The probabilities of the points in latent space belonging to different classes

1   Introduction

In this project, we aim to sample random points from the latent space of a Variational Autoencoder (VAE), decode them using the VAE's decoder, and utilize the resulting hybrid images to train a CNN-based classifier. However, a key challenge is determining the probabilities of these latent space points being associated with specific classes. This document outlines the method for computing these probabilities.

Note: This is a specific part of the project. If you are unsure of its context or have not reviewed the project outline, it is recommended to first refer to the project outline in the README file.

2   Definition and Assumptions

Let us define the following distributions and assumptions:

$p(x)$:   The data distribution.

$p(z)$:   The latent distribution, which we assume to follow a standard normal distribution, $\mathcal{N}(0, 1)$.

$p(z|x)$:   The posterior distribution or encoder distribution, which we assume to be a normal distribution parameterized by the mean and variance produced by the encoder, $\mathcal{N}(\mu, \sigma^2)$.

For each class, we assume that the points in the latent space are normally distributed with a class-specific mean and variance, denoted as μ_c and σ_c^2, respectively. Additionally, for individual points in the latent space belonging to each class, we denote their mean and variance as $\mu_{c,i}$ and $\sigma_{c,i}^2$, respectively, which can be computed using the VAE encoder.

Given a point z, we aim to determine the probability of it belonging to a particular class c, denoted as $p(c|z)$. This probability will serve as the expected value for training the CNN-based classifier.

3   Methodology

We start by finding the means μ_c and variances σ_c^2 of each class. The mean can be computed as follows:

$$\mu_c = \frac{1}{N_c} \sum_{i=1}^{N_c} \mu_{c,i}, \tag{1}$$

where N_c is the number of points in the latent space belonging to class c.

Finding the mean for each class is straightforward. However, computing the variance is more involved. We use the **Law of Total Variance** to compute the variance for each class. The **Law of Total Variance** states:

$$\text{Var}[X] = \mathbb{E}[\text{Var}[X|Y]] + \text{Var}[\mathbb{E}[X|Y]], \tag{2}$$

where X and Y are random variables. For our case, this can be rewritten as:

$$\sigma_c^2 = \frac{1}{N_c} \sum_{i=1}^{N_c} \sigma_{c,i}^2 + \frac{1}{N_c} \sum_{i=1}^{N_c} (\mu_{c,i} - \mu_c)^2, \tag{3}$$

where N_c is the number of points in the latent space belonging to class c.

After we have the means and variances for each class, we are going to compute the probability $p(c|z)$ for each class. We can use Bayes' theorem to compute the probability of a point z belonging to class c as follows:

$$p(c|z) = \frac{p(z|c) \cdot p(c)}{\sum p(z|c_i) \cdot p(c_i)}, \tag{4}$$

where we assume that $p(c)$ is the same for all classes. Therefore, we can simplify the equation as follows:

$$p(c|z) = \frac{p(z|c)}{\sum p(z|c_i)}, \tag{5}$$

Since $p(z|c)$ is a normal distribution, we cannot compute the exact value of $p(z|c)$. Instead, we compute it as a small range as follows:

$$p(z|c) = p(z + \epsilon \leq x \leq z - \epsilon),$$

where ϵ is a small value and $x \sim \mathcal{N}(\mu_c, \sigma_c^2)$. Then, we normalize to the standard normal distribution:

$$p(z|c) = p\left(\frac{z + \epsilon - \mu}{\sigma} \leq \frac{x - \mu}{\sigma} \leq \frac{z - \epsilon - \mu}{\sigma}\right),$$

which simplifies to:

$$p(z|c) = \Phi\left(\frac{z - \mu}{\sigma} + \gamma\right) - \Phi\left(\frac{z - \mu}{\sigma} - \gamma\right),$$

where Φ is the standard normal distribution function, and $\gamma = \frac{\epsilon}{\sigma}$, which represents a small value that can be randomly selected.

We can then compute the probability $p(c|z)$ for each class by replacing back the formula above.