ECE 239 AS

Lecture 2:

GAN setup: Has 2 parts, one is the generator network, and the other is the discriminator network. The discriminator network has to classify the manipulated datasets and correctly classify them.

The generator network has a random additive noise that is added to the images to make more samples of data. TO generate specific classes(variations of the same class, then we need to use a conditional GAN)

Generator: The mapping occurs int eh following manner : x’ = G(z) where G is a neural network that takes in the additive noise z. Its goal is to produce x’ from the training dataset p(x).

Discriminator: Performs mapping of D(x)- > [0,1]

Parameters: theta(G) and theta(D). As there are two different neural networks, they have two different loss functions. L(D) (theta(D), theta(G)) same for the generator network, however the caveat is that they can only change the parameters of their own model.

We treat this setup as a game, and the equilibrium that we reach from it is called the Nash Equilibrium. Both G and D know each other’s equilibrium strategies, and neither can gain anything by changing its strategy. This is a tuple (theta(D), theta(G)), where these parameters would be local minimas for their respective loss functions. Success probability of an Ideal GAN (0.5) on D(x).

Discriminator Loss function formula:

A close-up of a text

Description automatically generated

You can write this expectation as in integral for the calculation of the loss function with respect to pdata and pmodel

The estimator gets to change theta(d) to optimize this quantity(which is the loss function )

Loss will be zero if D(x) is 1 for all x ~ pdata and D(x’) = 0 for all x’ ~ pmodel , that is generated via x’ = G(z).

Need to set the derivative to zero for the L(D)(w.r.t. D(x) instead of theta(D).

Zero sum game: The sum of the losses of both the models is equal to zero, thus L(D) = -L(G). Solution for the zero-sum game is called the minimax solution(you minimize the maximum loss).

Thus, as the sign for the discriminator is negative, we do a gradient ascent step (argmax theta(d)).

Gradient descent step is done for the discriminator. Sample a minibatch of m noise samples {z(1), z(2)…..}. Perform SGD step and then perform the update steps.

KL Divergence: -> Way to compare the difference between the distributions of two variables

Ex~ p(x) log (P(x)/Q(x))

KL(P||Q) is always > = 0. When the distributions are the same, the KL divergence is 0. Also, it is non symmetric KL(P||Q) != KL(Q||P)

Summation Σ p(x) log(P(x)/Q(x)) over all x examples = Σ p(x) (-log(Q(x)/P(x))

E(f(x)) >= F(E(x))