# GNR 638 Assignment 2


Amruta Mahendra Parulekar (20d070009)
Hemant Hajare (20d070037)


### *Question 1*

*Calculation of number of multiplications for normal convolution vs depthwise separable convolution....Input 12 x 12 x 3, want output 8 x 8 x 256*

**Normal Convolution:**

Given:

- Input size: 12x12x3
- Output size: 8x8x256

Assuming:

- 5x5 kernel; stride 1; no padding; 256 such filters; 3 channels
- 3 input channels; 256 output channels

In normal convolution, each output pixel is computed by taking a dot product between the kernel weights and the corresponding input region.
So, for each output pixel, 5x5x3 multiplications (dot product) for each of the 256 filters.
Since we have 8x8 output pixels, the total number of multiplications would be:
Multiplications=8x8x256x(5x5x3) =1,228,800

**Depth-wise Separable Convolution:**

Given:

- Input size: 12x12x3
- Output size: 8x8x256

Assuming:

- 5x5 kernel; stride 1; no padding; 3 channels
- 3 input channels; 256 output channels

In a depthwise separable convolution, we first apply a depthwise convolution (convolve each input channel separately with its corresponding kernel) followed by a pointwise convolution (use 1x1 convolutions to mix the output channels)
For the depthwise convolution, assuming the same kernel size 5x5 and 3 input channels, we need 5x5x3 multiplications per output pixel.
For the pointwise convolution, we have 3x256 multiplications for each output pixel (assuming we use 1x1 kernels).
Since we have 8x8 output pixels, the total number of multiplications would be:
Multiplications=(8x8x5x5x3)+(8x8x3x256)=53952 (significantly lowered computation)

## *Question 2*

*The loss function if the relation between the output and input variables can be described through a Poisson distribution.*

**The Poisson distribution**

$$P_X(k) = \frac{e^{-(\lambda t)} * (\lambda t)^k}{k!} = Poisson(\lambda t)$$

If the relationship between the output and input variables follows a Poisson distribution, the typical loss function used in this scenario is the Poisson loss function.

**The Poisson loss function.**

The Poisson loss function is often used in count data modeling, where the output represents counts of events occurring within a fixed interval of time or space.

It is derived from the Poisson probability distribution, which is commonly used to model the number of events occurring in a fixed interval of time or space, given a constant rate of occurrence and independence of events.

Mathematically, the Poisson loss function can be expressed as:

$L(z, y) = z - y \cdot \log(z) + \log(y!)$

Where:

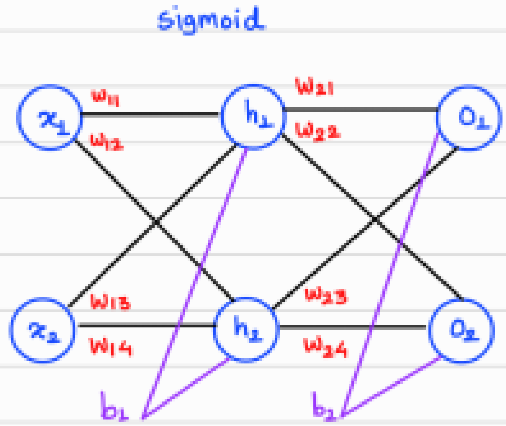
- z is the predicted value.
- $y$ is the actual value.

This loss function measures the discrepancy between the predicted counts and the actual counts while considering the characteristics of the Poisson distribution.

**A weighted mean-squared error**

A normal MSE loss will not work for this problem because output has a skew as compared to input. However, a weighted mean squared error will work to preserve the skew.

$$\mathrm{WMSE} = \frac{1}{n} \frac{\sum_{i=1}^{1} \mathrm{weights}_i \left(\widehat{\mathrm{predicted}}_i - \mathrm{actual}_i\right)^2}{\sum_{1=1}^{n} \mathrm{weights}_i}$$




# ANALYTIC GRADIENT CALCULATION

$$h = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \sigma\left\{ \begin{bmatrix} w_{11} & w_{13} \\ w_{12} & w_{14} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_1 \end{bmatrix} \right\} = \sigma(W_1 x + b_1)$$

$$O = \begin{bmatrix} O_1 \\ O_2 \end{bmatrix} = \sigma\left\{ \begin{bmatrix} w_{21} & w_{23} \\ w_{22} & w_{24} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} b_2 \\ b_2 \end{bmatrix} \right\} \quad \sigma(W_2 h + b_2)$$

$$O = \sigma(W_2\,\sigma(W_1 x + b_1) + b_2)$$

Writing explicitly

$$O_1 = \sigma\Big(w_{21}\,\sigma(w_{11}x_1 + w_{13}x_2 + b_1) + w_{23}\,\sigma(w_{12}x_1 + w_{14}x_2 + b_1) + b_2\Big)$$

$$O_2 = \sigma\Big(w_{22}\,\sigma(w_{11}x_1 + w_{13}x_2 + b_1) + w_{24}\,\sigma(w_{12}x_1 + w_{14}x_2 + b_1) + b_2\Big)$$

$$E = \frac{1}{2}(t_1 - O_1)^2 + \frac{1}{2}(t_2 - O_2)^2$$

Take

$w_{11} = 0.5$ $\quad w_{21} = 0.5$ $\quad x_1 = 1$ $\quad t_1 = 0.5$ $\quad b_1 = 1$

