HOMEWORK 6

>>NAME HERE<<
>>ID HERE<<

Instructions: Use this latex file as a template to develop your homework. Submit your homework on time as a single pdf file. Please wrap your code and upload to a public GitHub repo, then attach the link below the instructions so that we can access it. Answers to the questions that are not within the pdf are not accepted. This includes external links or answers attached to the code implementation. Late submissions may not be accepted. You can choose any programming language (i.e. python, R, or MATLAB). Please check Piazza for updates about the homework. It is ok to share the results of the experiments and compare them with each other.

1 Implementation: GAN (50 pts)

In this part, you are expected to implement GAN with MNIST dataset. We have provided a base jupyter notebook (gan-base.ipynb) for you to start with, which provides a model setup and training configurations to train GAN with MNIST dataset.

(a) Implement training loop and report learning curves and generated images in epoch 1, 50, 100. Note that drawing learning curves and visualization of images are already implemented in provided jupyter notebook. (20 pts)

1

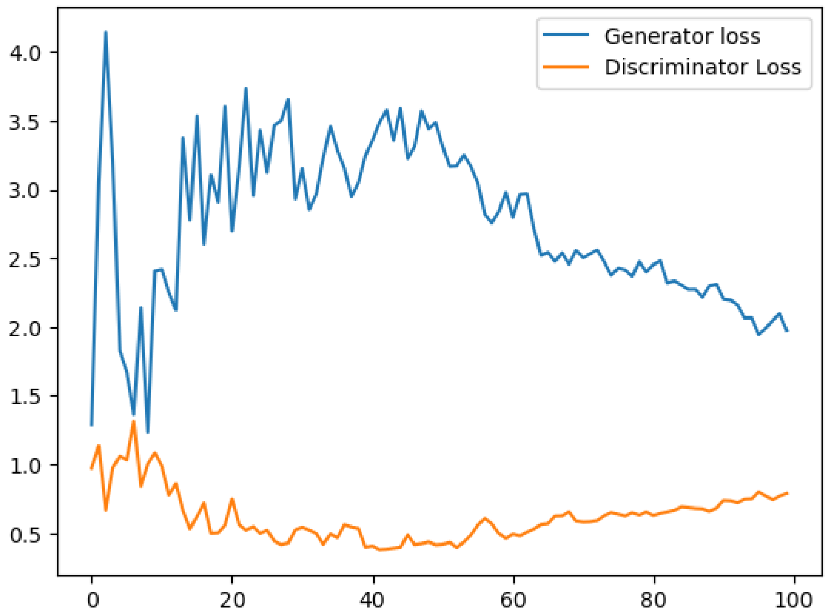
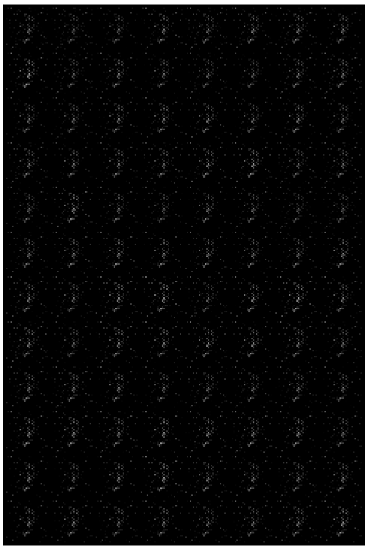
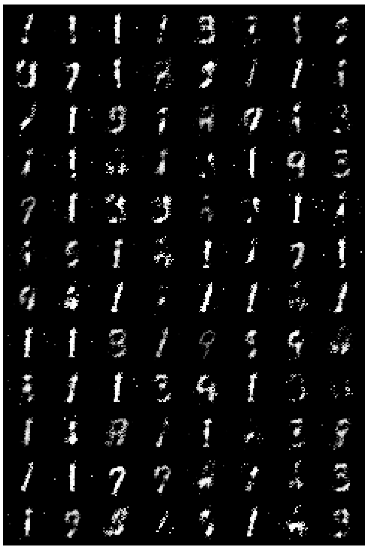


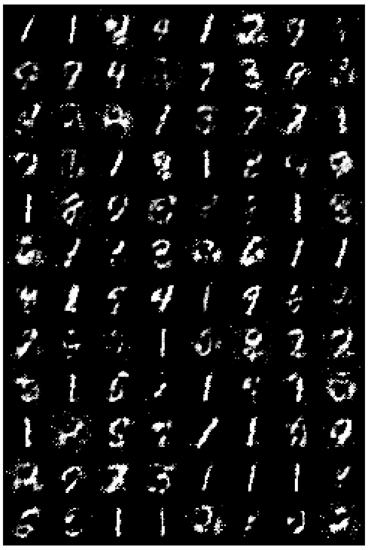
Figure 1: Learning curve



(a) Epoch 1



(b) Epoch 50



(c) Epoch 100

Figure 2: Generated images by G

(b) Replace the generator update rule as the original one in the slide,
"Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{n_z} \sum_{i=1}^{n_z} \log(1 - D(G(z^{(i)}))))$$

