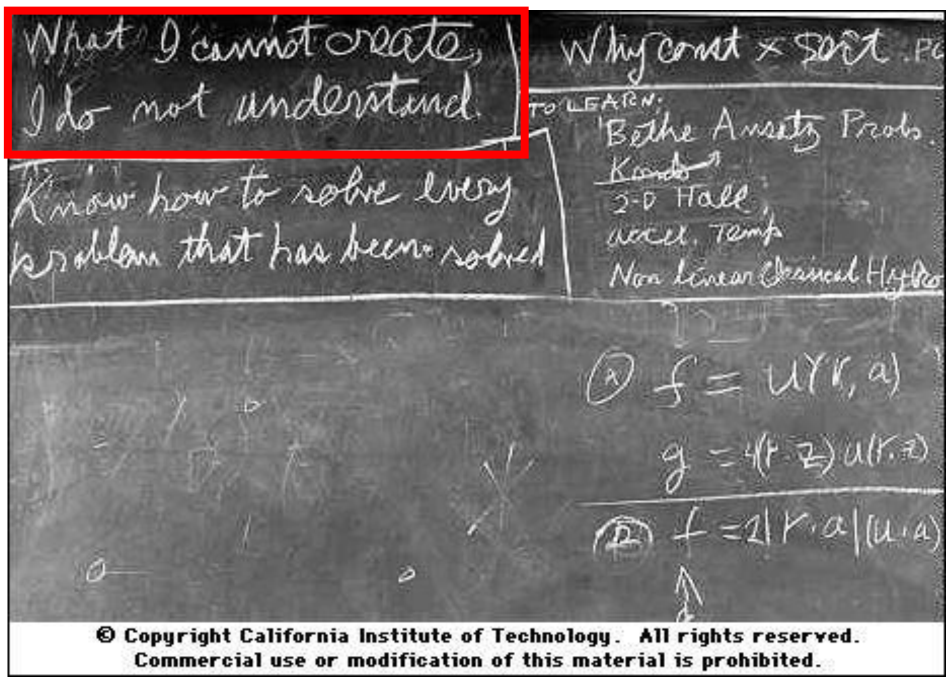
Unsupervised Learning: Generation

Creation

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

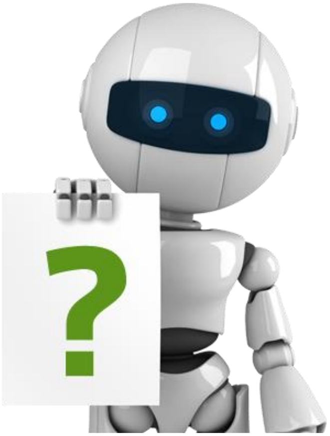


What I cannot create,
I do not understand.

Richard Feynman

https://www.quora.com/What-did-Richard-Feynman-mean-when-he-said-What-I-cannot-create-I-do-not-understand

Creation – Image Processing



Now

v.s.

In the future

Machine draws a cat

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

Generative Models

PixelRNN

Variational Autoencoder (VAE)

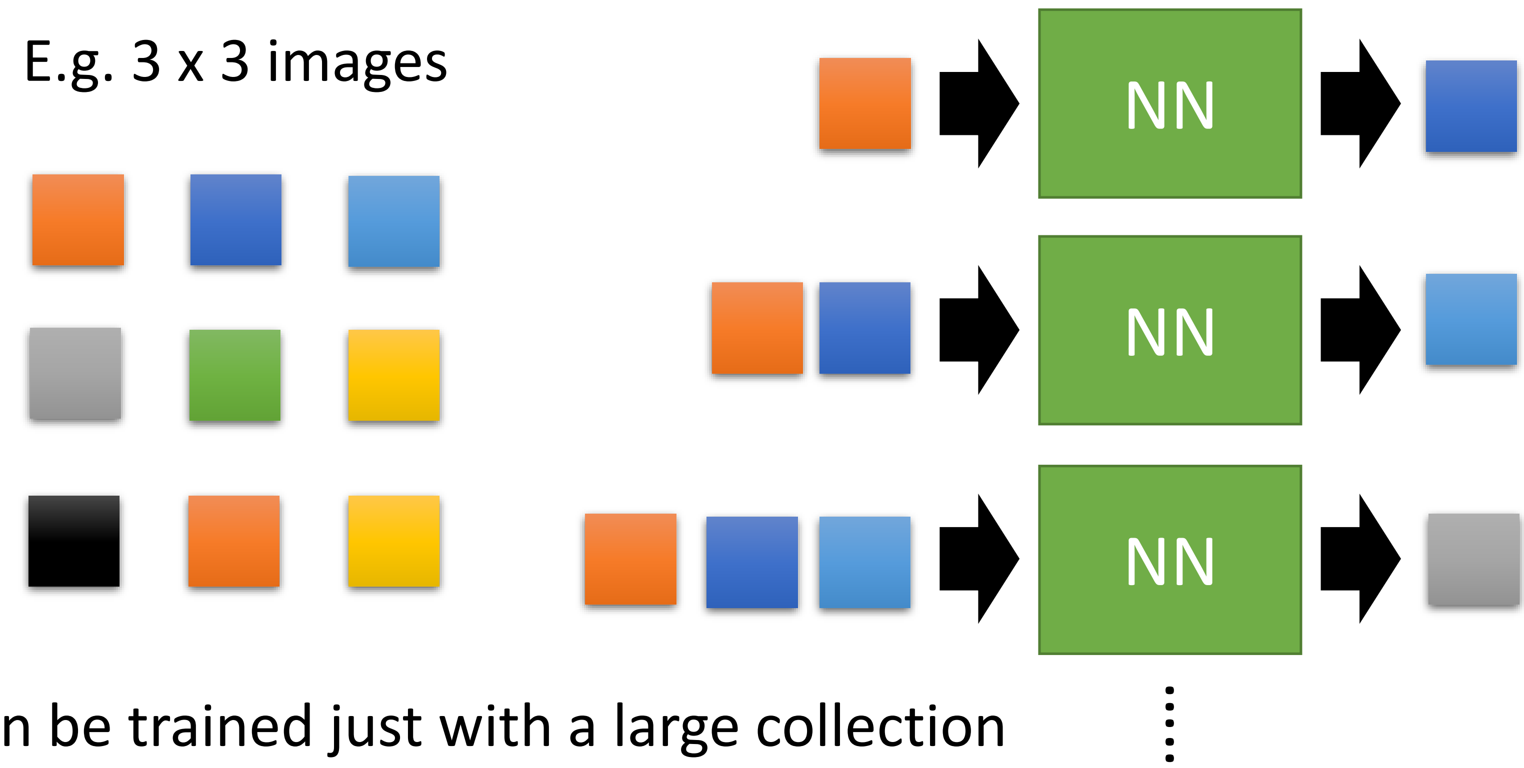
Generative Adversarial Network (GAN)

PixelRNN

Ref: Aaron van den Oord, Nal Kalchbrenner, Koray Kavukcuoglu, Pixel Recurrent Neural Networks, arXiv preprint, 2016

- To create an image, generating a pixel each time

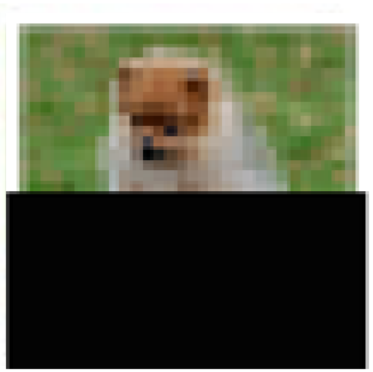
E.g. 3 x 3 images



Can be trained just with a large collection of images without any annotation

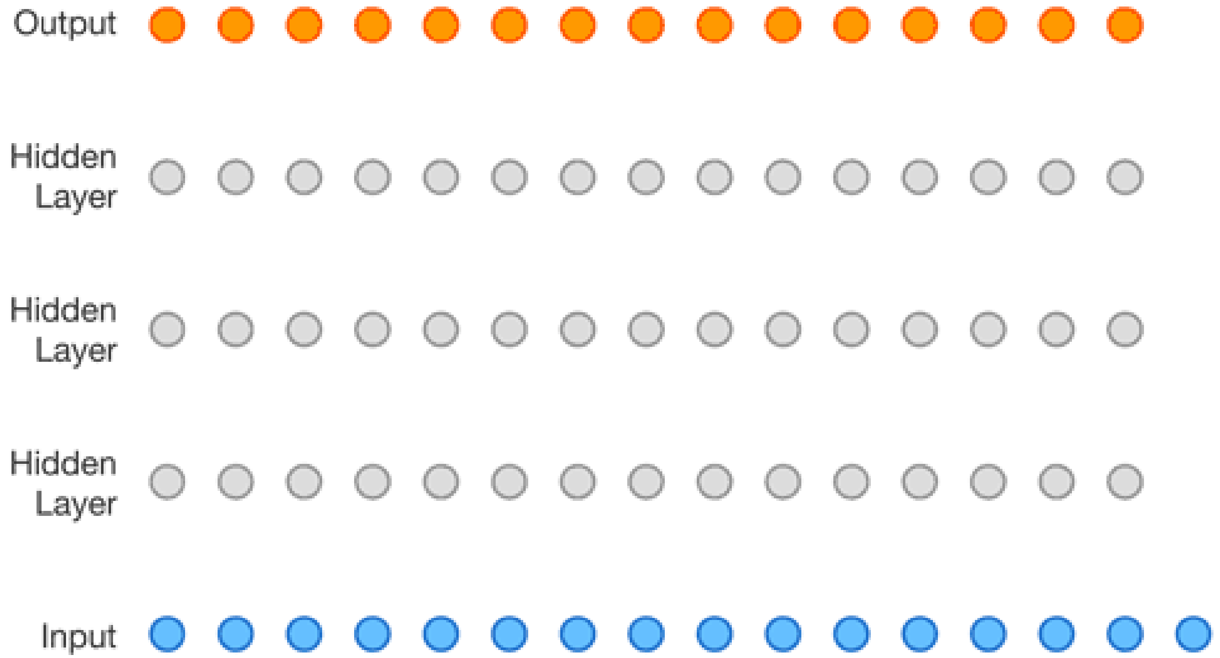
PixelRNN

Ref: Aaron van den Oord, Nal Kalchbrenner, Koray Kavukcuoglu, Pixel Recurrent Neural Networks, arXiv preprint, 2016



Real World

More than images ……



Audio: Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu, WaveNet: A Generative Model for Raw Audio, arXiv preprint, 2016

Video: Nal Kalchbrenner, Aaron van den Oord, Karen Simonyan, Ivo Danihelka, Oriol Vinyals, Alex Graves, Koray Kavukcuoglu, Video Pixel Networks , arXiv preprint, 2016

Practicing Generation Models: Pokémon Creation

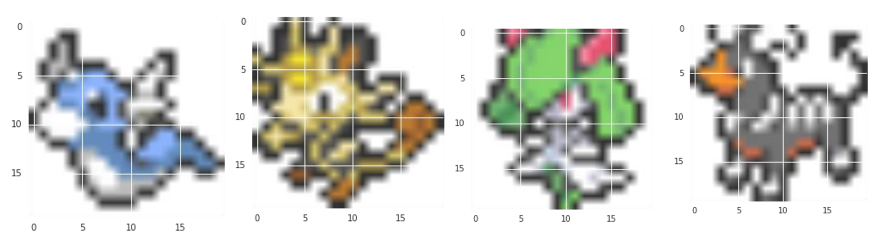
- Small images of 792 Pokémon's
  - Can machine learn to create new Pokémons?

