Homework 8   Sairam Srivatsavai

SSR226.

① Problem Solutions

Size of the Output of a Convolutional layer is given by

$$Size(O) = \frac{I - K + 2P}{S} + 1$$

I = Size of input image.

K = Size (width) of Kernels used in the Conv Layer,

N = Number of Kernels

S = Stride of Convolution operation.

P = Padding.

Size of a Max Pool layer.

$$O = \frac{I - P_s}{S} + 1$$

P_s = Poolsize.

Number of Parameters of a Convolution Layer.

$$W_c = K^2 \times C \times N$$

$$B_c = N$$

$$P_c = W_c + B_c.$$

W_c = Number of weights of the conv layer

B_c = Number of Biases of the Conv Layer

P_c = Number of parameters of the Conv Layer

K = Size (width) of Kernels used in Conv Layer

N = Number of Kernels

C = Number of Channels of the input image.

Number of parameters of a fully connected (FC) layer connected to a Conv Layer.

$$W_{cf} = O^2 \times N \times F$$

$$B_{cf} = F$$

$$P_{cf} = W_{cf} + B_{cf}$$

W_{cf} = Number of weights of a FC layer which is connected to a Conv layer.

B_{cf} = Number of biases of a FC Layer which is connected to a conv layer.

O = Size of the output image of the previous Conv layer

N = Number of Kernels in the previous Conv Layer.

F = Number of neurons in the FC Layer

→ Input = $64 \times 64 \times 3$

First convolution layer = $5 \times 5 \times 3 \times 32$.

__First Layer Convolution:__

$$\text{Output} = \frac{I - K + 2P}{S} + 1$$

$I = 64 \quad K = 5 \quad P = 0 \quad S = 1$

$$= \frac{64 - 5 + 2\times 0}{1} + 1$$

$= 59 + 1 = 60$.

$$\boxed{CNV_1 = 60 \times 60 \times 32}$$

No of trainable parameters =

$C = 3 \quad N = 32 \quad K = 5$

$W_c = K^2 \times C \times N = 5^2 \times 3 \times 32 = 2400$

$B_c = N = 32$

$$\boxed{P_c = W_c + B_c = 32 + 2400 = 2432}$$

blem.

Max Pool Layer 2x2

$$\text{Output image size} = \frac{I - P_s}{S} + 1$$

For 2D Max pooling

$$O = \frac{I}{2}$$

$$O = 30 \times 30 \times 32$$

There are no training parameters involved in MaxPool layer.

Convolutional Layer 2:

Convolutional $= 3 \times 3 \times 32 \times 64$ Stride $= 2$

$$\text{Output image size} = \frac{I - K + 2P}{S} + 1$$

$I = 30$ $K = 3$ $P = 0$ $S = 2$

$$= \frac{30 - 3 + 2 \times 0}{2} + 1$$

$$= 14$$

$$\text{Output} = 14 \times 14 \times 64.$$

Trainable Parameters =

$W_c = K^2 \times C \times N = 9 \times 32 \times 64$

$N = 64$ $C = 32$

$B_c = N = 64$

$P_c = W_c + B_c = 18496.$

Fully Connected Layer :

~~Number of Neuron in FC layer = 25×25×64 = 40000~~

$O = 25$

$N = 64$

$F = 10$

$$W_{CF} = O^2 \times N \times F$$

$$= 14 \times 14 \times 64 \times 10$$

$$= 125440$$

$$B_{CF} = F$$

$$= 10$$

$$P_{CF} = W_{CF} + B_{CF}$$

$$= 125440 + 10$$

$$= 125450$$

Total trainable parameters of Complete Architecture

$$= 146,378$$

The dimension of the weight Matrix of FC is = (12544, 10)

(2) Problem.

For RNN

Input Sequence $\{x_t\}$ with dimension n,

Hidden State $\{h_t\}$ with dimension K,

Output Sequence $\{y_t\}$ with dimension m,

The Number of trainable parameters =

$$= K(K+m) + K.$$

$$= K^2 + Km + K$$

(3) Problem

Let $0 < p, q < 1$ and $\epsilon = \log p - \log q$.

Show that

$$\frac{p-q}{p} = 1 - e^{-\epsilon}$$

Applying log on Both Sides

$$\cancel{\log}\left(\frac{p-q}{p}\right) = \cancel{\log}(1 - e^{-\epsilon})$$

$$\frac{p}{p} - \frac{q}{p} = 1 - e^{-\epsilon}.$$

$$\cancel{1} - \frac{q}{p} = \cancel{1} - e^{-\epsilon}$$

$$+\frac{q}{p} = +e^{-\epsilon}$$

Apply log on both Sides

$$\log\frac{q}{p} = \log(e^{-\epsilon})$$

$$\log q - \log p = -\epsilon \qquad \boxed{-\epsilon = -\epsilon}$$

$$\epsilon + O(\epsilon^2) = 1 - e^{-\epsilon}$$

$$\log p - \log q + \frac{\partial^2}{\partial x^2}\epsilon$$

$$\log p - \log q + \frac{\partial^2}{\partial x^2}\cancel{(\log p - \log q)} = 1 - e^{-\epsilon}.$$

$$\log p - \log q + O(\epsilon + \epsilon) = 1 - e^{-\epsilon}.$$

~~The KL divergence $D_{KL}(p(x)\|q(x)) = \int p(x) \log \frac{p(x)}{q(x)} dx$ is~~

~~$D_{KL}(p(x)\|q(x))$ is the expected amount of Information~~

ϵ is the error

$$\log p - \log q + O(\epsilon + \epsilon) = 1 - e^{-\epsilon}$$

Applying taylor series

$$\log p - \log q + O(\epsilon + \epsilon) = 1 - e^{-(\log p - \log q)}$$

The KL divergence $D_{KL}(p(x)\|q(x)) = \int p(x) \log \frac{p(x)}{q(x)} dx$ is the Continous form.

KL divergence gives the difference between the two probablity distribution.

$$= \int p(x)(\log p(x) - \log q(x)) dx$$

$$= \int p(x)\, \epsilon\, dx.$$

ϵ is the approximate error. So to if the error is zero only if both the distribution are same. and Probablitties are zero. This can be used to find error function.