Conditional gen Adversarial Nets Mirza (2014)

GANs were a novel way to train generative models

conditional version = feed data, y, we wish to condition

1 Introduction

GANs = Alt gen framework
↳ side step approx many intractable probabilistic computations

w/ Adversarial nets markov chains are not needed
↳ only backprop
↳ no inference required

unconditioned GANs = no control on modes of the data being generated

Conditioning allows direction in the generative process

Condition could be class labels

or even data from diff modality

2 Related work

3 Conditional Adversal Nets

gans

GANs learn generator dist $P_g(\cdot)$

gen builds map func from prior noise dist $P_z(z)$ to data space as $G(z; \theta_g)$

Disc = $D(x; \theta_d)$, Prob x came from traing D rather then P_g

G & D simul trained

Min min value func

Conditionals

GANs can be conditioned ~~so~~ if both gen & disc are conditioned on some extra info y

y can be any auxiliary information
- class labels, data from other modals

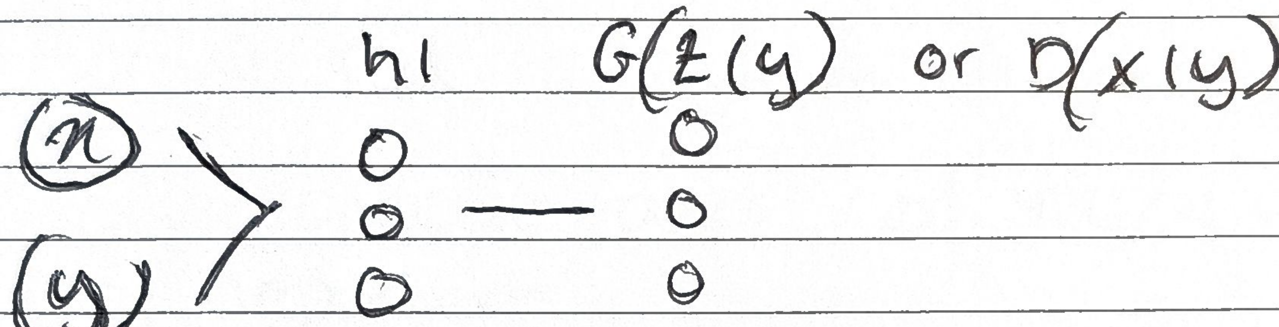
Perform ~~by~~ conditioning by feeding y into both Disc & gen as additional input layer

For the generator Noise $P_z(z)$ & y are combined in a hidden Representation

the framework is flexible on how this is achieved but paper uses single hidden layer of MLP to combined

for the discrim, x & y are presented as inputs to a function

↳ Paper uses MLP again



4 Experimental Results

conditional GAN on MNIST conditioned on their class labels hot-one encoded

gen net

z drawn from dist

$z + y$ mapped to hidden layer w/ relu

layer size 200 & 1000 respectively

& then to combined Relu layer of ~~1000~~ 1200

final sigmoid units to generate 784 MNIST sample

DISC Net

Map x to maxout layer 240 units, 5 Peis

Maps y to maxout 50 units, 5 Pieces

both hidden mapped to joint maxout
240 units, 4 Pieces

before fed to sigmoid layer

Sayer precise arch of Discrim is
not important
↳ as long as power = sufficent

Model trained using stochastic gradient D
- minibatch 100
- init lean rate of 0.1
  - Expo decreased to .000001

Dropout of 0.5 to both D & G

4.2 MultiModal

flickr = images w/ user generate labels
→ i.e tags

~~usage~~ user generated metadata is usn more descriptive & semantically close to human speech

→ rather then pure labels

need a way to normalize

→ conceptual word embedding are important here

the example in the paper ~~uses~~ Demos automated tagging of images

→ Multi-label predictions

generates a distribution of tag-vectors conditional on image features

- Pretrain convo imagenet model
- word rep embed

} Fix Mods & weigts During train