<u>Don't catch them! Create them!</u>

- Source of image: http://bulbapedia.bulbagarden.net/wiki/List_of_Pok%C3%A9mon_by_base_stats_(Generation_VI)

Original image is 40 x 40

Making them into 20 x 20

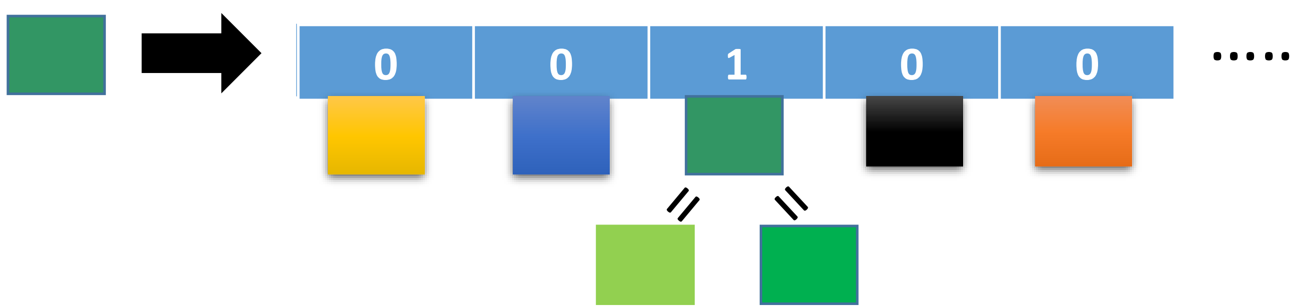


Practicing Generation Models: Pokémon Creation

- 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): http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2016/Pokemon_creation/image.rar
- Pixels (20 x 20): 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
    - http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2016/Pokemon_creation/colormap.txt

```
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 19 41 34 0 0 19 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 44 74 44 51 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 21 80 80 81 0 0 0 0 0 0 0 0
0 0 0 0 0 1 2 3 18 35 22 0 5 2 0 0 0 0 0 0
93 94 93 93 85 95 38 96 97 98 99 99 67 99 9
0 0 0 0 0 0 1 106 106 106 106 106 61 107 0
```

……

0

1

2

```
255 255 255
53 53 53
49 49 49
186 186 186
51 51 51
54 54 54
187 187 187
83 83 83
50 51 52
251 251 251
52 52 52
```

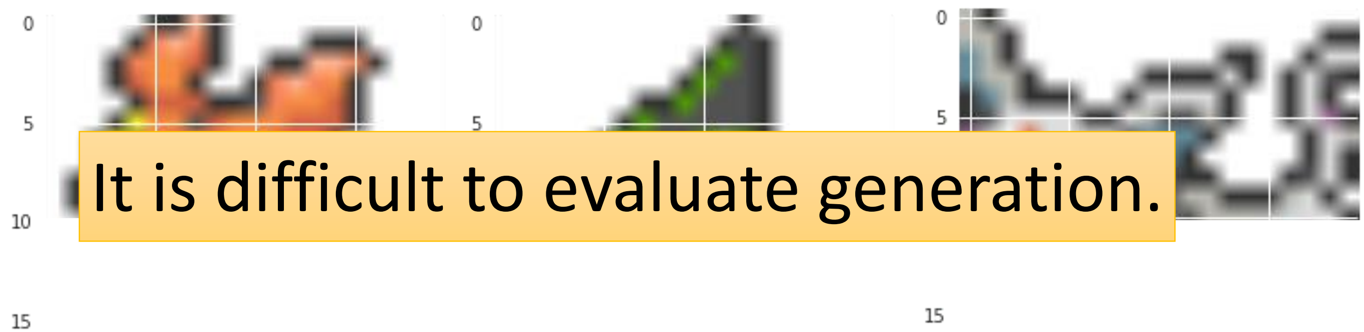
⋮

- Following experiment: 1-layer LSTM, 512 cells

Real Pokémon

Never seen by machine!

Cover 50%



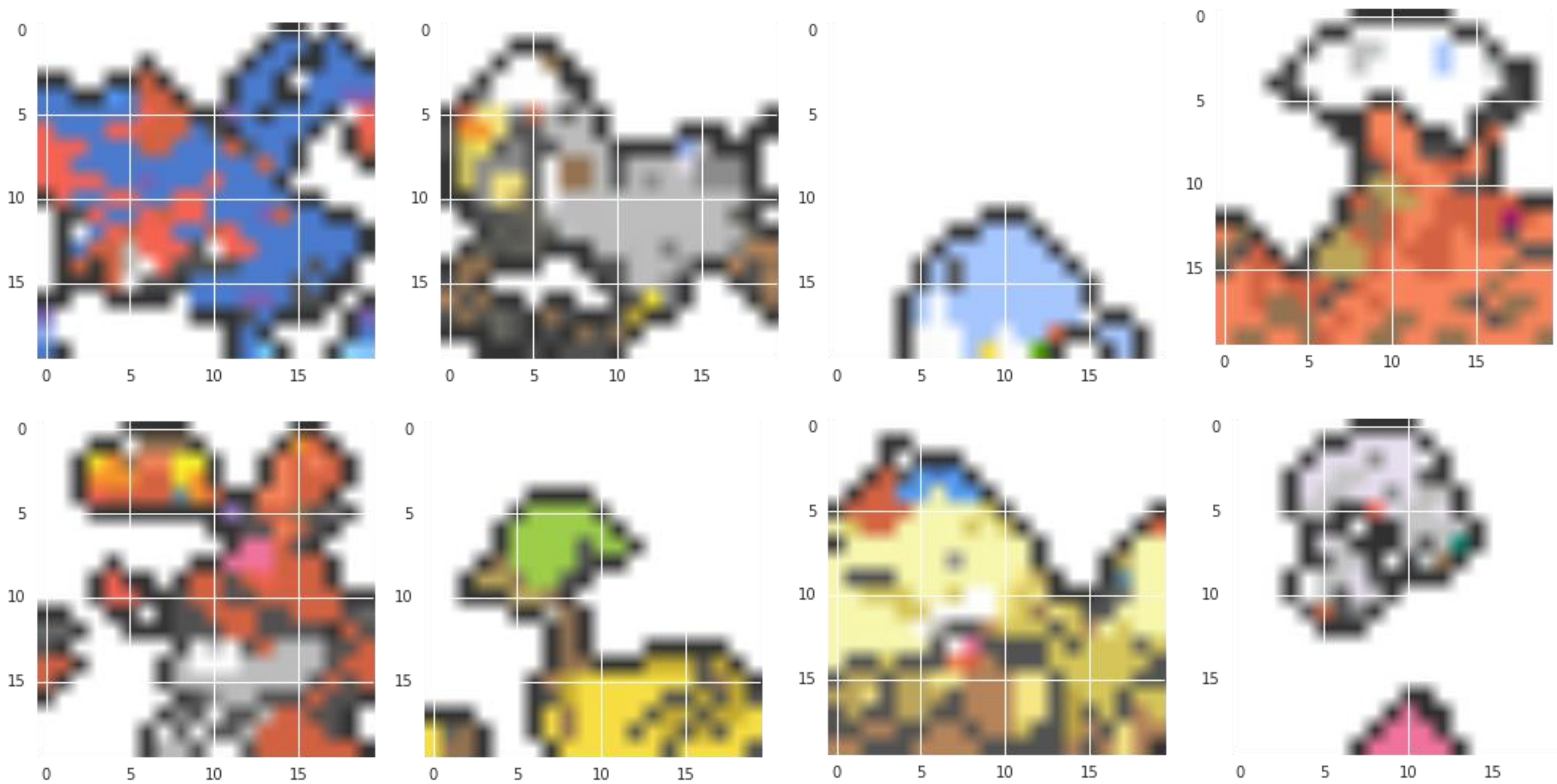
It is difficult to evaluate generation.