$w_{12} = 0.5$ $\quad w_{22} = 0.5$ $\quad x_2 = 1$ $\quad t_2 = 0.5$ $\quad b_2 = 1$

$w_{13} = 0.5$ $\quad w_{23} = 0.5$

$w_{14} = 0.5$ $\quad w_{24} = 0.5$

$$A_1 = w_{21}\sigma(w_{11}x_1 + w_{13}x_2 + b_1) + w_{23}\sigma(w_{12}x_1 + w_{14}x_2 + b_1) + b_2$$
$$A_2 = w_{22}\sigma(w_{11}x_1 + w_{13}x_2 + b_1) + w_{24}\sigma(w_{12}x_1 + w_{14}x_2 + b_1) + b_2$$
$$B = w_{11}x_1 + w_{13}x_2 + b_1$$
$$C = w_{12}x_1 + w_{14}x_2 + b_1$$

$$O_1 = \sigma(A_1) \qquad O_2 = \sigma(A_2)$$
$$A_1 = w_{21}\sigma(B) + w_{23}\sigma(C) + b_1 \qquad A_2 = w_{22}\sigma(B) + w_{23}\sigma(C) + b_2$$

$$\frac{dE}{dw_{11}} = (t_1 - O_1)\frac{dO_1}{dw_{11}} + (t_2 - O_2)\frac{dO_2}{dw_{11}}$$

$$= (t_1 - O_1)\frac{dO_1}{dA_1}\cdot\frac{dA_1}{\partial w_{11}} + (t_2 - O_2)\frac{dO_2}{dA_2}\cdot\frac{dA_2}{\partial w_{11}}$$

$$= (t_1 - O_1)\frac{dO_1}{dA_1}\frac{dA_1}{dB}\cdot\frac{dB}{dw_{11}} + (t_2 - O_2)\frac{dO_2}{dA_2}\frac{dA_2}{dB}\frac{dB}{dw_{11}}$$

$$= (t_1 - O_1)\sigma(A_1)(1-\sigma(A_1))\, w_{21}\sigma(B)(1-\sigma(B))\, x_1$$
$$+ (t_2 - O_2)\sigma(A_2)(1-\sigma(A_2))\, w_{22}\sigma(B)(1-\sigma(B))\, x_1$$

$$\frac{dE}{dw_{21}} = (t_1 - O_1)\frac{dO_1}{dw_{21}} + (t_2 - O_2)\frac{dO_2}{dw_{21}}$$
$$= (t_1 - O_1)\frac{dO_1}{dA_1}\cdot\frac{dA_1}{dw_{21}} + (t_2 - O_2)\frac{dO_2}{dA_2}\cdot\frac{dA_2}{dw_{21}}$$

$$= (t_1 - O_1)\sigma'(A_1)\sigma(B) + 0$$

$$= (t_1 - O_1)\sigma(A_1)(1-\sigma(A_1))\sigma(B)$$

In this case

$$B = 0.5(1) + 0.5(1) + 1 = 2$$
$$C = 0.5(1) + 0.5(1) + 1 = 2$$
$$A_1 = 0.5\,\sigma(2) + 0.5\,\sigma(2) + 1 = 1.88$$
$$A_2 = 0.5\,\sigma(2) + 0.5\,\sigma(2) + 1 = 1.88$$
$$O_1 = \sigma(1.88) = 0.87$$
$$O_2 = \sigma(1.88) = 0.87$$

$$\frac{dE}{dw_{11}} = (0.5-0.87)\cdot 0.87(1-0.87)\,0.5\cdot 0.88(1-0.88)\cdot 1 + (0.5-0.87)\cdot 0.87(1-0.87)\,0.5\cdot 0.88(1-0.88)\cdot 1$$
$$= (-0.37 \times 0.87 \times 0.13 \times 0.5 \times 0.88 \times 0.12) \times 2$$
$$= -2.209 \times 10^{-3} \times 2$$
$$= \boxed{-4.418 \times 10^{-3}} \approx -0.00443181$$

$$\frac{dE}{dw_{21}} = (0.5-0.87)\,0.87(1-0.87)\,0.88$$
$$= -0.37 \times 0.87 \times 0.13 \times 0.88$$
$$= \boxed{-0.036} \approx -0.0371$$

The values in red are values obtained by backpropogation