Unsupervised Learning: Generation
Creation

Drawing?

Writing Poems?
Creation

- Generative Models: https://openai.com/blog/generative-models/

What I cannot create, I do not understand.

Richard Feynman

Creation

Now

v.s.

In the future

Machine draws a cat

http://www.wikihow.com/Draw-a-Cat-Face
Generative Models

Component-by-component

Autoencoder

Generative Adversarial Network (GAN)
Component-by-component

• Image generation
  
  E.g. 3 x 3 images

  Can be trained just with a large collection of images without any annotation
Component-by-component

- Image generation
  - E.g. 3 x 3 images

Can be trained just with a large collection of images without any annotation
Practicing Generation Models: Pokémon Creation

- Small images of 792 Pokémon's
  - Can machine learn to create new Pokémon's?

Don't catch them! Create them!

- Source of image:
  http://bulbapedia.bulbagarden.net/wiki/List_of_Pokémon_by_base_stats_(Generation_VI)

Original image is 40 x 40
Making them into 20 x 20
Practicing Generation Models: Pokémon Creation

- **Tips (?)**
  - Each pixel is represented by 3 numbers (corresponding to RGB)
    - $R=50$, $G=150$, $B=100$
  - Each pixel is represented by a 1-of-N encoding feature

Clustering the similar color... 167 colors in total
Practicing Generation Models: Pokémon Creation

- Original image (40 x 40):
  [link](http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2016/Pokemon_creation/image.rar)

- Pixels (20 x 20):
  [link](http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2016/Pokemon_creation/pixel_color.txt)
  - Each line corresponds to an image, and each number corresponds to a pixel
  - [link](http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2016/Pokemon_creation/colormap.txt)

- Following experiment: 1-layer LSTM, 512 cells
Real Pokémon
Never seen by machine!

Cover 50%

It is difficult to evaluate generation.

Cover 75%
Pokémon Creation

Drawing from scratch

Need some randomness
PixelRNN


Real World
More than images ......


Generative Models

Component-by-component

Autoencoder

Generative Adversarial Network (GAN)
Auto-encoder

As close as possible

Randomly generate a vector as code
Review: Auto-encoder

-1.5

\[
\begin{bmatrix}
-1.5 \\
0
\end{bmatrix}
\]

0

1.5

\[
\begin{bmatrix}
1.5 \\
0
\end{bmatrix}
\]

-1.5

\[
\begin{bmatrix}
-1.5 \\
0
\end{bmatrix}
\]

2D

code

NN Decoder

NN Decoder

NN Decoder
Review: Auto-encoder
Auto-encoder

From a normal distribution

Minimize reconstruction error

VAE

Minimize

\[ \sum_{i=1}^{3} (\exp(\sigma_i) - (1 + \sigma_i) + (m_i)^2) \]
Why VAE?

**Intuitive Reason**

The variance of noise is automatically learned

What will happen if we only minimize reconstruction error?

\[ \sum_{i=1}^{3} (\exp(\sigma_i) - (1 + \sigma_i) + (m_i)^2) \]
Why VAE?

Intuitive Reason

We want $\sigma_i$ close to 0 (variance close to 1)

What will happen if we only minimize reconstruction error?

Minimize

$$\sum_{i=1}^{3} (\exp(\sigma_i) - (1 + \sigma_i) + (m_i)^2)$$

L2 regularization
Why VAE?