Cover 75%

Pokémon Creation

Drawing from scratch

Need some randomness



Generative Models

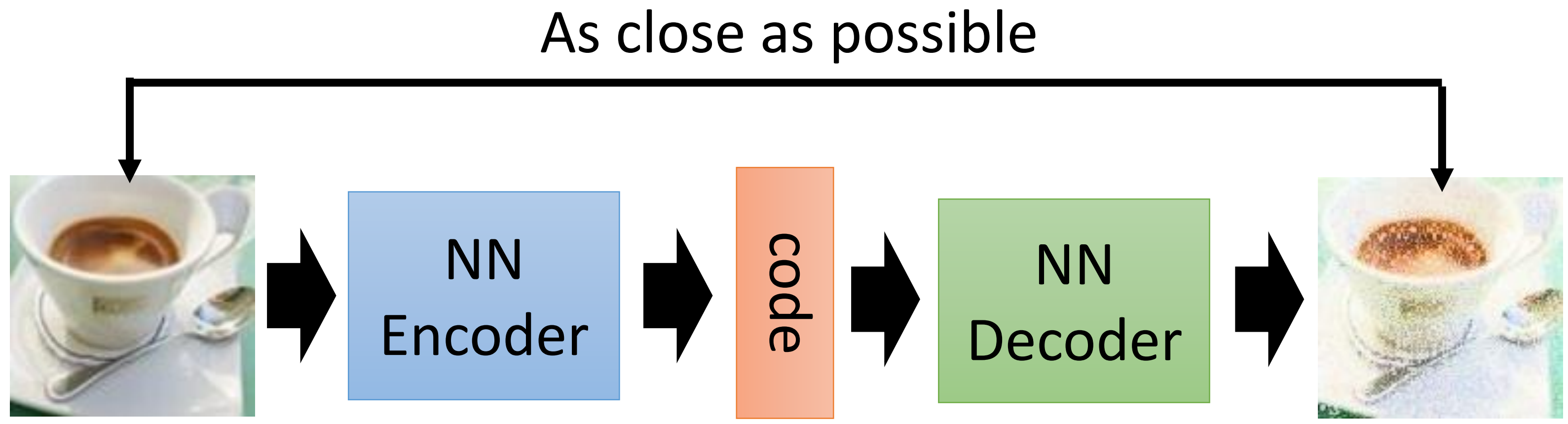
PixelRNN

Variational Autoencoder (VAE)

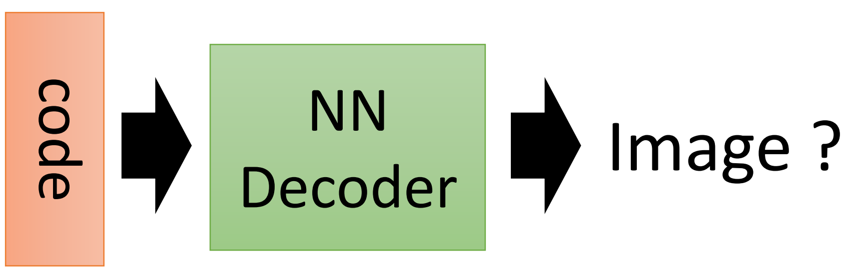
Diederik P Kingma, Max Welling, Auto-Encoding Variational Bayes, arXiv preprint, 2013

Generative Adversarial Network (GAN)

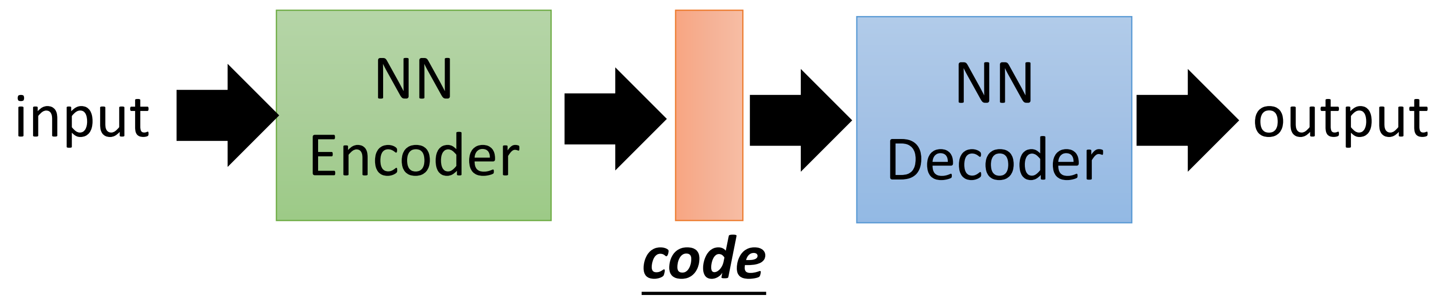
Auto-encoder



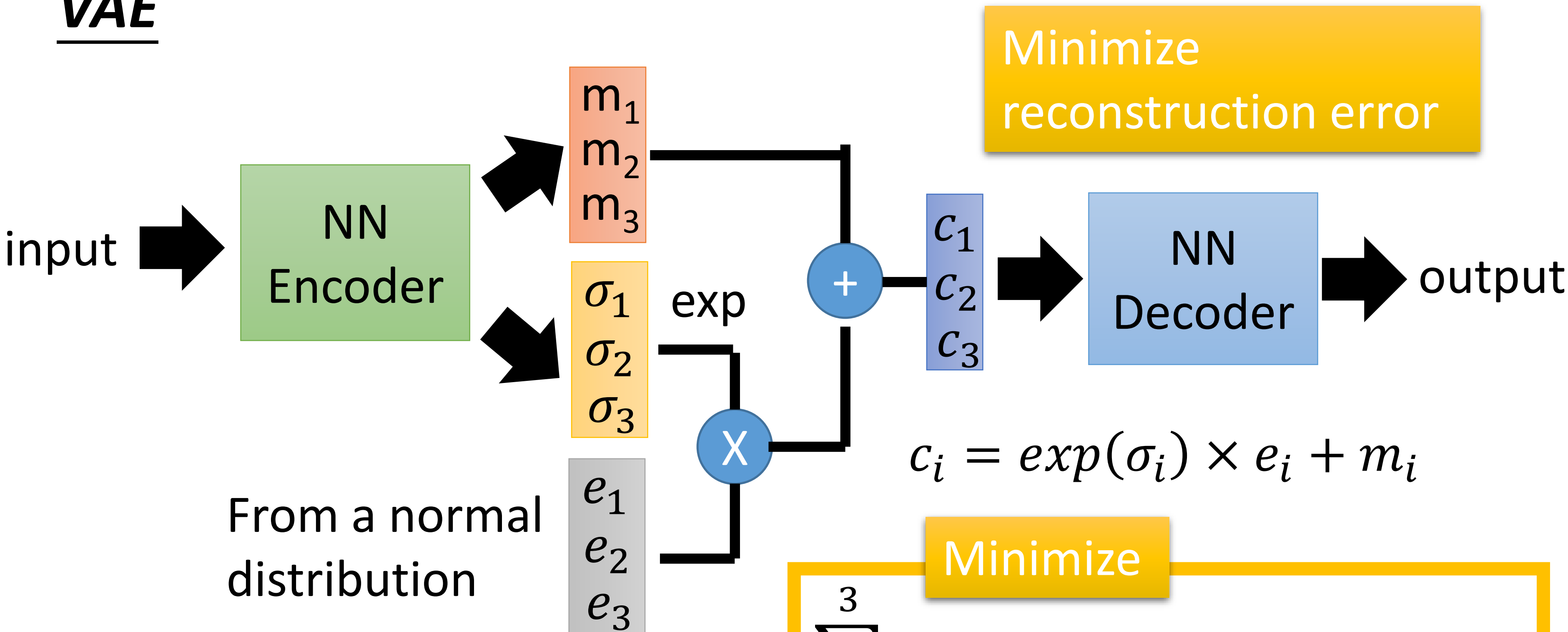
Randomly generate
a vector as code



Auto-encoder



VAE



$$c_i = exp(\sigma_i) \times e_i + m_i$$

Minimize

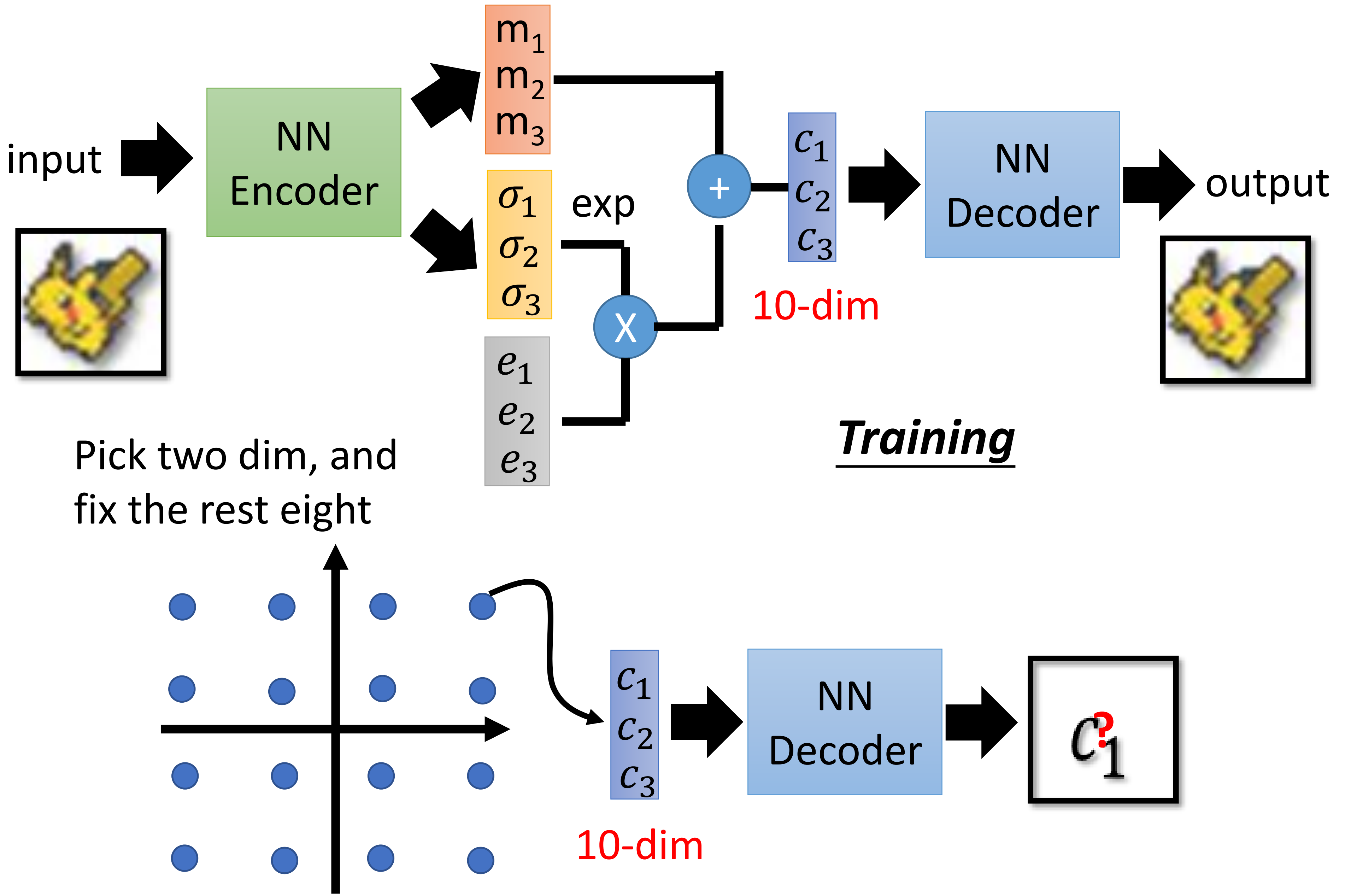
$$\sum_{i=1}^{3} (exp(\sigma_i) - (1 + \sigma_i) + (m_i)^2)$$