" , and report learning curves and generated images in epoch 1, 50, 100. Compare the result with (a). Note that it may not work. If training does not work, explain why it doesn't work.
You may find this helpful: https://jonathan-hui.medium.com/gan-what-is-wrong-with-the-gan-cost-function-6f594162ce01 (10 pts)

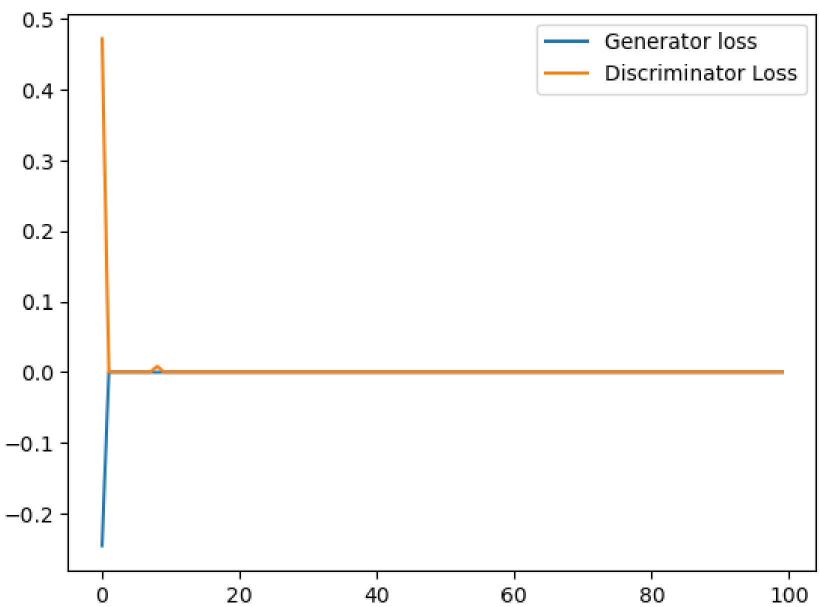
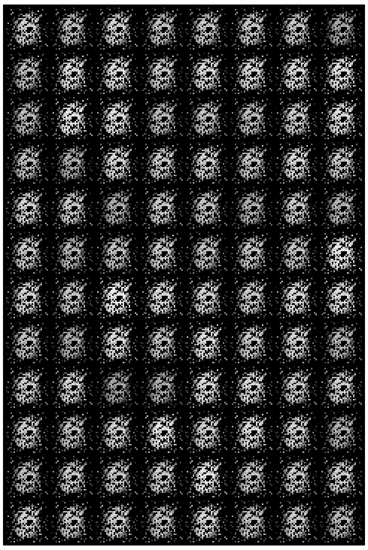
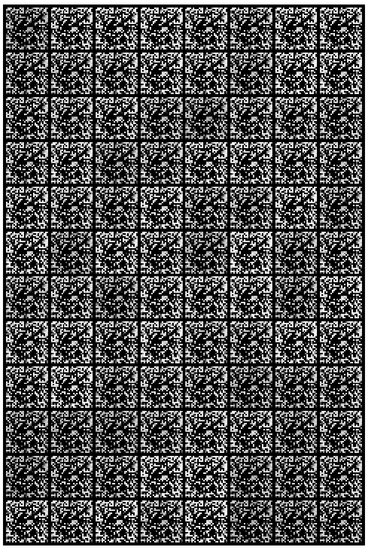


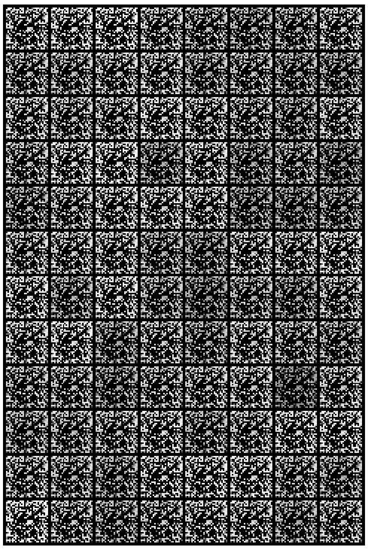
Figure 3: Learning curve



(a) Epoch 1



(b) Epoch 50



(c) Epoch 100

Figure 4: Generated images by G

As the performance of the discriminator improves, it will output lower and lower probabilities that a fake image is real. Since $D(G(z)) \to 0$, then $\log(1 - D(G(z))) \to 0$ which causes the gradients of the generator network to vanish.

