TRY LPIPS

Discriminators (recognize a distinction; to differentiate) try to maximize the distance between two classes 0 and 1.

Critics try to minimize the distance between two distributions.

The sigmoid function transforms logits (raw, unnormalized output values made by last layer of network before activation) into probabilities. Sigmoid squashes between 0 and 1.

Wasserstein GANs introduce a different loss

* Called critics instead of discriminators.
* The sigmoid function is removed from the discriminator.

Measuring dissimilarity or the distance between probability distributions.

* Kullback-Leibler Divergence measures how one probability distribution *p* diverges from a second expected probability distribution *q*.
  + Asymmetric. P|Q works but Q|P produces different results.

Norm of the gradient of the critic needs to be less than or equal to 1 to satisfy the Lipschitz constraint.

||f||L<1 is a constraint on the critic, f, to enforce the Lipschitz constraint. This wants to keep the norm of the critic to less than or equal to 1.

We take a sample, x, from the real data probability distribution. We pass x through the critic, f. We repeat this for a sample from the fake data probability distribution. We then want to minimize the distance between these distributions.

The GAN loss takes a sample x from the real data. The real data is passed through the discriminator. We also sample latent noise z. We pass that through the generator to create a fake data sample which is then passed through the discriminator. We take the inverse of it and add both losses together as a GAN loss.

Qs

* Why do WGANs take longer training times?