EVALUATING GENERATIVE NETWORKS USING GAUSSIAN MIXTURES OF IMAGE FEATURES

Anonymous authors
Paper under double-blind review

ABSTRACT

We develop a measure for evaluating the performance of generative networks given two sets of images. A popular performance measure currently used to do this is the Fréchet Inception Distance (FID). However, FID assumes that images featurized using the penultimate layer of Inception-v3 follow a Gaussian distribution. This assumption allows FID to be easily computed, since FID uses the 2-Wasserstein distance of two Gaussian distributions fitted to the featurized images. However, we show that Inception-v3 features of the ImageNet dataset are not Gaussian; in particular, each marginal is not Gaussian. To remedy this problem, we model the featurized images using Gaussian mixture models (GMMs) and compute the 2-Wasserstein distance restricted to GMMs. We define a performance measure, which we call WaM, on two sets of images by using Inception-v3 (or another classifier) to featurize the images, estimate two GMMs, and use the restricted 2-Wasserstein distance to compare the GMMs. We experimentally show the advantages of WaM over FID, including how FID is more sensitive than WaM to image perturbations. By modelling the non-Gaussian features obtained from Inception-v3 as GMMs and using a GMM metric, we can more accurately evaluate generative network performance.