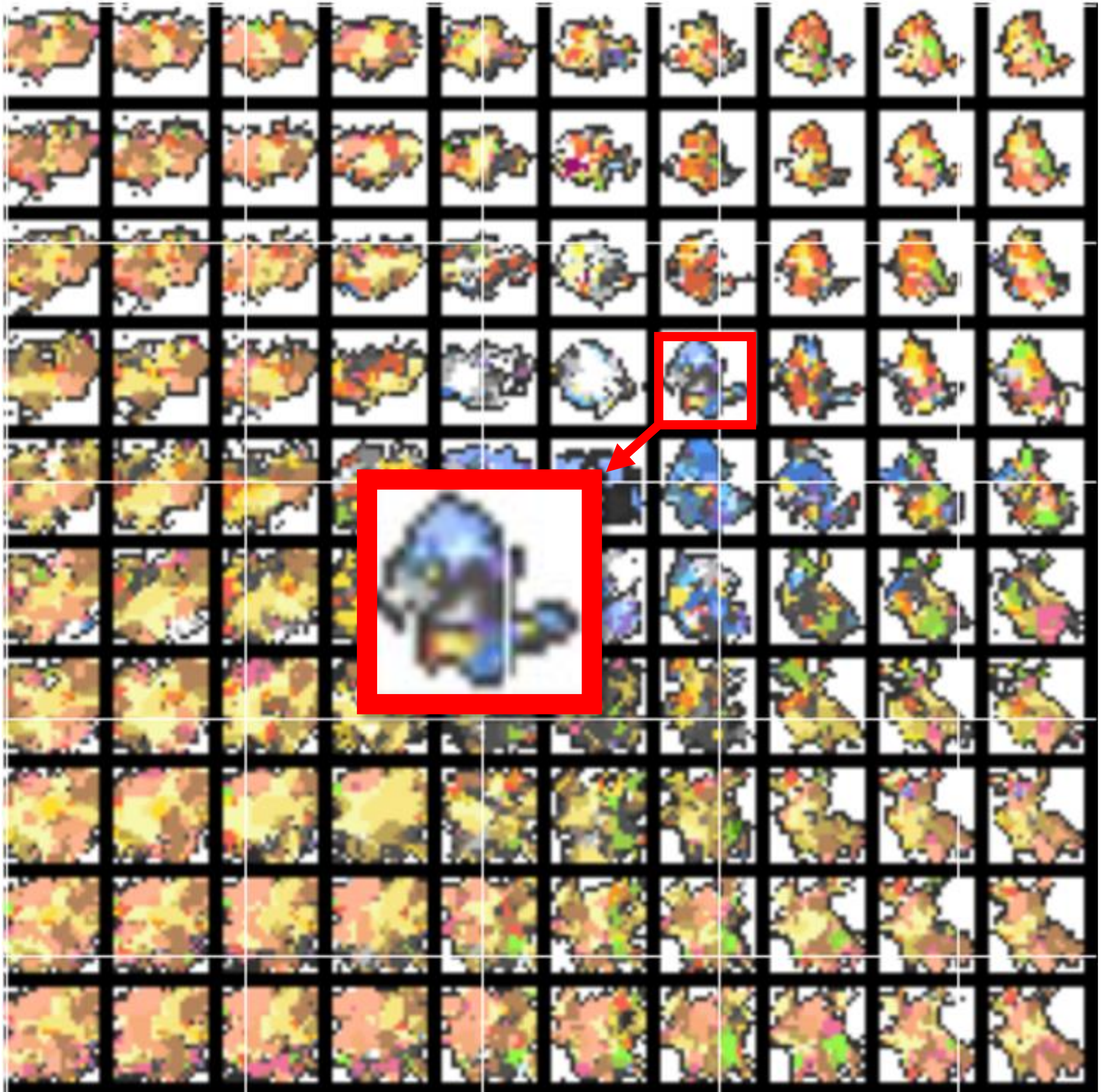
Cifar-10

https://github.com/openai/iaf

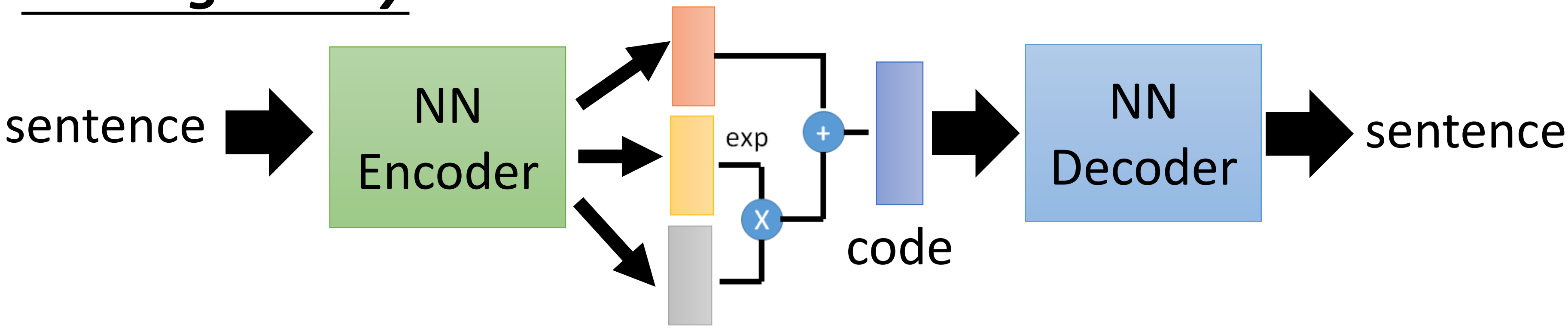
Source of image: https://arxiv.org/pdf/1606.04934v1.pdf