(c) Except the method that we used in (a), how can we improve training for GAN? Implement that and report your setup, learning curves, and generated images in epoch 1, 50, 100. This question is an open-ended question and you can choose whichever method you want. (20 pts)

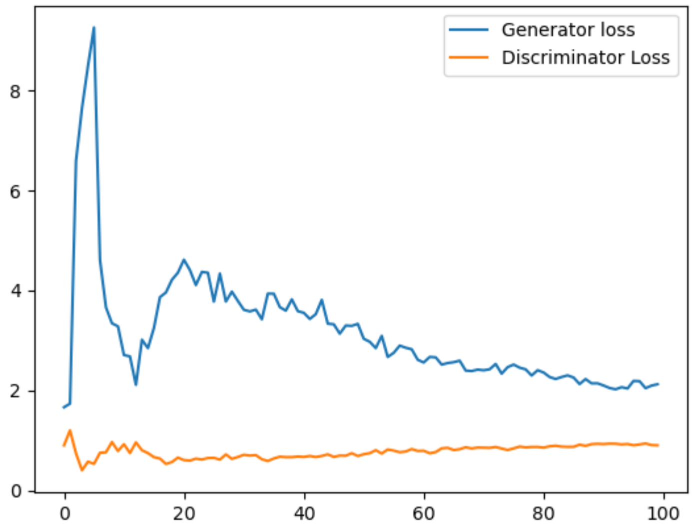


Figure 5: Learning curve

(a) Epoch 1 (b) Epoch 50 (c) Epoch 100

Figure 6: Generated images by G

One-sided label smoothing is a form a regularization for GANs that uses target labels that are slightly less than 1 (0.9 in this case) in order to avoid overconfidence in the discriminator. Overconfidence is a problem because if the discriminator learns to rely on a small subset of features to identify real images, the generator will exploit this and only adjust that subset of features, hindering learning in the long run.

2 Directed Graphical Model [25 points]

Consider the directed graphical model (aka Bayesian network) in Figure 7.

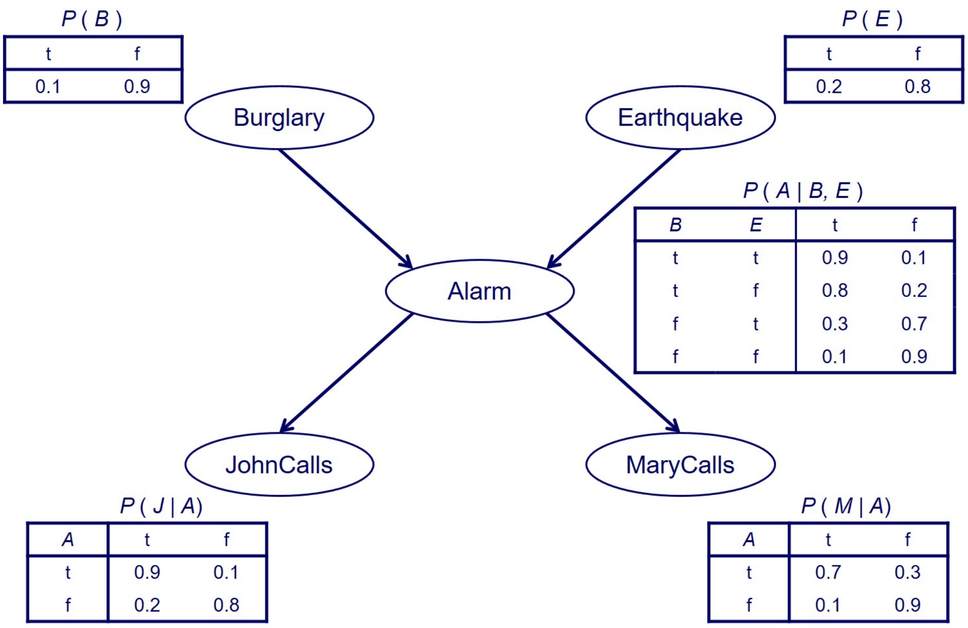


Figure 7: A Bayesian Network example.