*Intuitive Reason*

decode → code → decode

code

code

noise

noise
Warning of Math
**Gaussian Mixture Model**

\[ P(x) = \sum_{m} P(m)P(x|m) \]

How to sample?

\[ m \sim P(m) \] (multinomial)

\[ x|m \sim N(\mu^m, \Sigma^m) \]

Each \( x \) you generate is from a mixture Distributed representation is better than cluster.
z \sim N(0, I)

x|z \sim N(\mu(z), \sigma(z))

z is a vector from normal distribution

Each dimension of z represents an attribute

$P(x) = \int_{\mathbb{R}^d} P(z)P(x|z)dz$

Even though z is from $N(0, I)$, $P(x)$ can be very complex

Infinite Gaussian
Maximizing Likelihood

\[ P(x) = \int_{x} P(z)P(x|z)dz \]

\[ L = \sum_{x} \log P(x) \]

\[ P(z) \text{ is normal distribution} \]

\[ x|z \sim N(\mu(z), \sigma(z)) \]

\[ \mu(z), \sigma(z) \text{ is going to be estimated} \]

Tuning the parameters to maximize likelihood \( L \)

We need another distribution \( q(z|x) \)

\[ z|x \sim N(\mu'(x), \sigma'(x)) \]
Maximizing Likelihood

\[ P(x) = \int P(z)P(x|z)dz \]

\[ L = \sum_x \log P(x) \]

\[ \log P(x) = \int q(z|x) \log P(x)dz \]

\[ = \int q(z|x) \log \left( \frac{P(z,x)}{P(z|x)} \right)dz = \int q(z|x) \log \left( \frac{P(z,x)q(z|x)}{q(z|x)P(z|x)} \right)dz \]

\[ = \int q(z|x) \log \left( \frac{P(z,x)}{q(z|x)} \right)dz + \int q(z|x) \log \left( \frac{q(z|x)}{P(z|x)} \right)dz \]

\[ \geq \int q(z|x) \log \left( \frac{P(x|z)P(z)}{q(z|x)} \right)dz \]

P(z) is normal distribution

\[ x|z \sim N(\mu(z), \sigma(z)) \]

\[ \mu(z), \sigma(z) \text{ is going to be estimated} \]

q(z|x) can be any distribution

Maximizing the likelihood of the observed x

KL(q(z|x)||P(z|x)) \geq 0

Lower bound \( L_b \)
Maximizing Likelihood

\[ \log P(x) = L_b + KL(q(z|x) || P(z|x)) \]

\[ L_b = \int q(z|x) \log \left( \frac{P(x|z)P(z)}{q(z|x)} \right) dz \]

Find \( P(x|z) \) and \( q(z|x) \) maximizing \( L_b \)

\( q(z|x) \) will be an approximation of \( p(z|x) \) in the end
Maximizing Likelihood

\[ P(x) = \int P(z)P(x|z)dz \]
\[ L = \sum_x logP(x) \]

\[ L_b = \int q(z|x)log\left(\frac{P(z,x)}{q(z|x)}\right)dz = \int q(z|x)log\left(\frac{P(x|z)P(z)}{q(z|x)}\right)dz \]

\[ = \int q(z|x)log\left(\frac{P(z)}{q(z|x)}\right)dz + \int q(z|x)logP(x|z)dz \]

\[-KL(q(z|x)||P(z))\]

P(z) is normal distribution
\[ x|z \sim N(\mu(z), \sigma(z)) \]
\[ \mu(z), \sigma(z) \text{ is going to be estimated} \]

Maximizing the likelihood of the observed x

\[ z|x \sim N(\mu'(x), \sigma'(x)) \]

\[ x \rightarrow \text{NN}' \]

\[ \mu'(x) \]

\[ \sigma'(x) \]
**Connection with Network**

Minimizing \( KL(q(z|x) || P(z)) \)

Minimizing \( \sum_{i=1}^{3} (exp(\sigma_i) - (1 + \sigma_i) + (m_i)^2) \)

Maximizing

\[
\int q(z|x) \log P(x|z) dz = E_{q(z|x)}[\log P(x|z)]
\]

This is the auto-encoder

(Refer to the Appendix B of the original VAE paper)
End of Warning
Pokémon Creation

Input → NN Encoder

m₁ m₂ m₃

σ₁ σ₂ σ₃

e₁ e₂ e₃

exp → +

10-dim

Training

10-dim

Pick two dim, and fix the rest eight

C₁ C₂ C₃

NN Decoder

10-dim

Output
Writing Poetry

Code Space

sentence \(\rightarrow\) NN Encoder \(\rightarrow\) code \(\rightarrow\) NN Decoder \(\rightarrow\) sentence

- i went to the store to buy some groceries.
- "come with me," she said.
- "don't worry about it," she said.