Pokémon Creation

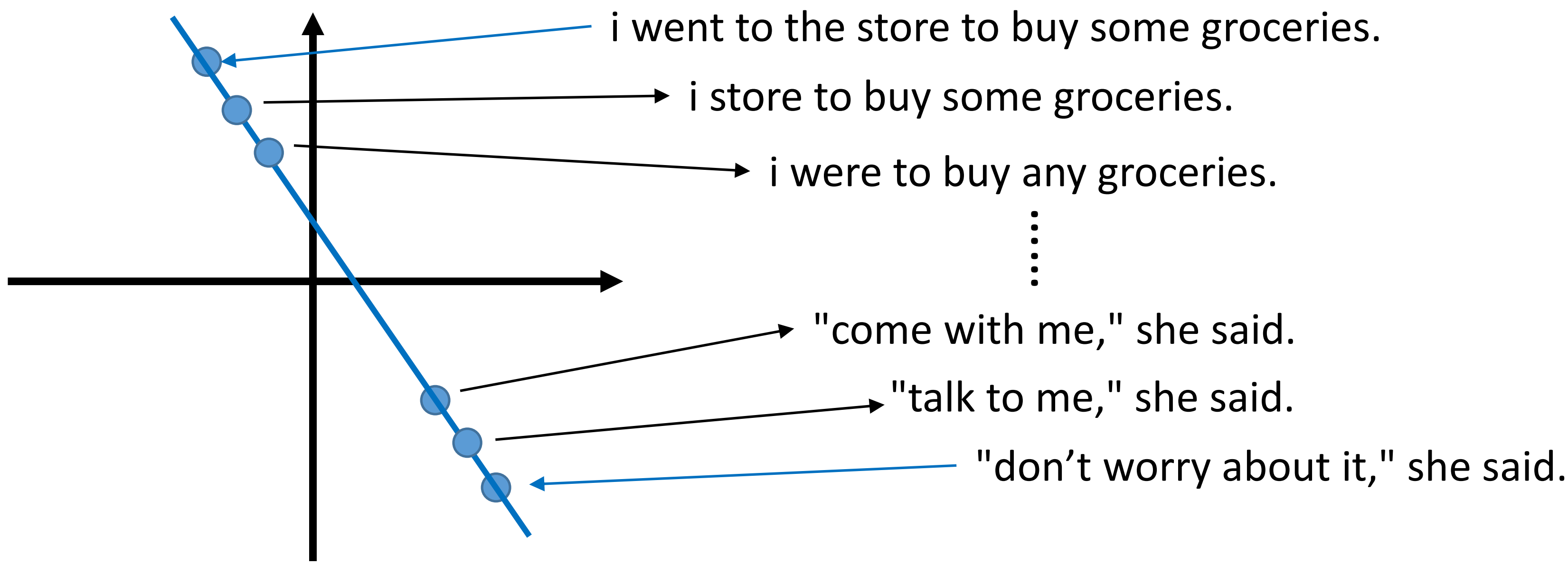




Writing Poetry



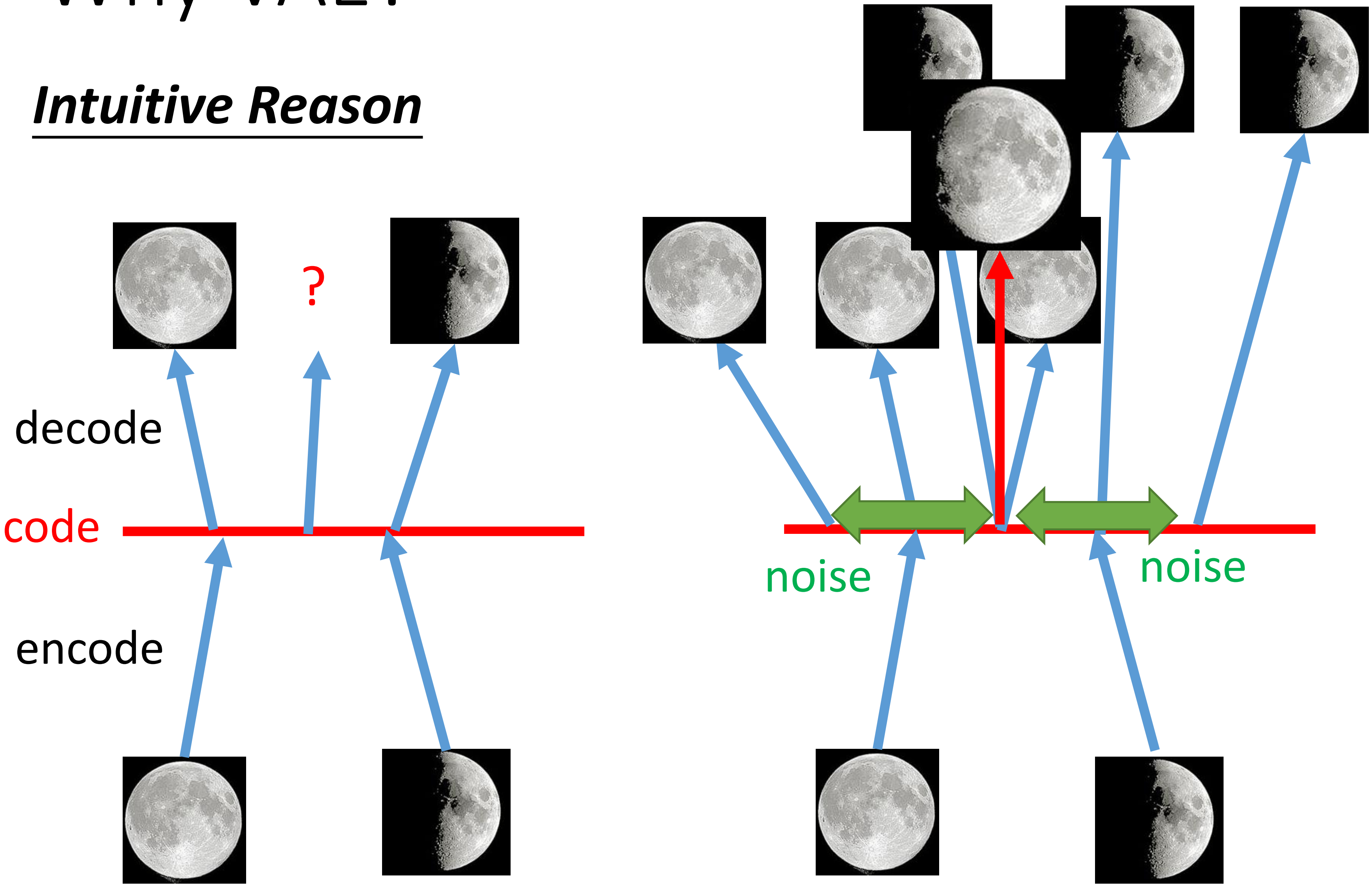
Code Space



Ref: http://www.wired.co.uk/article/google-artificial-intelligence-poetry
Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew M. Dai, Rafal Jozefowicz, Samy Bengio, Generating Sentences from a Continuous Space, arXiv prepring, 2015

Why VAE?

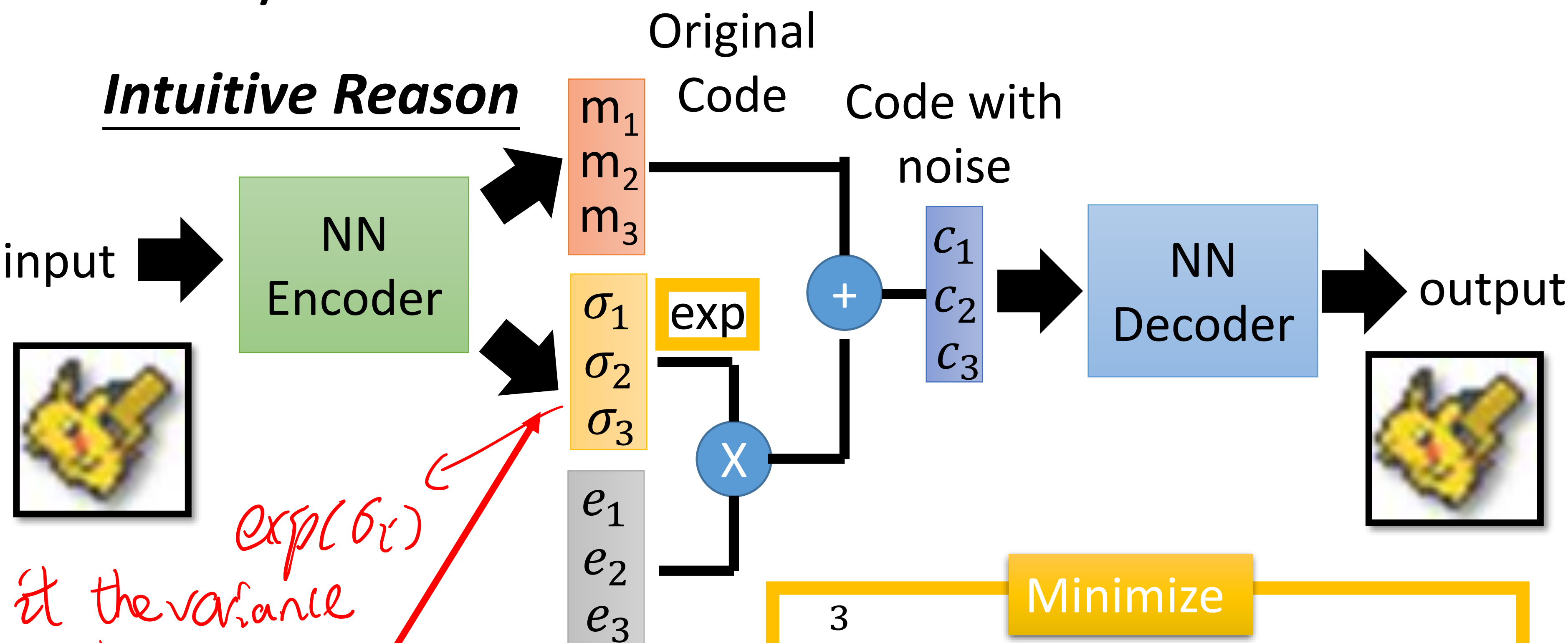
Intuitive Reason



Why VAE?

What will happen if we only minimize reconstruction error?

Intuitive Reason



$\exp(\sigma_i)$

it the variance
not σ_i

The variance of noise is automatically learned

Minimize

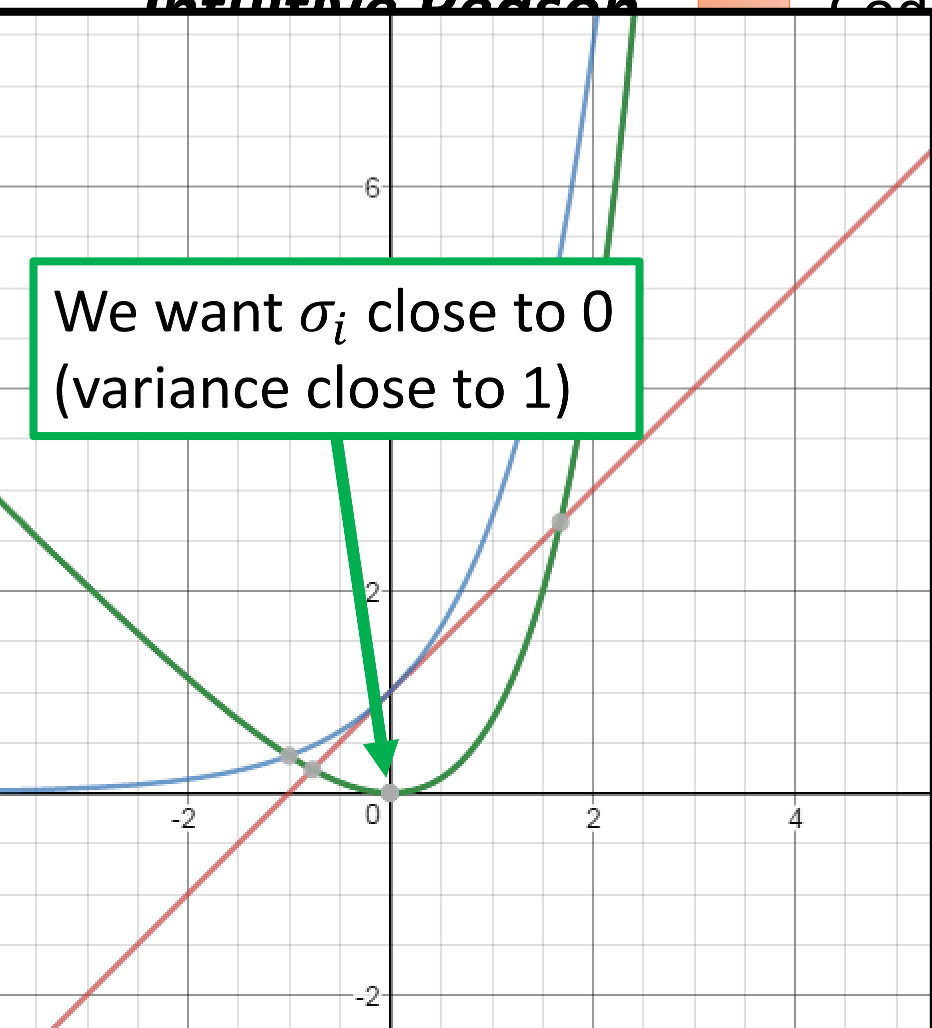
$$\sum_{i=1}^{3}\left(exp(\sigma_i) - (1+\sigma_i) + (m_i)^2\right)$$

l2 Norm

Why VAE?

What will happen if we only minimize reconstruction error?

