$$0, S, 2S, 3S, \ldots, kS \leq J + 2P - M$$

$L =: J + 2P - M$. The largest k is $\left\lfloor \frac{L}{S} \right\rfloor$.

10.2.5 Multi-dimensional convolution

- Input Tensor $= J \times K \times C$ no. of channels
- Kernel $= M \times M \times C$

(This will have <u>M^2C</u> weights, regardless of the input size).

- To build more flexible models, we can include multiple such filters.
- Kernel (multidimensional) $= M \times M \times C_{in} \times C_{out}$
  (This will have <u>$(M^2 C_{in} + 1) C_{out}$</u> weights

10.2.6 Pooling

- Max Pooling
- Average Pooling