Ref: [http://www.wired.co.uk/article/google-artificial-intelligence-poetry](http://www.wired.co.uk/article/google-artificial-intelligence-poetry)
Problems of VAE

- It does not really try to simulate real images

VAE may just memorize the existing images, instead of generating new images.
Generative Models

Component-by-component

Autoencoder

Generative Adversarial Network (GAN)

Cifar-10

• Which one is machine-generated?

Ref: https://openai.com/blog/generative-models/
Yann LeCun’s comment

What are some recent and potentially upcoming breakthroughs in unsupervised learning?

Yann LeCun, Director of AI Research at Facebook and Professor at NYU
Written Jul 29 · Upvoted by Joaquin Quiñonero Candela, Director Applied Machine Learning at Facebook and Huang Xiao

Adversarial training is the coolest thing since sliced bread.

I’ve listed a bunch of relevant papers in a previous answer.

Expect more impressive results with this technique in the coming years.

What’s missing at the moment is a good understanding of it so we can make it work reliably. It’s very finicky. Sort of like ConvNet were in the 1990s, when I had the reputation of being the only person who could make them work (which wasn’t true).
The most important one, in my opinion, is adversarial training (also called GAN for Generative Adversarial Networks). This is an idea that was originally proposed by Ian Goodfellow when he was a student with Yoshua Bengio at the University of Montreal (he since moved to Google Brain and recently to OpenAI).

This, and the variations that are now being proposed is the most interesting idea in the last 10 years in ML, in my opinion.

Evolution

Butterflies are not brown
Butterflies do not have veins

Kallima inachus
The evolution of generation

NN Generator v1

Disriminator v1

Real images:

NN Generator v2

Disriminator v2

NN Generator v3

Disriminator v3
Basic Idea of GAN

• The data we want to generate has a distribution $P_{data}(x)$.
Basic Idea of GAN

• A generator G is a network. The network defines a probability distribution.

It is difficult to compute $P_G(x)$
We do not know what the distribution looks like.

https://blog.openai.com/generative-models/
Basic Idea of GAN

It can be proofed that the *loss the discriminator* related to *JS divergence*. 
Basic Idea of GAN

• **Next step:**
  • Updating the parameters of generator
  • To minimize the JS divergence

The output be classified as “real” (as close to 1 as possible)

Generator + Discriminator = a network

Using gradient descent to update the parameters in the generator, but fix the discriminator
GAN – 二次元人物頭像鍊成

Source of images: https://zhuanlan.zhihu.com/p/24767059
DCGAN: https://github.com/carpedm20/DCGAN-tensorflow
GAN – 二次元人物頭像鍊成

100 rounds
GAN – 二次元人物頭像鍊成

1000 rounds
GAN – 二次元人物頭像鍊成

2000 rounds
GAN – 二次元人物頭像鍊成

5000 rounds
GAN – 二次元人物頭像鍊成

10,000 rounds
GAN - 二次元人物頭像鍊成

20,000 rounds
GAN – 二次元人物頭像鍊成

50,000 rounds
Why GAN is hard to train?