Intuitive Reason



Original Code

Code with noisy

c_1 c_2 c_3

NN Decoder

output

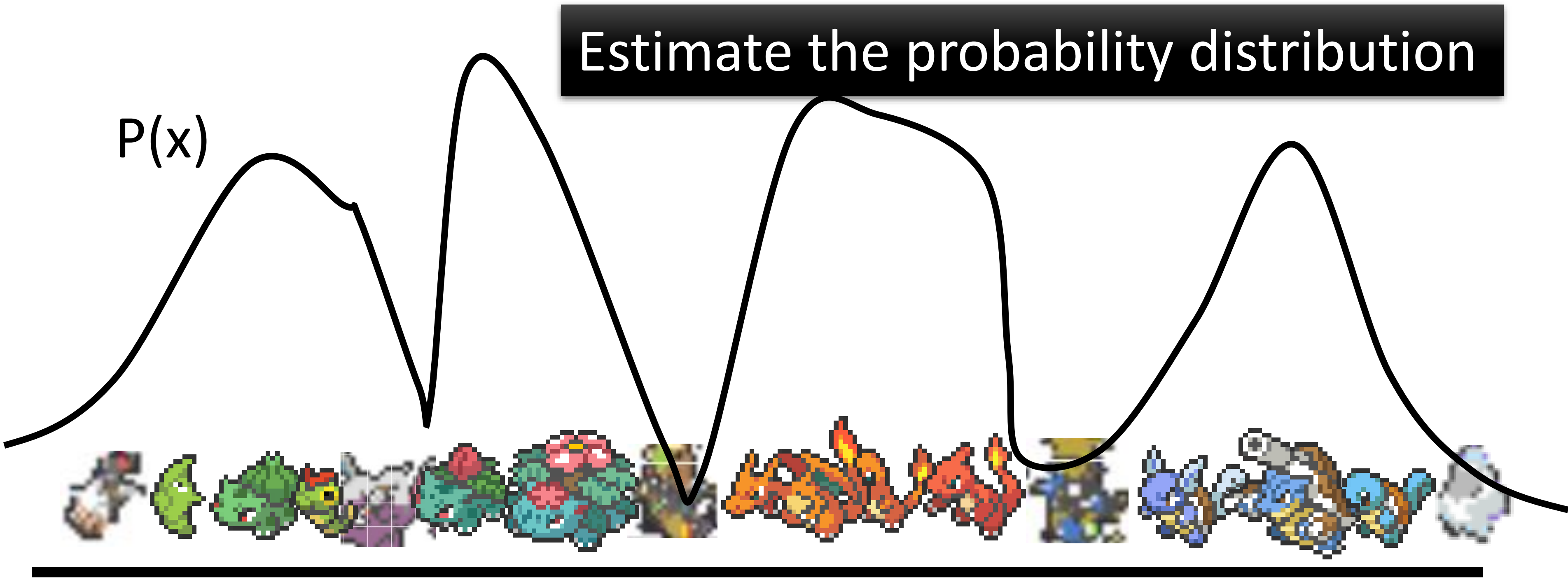
Minimize

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

L2 regularization

Why VAE?

- Back to what we want to do



Each Pokémon is a point x in the space

Gaussian Mixture Model

$$P(x) = \sum_m P(m)P(x|m)$$

How to sample?

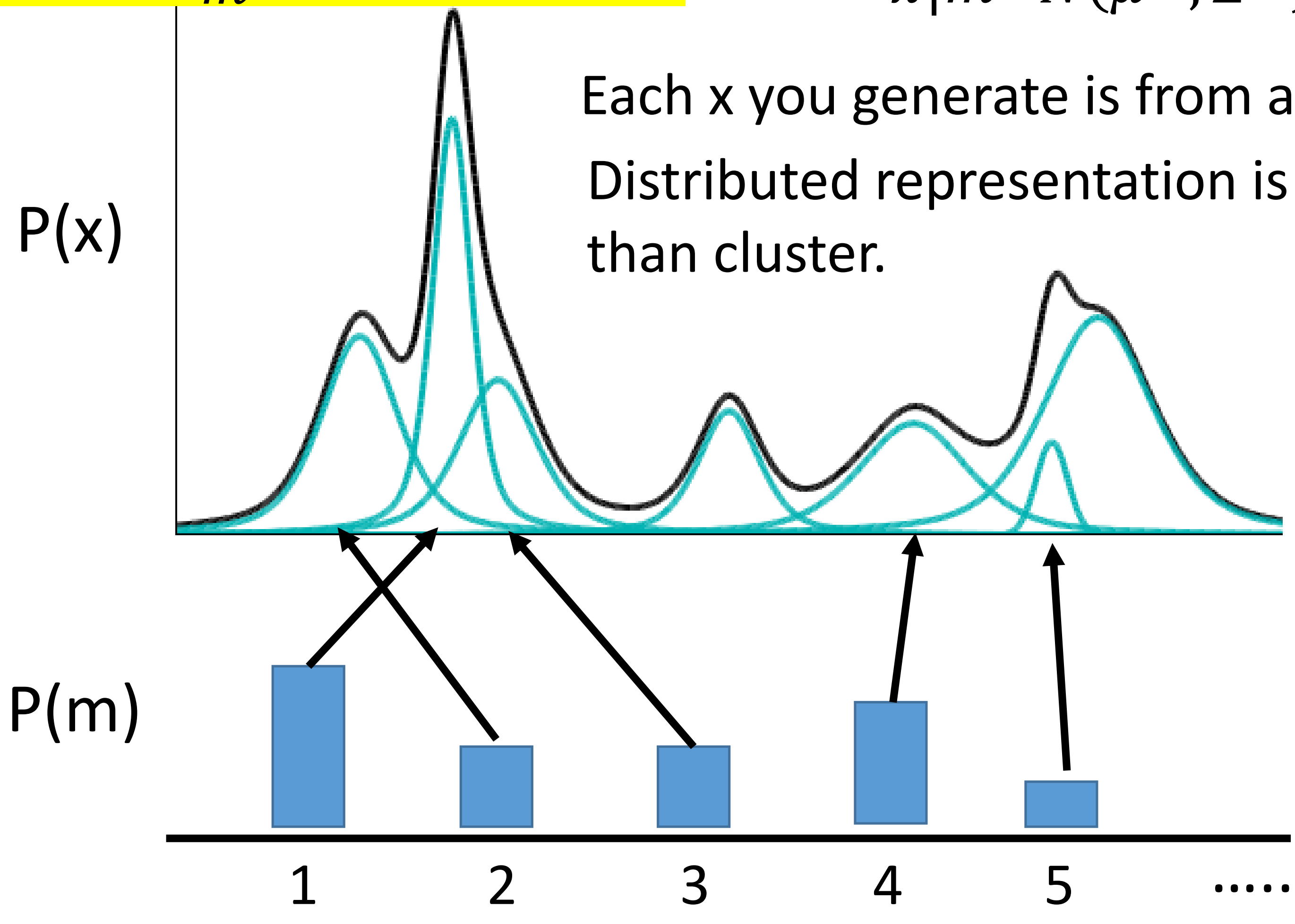
$m \sim P(m)$ (multinomial)

m is an integer

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

Each x you generate is from a mixture

Distributed representation is better than cluster.



