Computer Vision
Assignment-1

Theory

Ans.1:- Given: No. of Classes = K
True Class Label as One-hot encoded vector $y \in \{0,1\}^K$, $\|y\|_1 = 1$
Prob. Distribution over class
$q = [q_1, q_2, q_3 \ldots q_K]$, $\sum_{i=1}^{K} q_i = 1$, $q_i \geq 0$

Label Smoothing :->
Add ϵ/K to each incorrect class
& Add $1-\epsilon+\epsilon/K$ to each correct class

(a) $H(p,q) = E_p[-\log q(X)]$

, where p & q are two arbitrary distributions;
X is a Random Variable;
H(p,q) is the cross-entropy b/w p & q.

It is given that, for label smoothing, regularisation is done by adding a small probability in order to reward the model for correct predictions & penalise it for the incorrect ones.
Mathematically,

$$p_i = \begin{cases} \epsilon/K & , i \neq y \\ 1-\epsilon+\epsilon/K & , i = y \end{cases}$$

$\therefore$ Cross Entropy Loss with Label Smoothing, $H(y,q)$:-

$$H(y,q) = -\sum_{i=1}^{K} p_i \log q_i$$

$$= -\left(\left(1-\epsilon+\frac{\epsilon}{K}\right)\log q_y\right) - \sum_{i \neq y} \frac{\epsilon}{K} \log q_i$$

$$= -\frac{\epsilon}{K}\sum_{i=1}^{K} \log q_i + (\epsilon - 1)\log q_y$$

$$= -\left(\frac{\epsilon}{K}\sum_{i=1}^{K} \log q_i - (1-\epsilon)\log q_y\right)$$

Classmate
Date
Page

(b) Effect of Label Smoothing :→

* Prevents Overfitting by adding a regularisation term to the loss function to penalise it for incorrect predictions.

* Since Label Smoothing is aimed at reducing overfitting, it acts as a regularisation technique where it ensure the model logits are constraine-d within a finite range, thus avoiding the chance of extreme probabilities, thereby making the model more generalised (less biased) which eventually results in efficient/stable training.

Ans 2 :- Given :→ Two Univariate Gaussian Distributions

$$p(x) = \mathcal{N}(\mu_p, \sigma_p^2)$$
$$q(x) = \mathcal{N}(\mu_q, \sigma_q^2)$$

(a) Cross Entropy b/w $p(x)$ & $q(x)$ as an expectation :→

$$H(p, q) = -E_{p(x)}[\log q(x)]$$

(b) We know that,

$$q(x) = \frac{1}{\sqrt{2\pi\sigma_q^2}} e^{-\frac{(x-\mu_q)^2}{2\sigma_q^2}} \quad \text{--- (1)}$$

, acc. to the formula of Gaussian Distribution

$$\therefore, H(p,q) = -E_{p(x)}[\log q(x)]$$

$$= -E_{p(x)}\left[\log\left(\frac{1}{\sqrt{2\pi\sigma_q^2}} e^{-\frac{(x-\mu_q)^2}{2\sigma_q^2}}\right)\right] \quad (\text{from } (1))$$

$$= -E_{p(x)}\left[-\frac{1}{2}\log(2\pi\sigma_q^2) - \frac{(x-\mu_q)^2}{2\sigma_q^2}\right]$$

$$= \frac{1}{2}\log(2\pi\sigma_q^2) + \frac{1}{2\sigma_q^2} E_{p(x)}[(x-\mu_q)^2]$$

$$= \frac{1}{2}\left[\log(2\pi\sigma_q^2) + \frac{1}{\sigma_q^2} E_{p(x)}[(x-\mu_q)^2]\right]$$

$$\because, p(x) = \mathcal{N}(\mu_p, \sigma_p^2)$$

$$\therefore, H(p,q) = \frac{1}{2}\left[\log(2\pi\sigma_q^2) + \frac{\sigma_p^2 + (\mu_p-\mu_q)^2}{\sigma_q^2}\right]$$

(c) For $\sigma_p = \sigma_q = \sigma$

$$H(p,q) = \frac{1}{2}\left[\log(2\pi\sigma^2) + \frac{\sigma^2 + (\mu_p - \mu_q)^2}{\sigma^2}\right]$$

$$= \frac{1}{2}\left[\log(2\pi\sigma^2) + 1 + \frac{(\mu_p - \mu_q)^2}{\sigma^2}\right]$$

If $p = q$,

$$H(p) = \frac{1}{2}\left[\log(2\pi\sigma^2) + 1\right] \quad (1)$$

Using (1), we can infer

$$H(p,q) = H(p) + \frac{(\mu_p - \mu_q)^2}{2\sigma^2}$$

Thus, $H(p,q)$ is a sum of $H(p)$ & a regularisation term when variances are equal.
The extent of regularisation increases as the size of the penalty increases when the mean difference b/w the two distribution increase.

Ans 3:- Given:→ Dilation factor/rate = r
↳ Specifies spacing b/w Kernel Elements

(a) 1 D Conv. with
Kernel size = K
Dilation factor = r
Layers of stacked convolutions = L

Let R_L be the Receptive field of layer, 'L'

We know that, the Receptive field of a layer (say 'L') is given by:

$$R_L = R_{L-1} + (K-S)\prod_{i=0}^{L-1} r_i$$

, where R_L : Receptive field of Layer 'L'
R_{L-1} : Receptive field of Layer 'L-1'
K : Kernel size
S : Stride
r_i : Dilation factor of Layer 'i'

∵, Stride is not mentioned, we assume a default value '1'

∴, $$R_L = R_{L-1} + (K-1)\prod_{i=0}^{L-1} r_i$$

As Dilation factor, 'r', grows layer-by-layer (exponentially),

$$R_L = 1 + (K-1)\prod_{i=0}^{L-1} 2^i \quad , \quad r = 2^0, 2^1, 2^2, \ldots, 2^{L-1}$$

$$R_L = 1 + (K-1)(2^L - 1) \qquad \boxed{\because \prod_{i=0}^{L-1} 2^i = 2^L - 1}$$

$\therefore$, Receptive field grows exponentially with respect to r (the Dilation factor/rate).

(b) 2D Conv. with
Kernel size = $K \times K$

Since, Receptive field does not change for either height or width,

For Height,

$$R_L = 1 + (K-1)(2^L - 1)$$

& For Width,

$$R_L = 1 + (K-1)(2^L - 1)$$

$\therefore$, Receptive field in 2D $\Rightarrow$

$$H \times W \Rightarrow R_L \times R_L = [1 + (K-1)(2^L - 1)] \times [1 + (K-1)(2^L - 1)]$$

$$= [1 + (K-1)(2^L - 1)]^2$$

$\therefore$, In 2D as well,
the Receptive field grows exponentially with respect to L.

(c) Computational Complexity of a Standard $K \times K$ convolution

Let the feature map be of the size $H \times W$

Assume Stride, s, be equal to 1.

$\therefore$, For every pixel in Output,
# of Multiplications $= K^2$

Total # of Operations $= K^2 \times (H \times W)$

Computational complexity of a Dilated $K \times K$ convolution

Let,
Stride $(s) = 1$
Dilation factor / rate $= r$

For every pixel in Output,
# of Multiplications $= K^2$, as only a few locations are jumped / skipped in convolution.

$\therefore$, Total # of Operations $= K^2 \times (H \times W)$

So, # of Multiply-Add Operations for both case $= K^2 \times (H \times W)$