classmate

Date
Page

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:- ~~(a)~~ 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

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

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

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

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

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

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

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

classmate

Date

Page

(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 $p = q$,

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

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.