Inception Distance

A Preprint

1 Main

Fréchet Inception Distance is an improvment above Inception score which is defined as

$$\mathrm{IS} = \exp(\mathbb{E}_x[KL(p(y|x)||p(y))]), \text{ where}$$

$$p(y|x) = \text{the conditional label distribution predicted for image x} \tag{1}$$

$$p(y) = \int p(y|x) \cdot p(x)dx \text{ the marginal class distribution over all generated images} \tag{2}$$

It measures how well the distribution of predicted labels for generated images

2 Fréchet Inception Distance

Let $p(x)$ denote the distribution of generated samples and $p_w(x)$ denote the distribution of real samples. These two distributions are equal (up to a non-measurable set) if and only if their expectations match for all functions $f(x)$ in a spanning set of the function space:

$$\int p(x)f(x)\,dx = \int p_w(x)f(x)\,dx.$$

These expectations are often approximated using moments or cumulants, where $f(x)$ is chosen to be a polynomial. The first moment corresponds to the mean, and the second moment to the covariance.

Instead of raw data x, one may consider the activations from the coding layer of a pretrained Inception model, which provide more semantically meaningful features. Assuming the coding activations follow a multivariate Gaussian distribution, the comparison between generated and real data reduces to comparing two Gaussians. The Fréchet distance (also known as Wasserstein-2 distance) between two Gaussians is then used, resulting in the Fréchet Inception Distance (FID):

$$\mathrm{FID}^2((m, C), (m_w, C_w)) = \|m - m_w\|_2^2 + \mathrm{Tr}\left(C + C_w - 2(CC_w)^{1/2}\right),$$

where (m, C) and (m_w, C_w) are the means and covariances of the coding layer features from the generated and real distributions, respectively.

Inception Score vs Fréchet Inception Distance

The Inception Score (IS) is defined as:

$$\mathrm{IS} = \exp\left(\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K} p(y_k \mid x_i)\log\frac{p(y_k \mid x_i)}{p(y_k)}\right) \text{ same as was presented in begining but without KL directly writen,}$$

where m is the number of generated samples and K is the number of classes.

Unlike the FID, which is a distance metric, the Inception Score is a score. To allow for direct comparison, a transformation of IS into a distance-like measure, termed the Inception Distance (IND), is proposed.

This transformation is possible because the Inception Score has a known upper bound. If $p(y_k \mid x_i) = 0$, the corresponding term in the sum is defined to be zero. Furthermore, the log term is bounded as:

$$\log\frac{p(y_k \mid x_i)}{p(y_k)} \le \log\frac{1}{1/m} = \log m.$$

Using this bound, the Inception Score can be bounded as follows:

$$\begin{aligned}\mathrm{IS} &= \exp\left(\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K} p(y_k \mid x_i)\log\frac{p(y_k \mid x_i)}{p(y_k)}\right)\\ &\le \exp\left(\log m \cdot \frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K} p(y_k \mid x_i)\right)\\ &= \exp\left(\log m \cdot \frac{1}{m}\sum_{i=1}^{m} 1\right) = m.\end{aligned}$$

The upper bound m is achieved when $m \le K$ and each sample is perfectly classified into a unique class.

Based on this, the Inception Distance (IND) is defined as:

$$\mathrm{IND} = m - \mathrm{IS}.$$

This value equals zero in the ideal case where every generated sample belongs to a different class and is classified with certainty. Like FID, IND increases with the level of disturbance or deviation from the ideal.