VAE

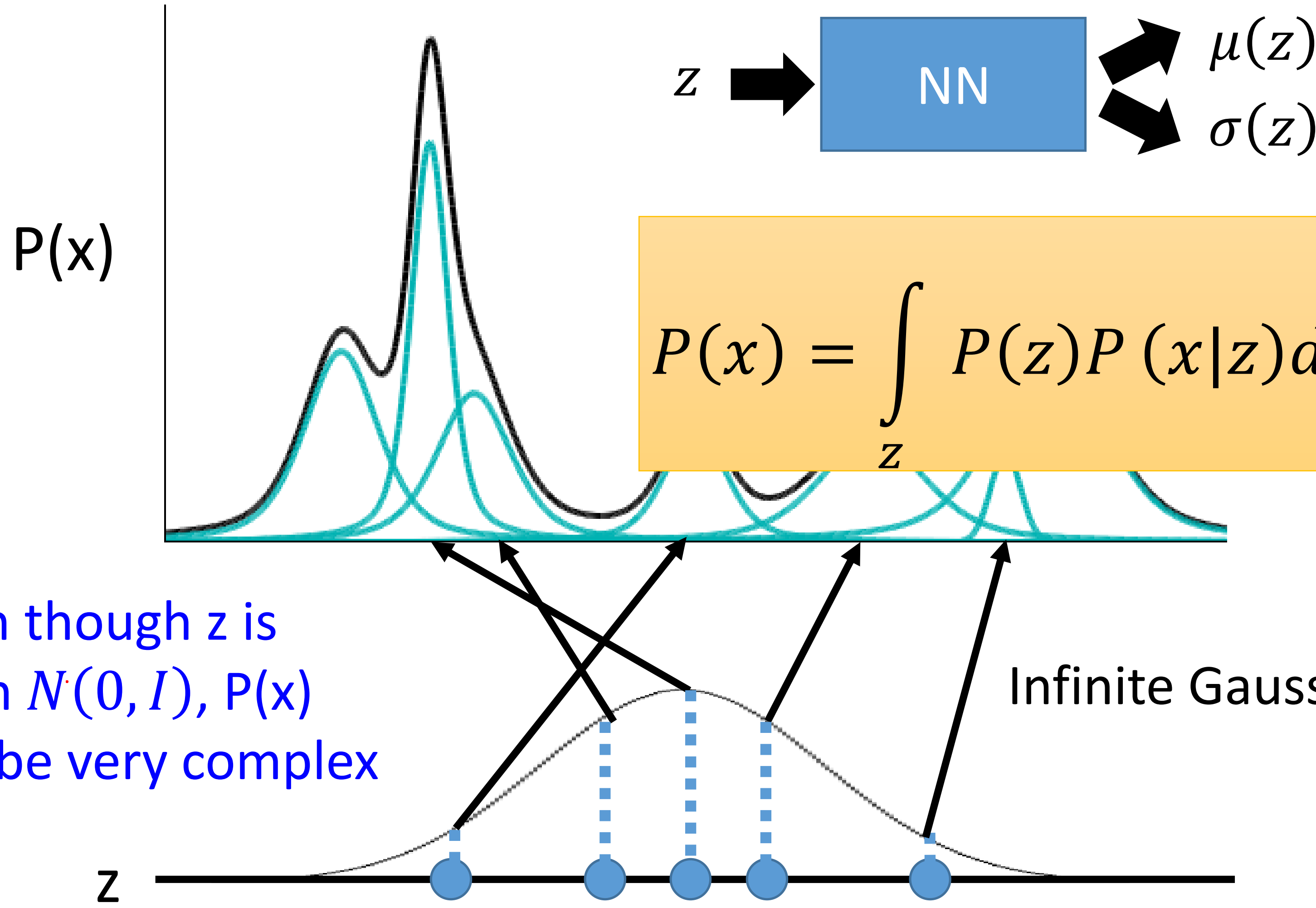
$z \sim N(0, I)$

z is a vector from normal distribution

$x|z \sim N\left(\mu(z), \sigma(z)\right)$

Each dimension of z represents an attribute

$$P(x) = \int_z P(z) P(x|z) dz$$



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

Maximizing Likelihood

$$P(x) = \int_z P(z)P(x|z)dz$$

$$L = \sum_x logP(x)$$

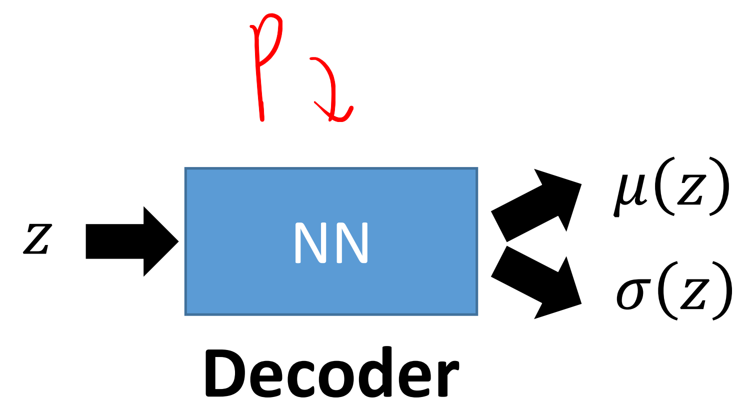
P(z) is normal distribution

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

$\mu(z), \sigma(z)$ is going to be estimated

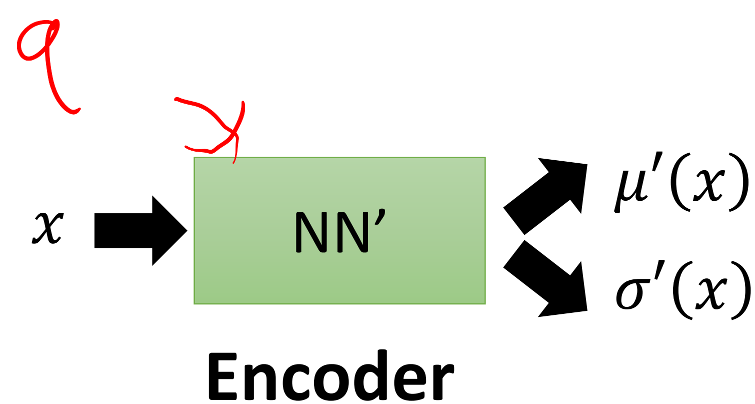
Maximizing the likelihood of the observed x

Tuning the parameters to maximize likelihood L



We need another distribution q(z|x)

$$z|x \sim N\big(\mu'(x), \sigma'(x)\big)$$



Maximizing Likelihood

P(z) is normal distribution

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

$\mu(z), \sigma(z)$ is going to be estimated

$$P(x) = \int_z P(z)P(x|z)dz$$

$$L = \sum_x logP(x)$$

Maximizing the likelihood of the observed x

$\int log P(x) dz = 1$ left = right

$$logP(x) = \int_z q(z|x)logP(x)dz$$

q(z|x) can be any distribution

$$= \int_z q(z|x)log\left(\frac{P(z,x)}{P(z|x)}\right)dz = \int_z q(z|x)log\left(\frac{P(z,x)}{q(z|x)}\frac{q(z|x)}{P(z|x)}\right)dz$$

$$= \int_z q(z|x)log\left(\frac{P(z,x)}{q(z|x)}\right)dz + \underbrace{\int_z q(z|x)log\left(\frac{q(z|x)}{P(z|x)}\right)dz}_{KL(q(z|x)||P(z|x))}$$

≥ 0

$$\geq \int_z q(z|x)log\left(\frac{P(x|z)P(z)}{q(z|x)}\right)dz$$

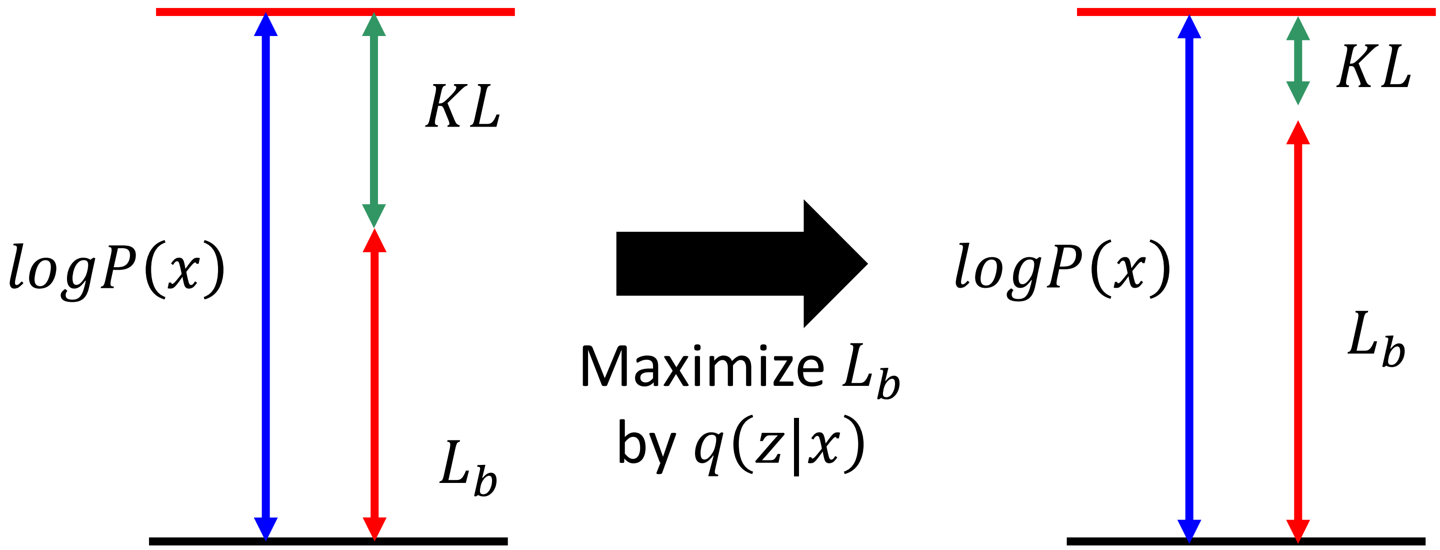
lower bound L_b

Maximizing Likelihood

$$logP(x) = L_b + KL\left(q(z|x)||P(z|x)\right)$$

$$L_b = \int_z q(z|x)log\left(\frac{P(x|z)P(z)}{q(z|x)}\right)dz$$

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

$$P(x) = \int_z P(z)P\,(x|z)dz$$

P(z) is normal distribution

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

$\mu(z), \sigma(z)$ is going to be estimated

$$L = \sum_x logP(x)$$

Maximizing the likelihood of the observed x

$$L_b = \int_z q(z|x)log\left(\frac{P(z,x)}{q(z|x)}\right)dz = \int_z q(z|x)log\left(\frac{P(x|z)P(z)}{q(z|x)}\right)dz$$

$$= \underbrace{\int_z \boxed{q(z|x)}log\left(\frac{P(z)}{\boxed{q(z|x)}}\right)dz}_{-KL\left(q(z|x)||P(z)\right)} + \int_z q(z|x)logP(x|z)dz$$

$$z|x \sim N\big(\mu'(x), \sigma'(x)\big)$$

Connection with Network

Minimize

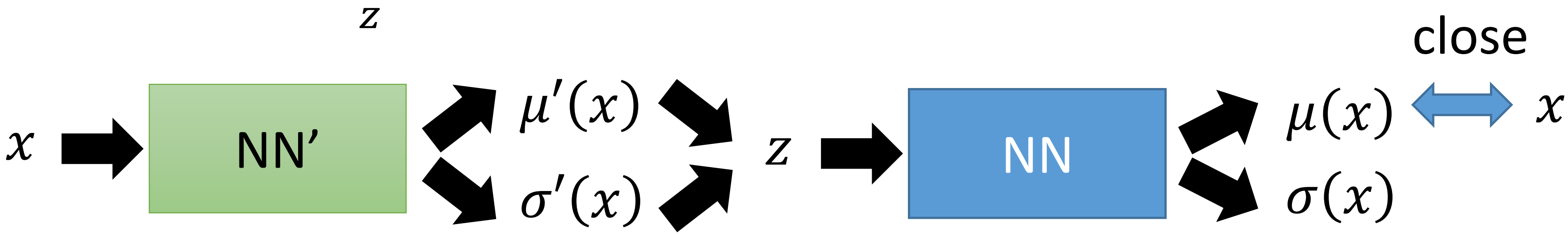
$$\sum_{i=1}^{3}\left(exp(\sigma_i) - (1 + \sigma_i) + (m_i)^2\right)$$

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

(Refer to the Appendix B of the original VAE paper)