回到演化的比喻 ......
Why GAN is hard to train?

\[ J S(P_{G0} || P_{data}) = \log 2 \]

\[ J S(P_{G50} || P_{data}) = \log 2 \]

Not really better ....

\[ J S(P_{G100} || P_{data}) = 0 \]
Using Wasserstein distance instead of JS divergence
Real poems: 床前明月光，疑似地上霜，舉頭望明月，低頭思故鄉。
WGAN – 唐詩鍊成

由 李仲翊 同學提供實驗結果

升雲白遲丹齋取，此酒新巷市入頭。黃道故海歸中後，不驚入得韻子門。
據口容章蕃翎翎，邦貸無遊隔將毬。外蕭曾臺遙出畧，此計推上呂天夢。
新來寶伎泉，手雪泓臺蓑。曾子花路魏，不謀散薦船。
功持牧度機邈爭，不躚官嬉牧涼散。不迎白旅今掩冬，盡蘸金祇可停。
玉十洪沄爭春風，溪子風佛挺橫鞋。盤盤稅焰先花齋，誰過飘鶴一丞憧。
海人依野庇，為阻例沉迥。座花不佐樹，弟闌十名儂。
入維當興日世瀕，不評皺。頭醉空其杯，駸園凋送頭。
鉢笙動春枝，寶叅潔長知。官爲宻爛去，絆粒薛一靜。
吾涼腕不楚，縱先待旅知。楚人縱酒待，一蔓飄聖猜。
折幕故癘應韻子，徑頭霜瓊老徑徑。尚錯春鐭熊悽梅，去吹依能九將香。
通可矯目鸚須浄，丹逕挈花一抵嫖。外子當目中前醒，迎日幽筆鈎弧前。
庭愛四樹人庭好，無衣服仍繡秋州。更怯風流欲鵎雲，帛陽舊據:uint婷儂。
Moving on the code space

Moving on the code space

- Ref: http://qiita.com/mattya/items/e5bfe5e04b9d2f0b0bd47
Text to Image


Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaolei Huang, Xiaogang Wang, Dimitris Metaxas, “StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks”, arXiv prepring, 2016

<table>
<thead>
<tr>
<th>Caption</th>
<th>Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>this flower has white petals and a yellow stamen</td>
<td><img src="image1.jpg" alt="Flower Images" /></td>
</tr>
<tr>
<td>the center is yellow surrounded by wavy dark purple petals</td>
<td><img src="image2.jpg" alt="Flower Images" /></td>
</tr>
<tr>
<td>this flower has lots of small round pink petals</td>
<td><img src="image3.jpg" alt="Flower Images" /></td>
</tr>
</tbody>
</table>

"red flower with black center"
Text to Image

• E.g. 根據文字敘述畫出動漫人物頭像

Red hair, long hair

Black hair, blue eyes

Blue hair, green eyes
Image-to-image Translation

Image-to-image Translation - Results
Cycle GAN
https://arxiv.org/abs/1703.10593
Disco GAN

(a) Learning cross-domain relations without any extra label

(b) Handbag images (input) & Generated shoe images (output)

(c) Shoe images (input) & Generated handbag images (output)

https://arxiv.org/abs/1703.05192
機械学習で美少女化～あるいはNEW GAME!の世界

http://qiita.com/Hiking/items/8d36d9029ad1203aac55
So many GANs
…… Just name a few

<table>
<thead>
<tr>
<th>Modifying the Optimization of GAN</th>
<th>Different Structure from the Original GAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>fGAN</td>
<td>Conditional GAN</td>
</tr>
<tr>
<td>WGAN</td>
<td>Semi-supervised GAN</td>
</tr>
<tr>
<td>Least-square GAN</td>
<td>InfoGAN</td>
</tr>
<tr>
<td>Loss Sensitive GAN</td>
<td>BiGAN</td>
</tr>
<tr>
<td>Energy-based GAN</td>
<td>Cycle GAN</td>
</tr>
<tr>
<td>Boundary-seeking GAN</td>
<td>Disco GAN</td>
</tr>
<tr>
<td>Unroll GAN</td>
<td>VAE-GAN</td>
</tr>
<tr>
<td>......</td>
<td>......</td>
</tr>
</tbody>
</table>
Acknowledgement

• 感謝 Ryan Sun 來信指出投影片上的錯字