Compute $P(B = t \mid E = f, J = t, M = t)$ and $P(B = t \mid E = t, J = t, M = t)$. (10 points for each) These are the conditional probabilities of a burglar in your house (yikes!) when both of your neighbors John and Mary call you and say they hear an alarm in your house, but without or with an earthquake also going on in that area (what a busy day), respectively.

We perform inference by enumeration using the probability tables provied.

$$
\begin{aligned}
P(b \mid \neg e, j, m) &= \frac{P(b, \neg e, j, m)}{P(\neg e, j, m)} \\
&= \frac{P(b, \neg e, j, m)}{P(b, \neg e, j, m) + P(\neg b, \neg e, j, m)} \\
&= \frac{P(b)P(\neg e)\sum_A P(A \mid b, \neg e)P(j \mid A)P(m \mid A)}{P(b)P(\neg e)\sum_A P(A \mid b, \neg e)P(j \mid A)P(m \mid A) + P(\neg b)P(\neg e)\sum_A P(A \mid \neg b, \neg e)P(j \mid A)P(m \mid A)} \\
&\approx 0.4107
\end{aligned}
$$

$$
\begin{aligned}
P(b \mid e, j, m) &= \frac{P(b, e, j, m)}{P(e, j, m)} \\
&= \frac{P(b, e, j, m)}{P(b, e, j, m) + P(\neg b, e, j, m)} \\
&= \frac{P(b)P(e)\sum_A P(A \mid b, e)P(j \mid A)P(m \mid A)}{P(b)P(e)\sum_A P(A \mid b, e)P(j \mid A)P(m \mid A) + P(\neg b)P(e)\sum_A P(A \mid \neg b, e)P(j \mid A)P(m \mid A)} \\
&\approx 0.00006203
\end{aligned}
$$

3 Chow-Liu Algorithm [25 pts]

Suppose we wish to construct a directed graphical model for 3 features X, Y, and Z using the Chow-Liu algorithm. We are given data from 100 independent experiments where each feature is binary and takes value T or F. Below is a table summarizing the observations of the experiment:

X	Y	Z	Count
T	T	T	36
T	T	F	4
T	F	T	2
T	F	F	8
F	T	T	9
F	T	F	1
F	F	T	8
F	F	F	32

1. Compute the mutual information $I(X, Y)$ based on the frequencies observed in the data. (5 pts)

$$
\begin{aligned}
I(X;Y) &= \sum_{x \in \{T,F\}} \sum_{y \in \{T,F\}} p(x,y) \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right) \\
&= \frac{2}{5} \log_2 \left(\frac{2/5}{1/4} \right) + \frac{1}{10} \log_2 \left(\frac{1/10}{1/4} \right) + \frac{1}{10} \log_2 \left(\frac{1/10}{1/4} \right) + \frac{2}{5} \log_2 \left(\frac{2/5}{1/4} \right) \\
&\approx 0.2781
\end{aligned}
$$

2. Compute the mutual information $I(X, Z)$ based on the frequencies observed in the data. (5 pts)

$$
\begin{aligned}
I(X;Z) &= \sum_{x\in\{T,F\}} \sum_{z\in\{T,F\}} p(x,z) \log_2 \left(\frac{p(x,z)}{p(x)p(z)} \right) \\
&= \frac{19}{50} \log_2 \left(\frac{19/50}{11/40} \right) + \frac{17}{100} \log_2 \left(\frac{17/100}{11/40} \right) + \frac{3}{25} \log_2 \left(\frac{3/25}{9/40} \right) + \frac{33}{100} \log_2 \left(\frac{33/100}{9/40} \right) \\
&\approx 0.1328
\end{aligned}
$$

3. Compute the mutual information $I(Z, Y)$ based on the frequencies observed in the data. (5 pts)

$$
\begin{aligned}
I(Y;Z) &= \sum_{y\in\{T,F\}} \sum_{z\in\{T,F\}} p(y,z) \log_2 \left(\frac{p(y,z)}{p(y)p(z)} \right) \\
&= \frac{9}{20} \log_2 \left(\frac{9/20}{11/40} \right) + \frac{1}{10} \log_2 \left(\frac{1/10}{11/40} \right) + \frac{1}{20} \log_2 \left(\frac{1/20}{9/40} \right) + \frac{2}{5} \log_2 \left(\frac{2/5}{9/40} \right) \\
&\approx 0.3973
\end{aligned}
$$

4. Which undirected edges will be selected by the Chow-Liu algorithm as the maximum spanning tree? (5 pts)

   The Chow-Liu algorithm will first select edge (Y, Z) and then select edge (X, Y).

5. Root your tree at node X, assign directions to the selected edges. (5 pts)

$$X \to Y \to Z$$

References