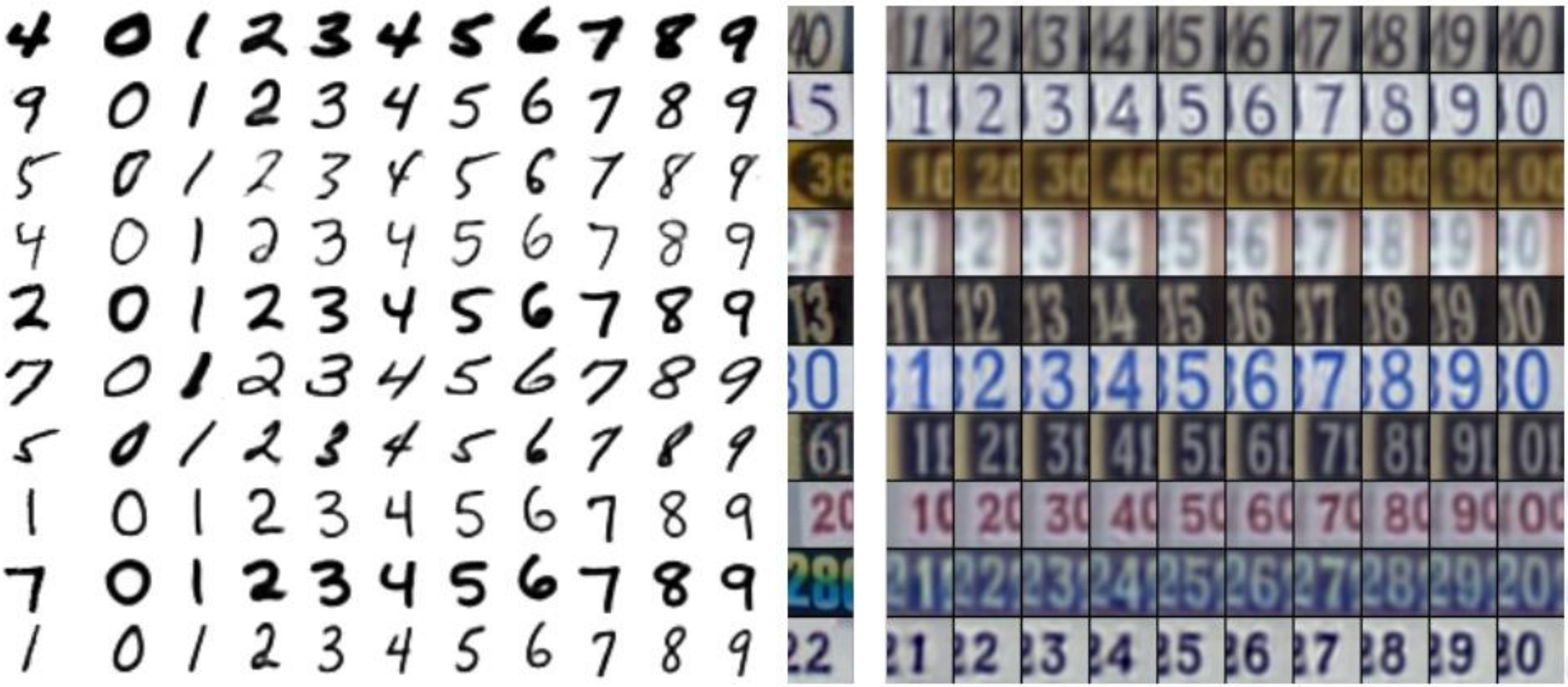
Maximizing

$$\int_z q(z|x)logP(x|z)dz \;\; = E_{q(z|x)}[logP(x|z)]$$



This is the auto-encoder

Conditional VAE



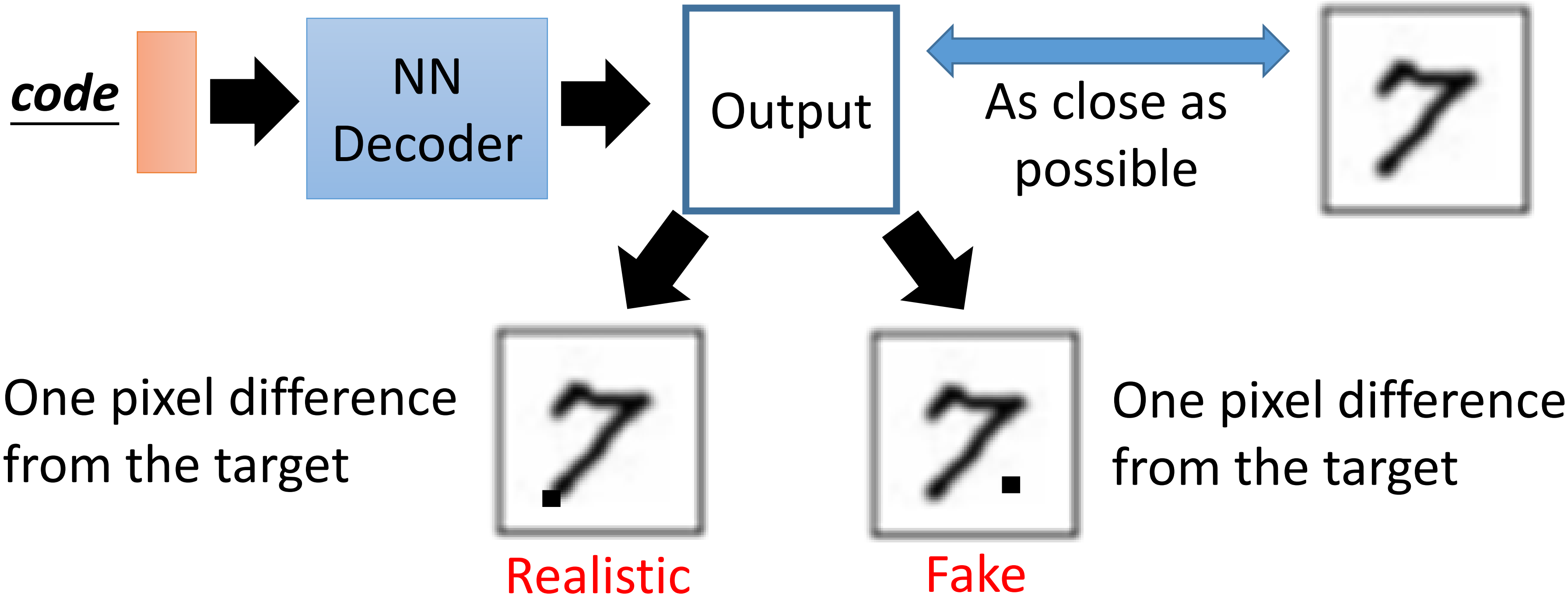
https://arxiv.org/pdf/1406.5298v2.pdf

To learn more …

- Carl Doersch, Tutorial on Variational Autoencoders
- Diederik P. Kingma, Danilo J. Rezende, Shakir Mohamed, Max Welling, "Semi-supervised learning with deep generative models." *NIPS,* 2014.
- Sohn, Kihyuk, Honglak Lee, and Xinchen Yan, "Learning Structured Output Representation using Deep Conditional Generative Models." *NIPS,* 2015.
- Xinchen Yan, Jimei Yang, Kihyuk Sohn, Honglak Lee, "Attribute2Image: Conditional Image Generation from Visual Attributes", ECCV, 2016
- Cool demo:
  - http://vdumoulin.github.io/morphing_faces/
  - http://fvae.ail.tokyo/

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

PixelRNN

Variational Autoencoder (VAE)

Generative Adversarial Network (GAN)

Ian J. Good fellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative Adversarial Networks, arXiv preprint 2014

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).

https://www.quora.com/What-are-some-recent-and-potentially-upcoming-breakthroughs-in-unsupervised-learning

Yann LeCun’s comment

What are some recent and potentially upcoming breakthroughs in deep 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 Nikhil Garg, I lead a team of Quora engineers working on ML/NLP problems

……

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.

https://www.quora.com/What-are-some-recent-and-potentially-upcoming-breakthroughs-in-deep-learning

擬態的演化

http://peellden.pixnet.net/blog/post/40406899-2013-%E7%AC%AC%E5%9B%9B%E5%AD%A3%EF%BC%8C%E5%86%AC%E8%9D%B6%E5%AF%82%E5%AF%A5

棕色

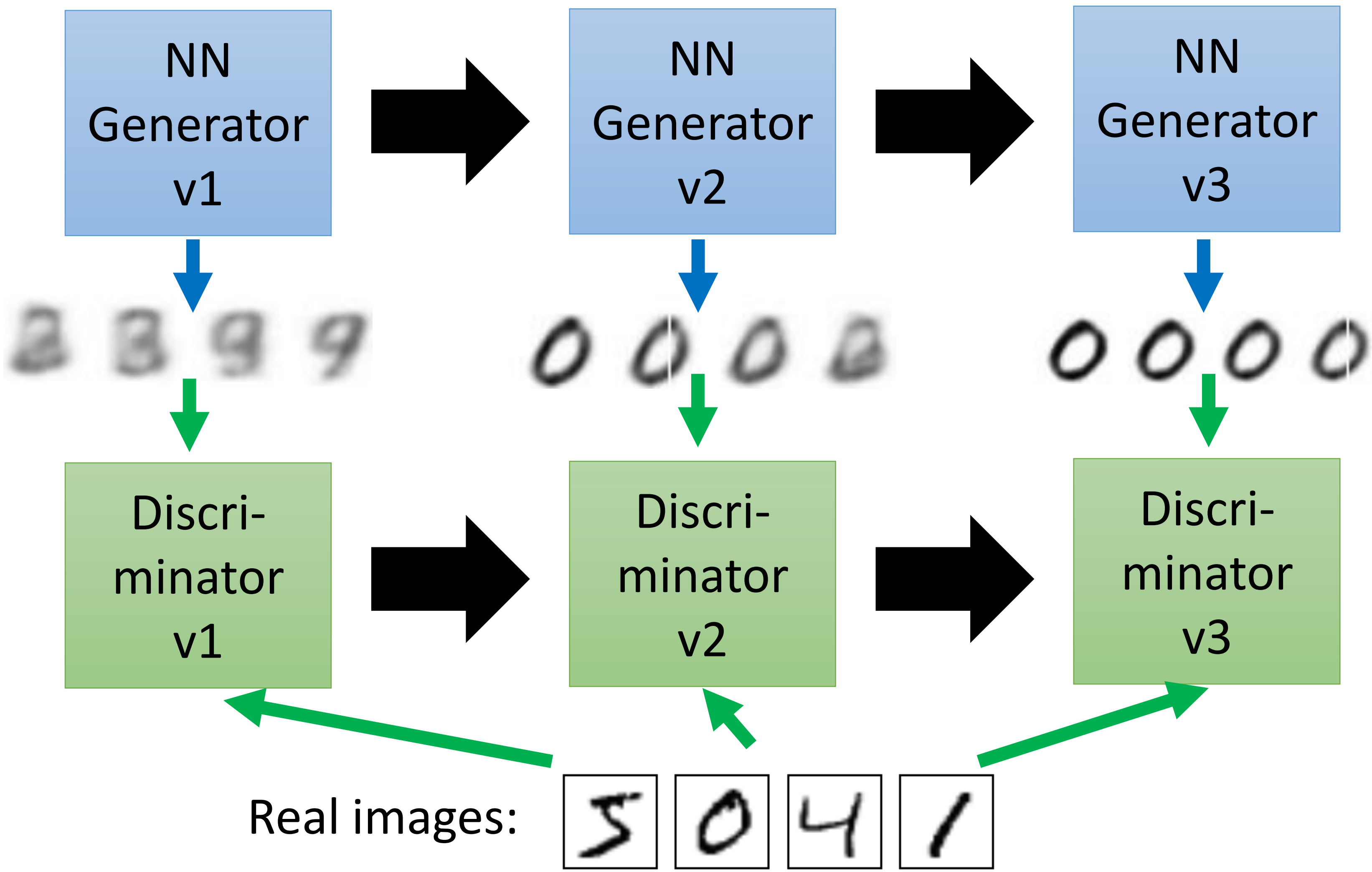
葉脈

蝴蝶不是棕色

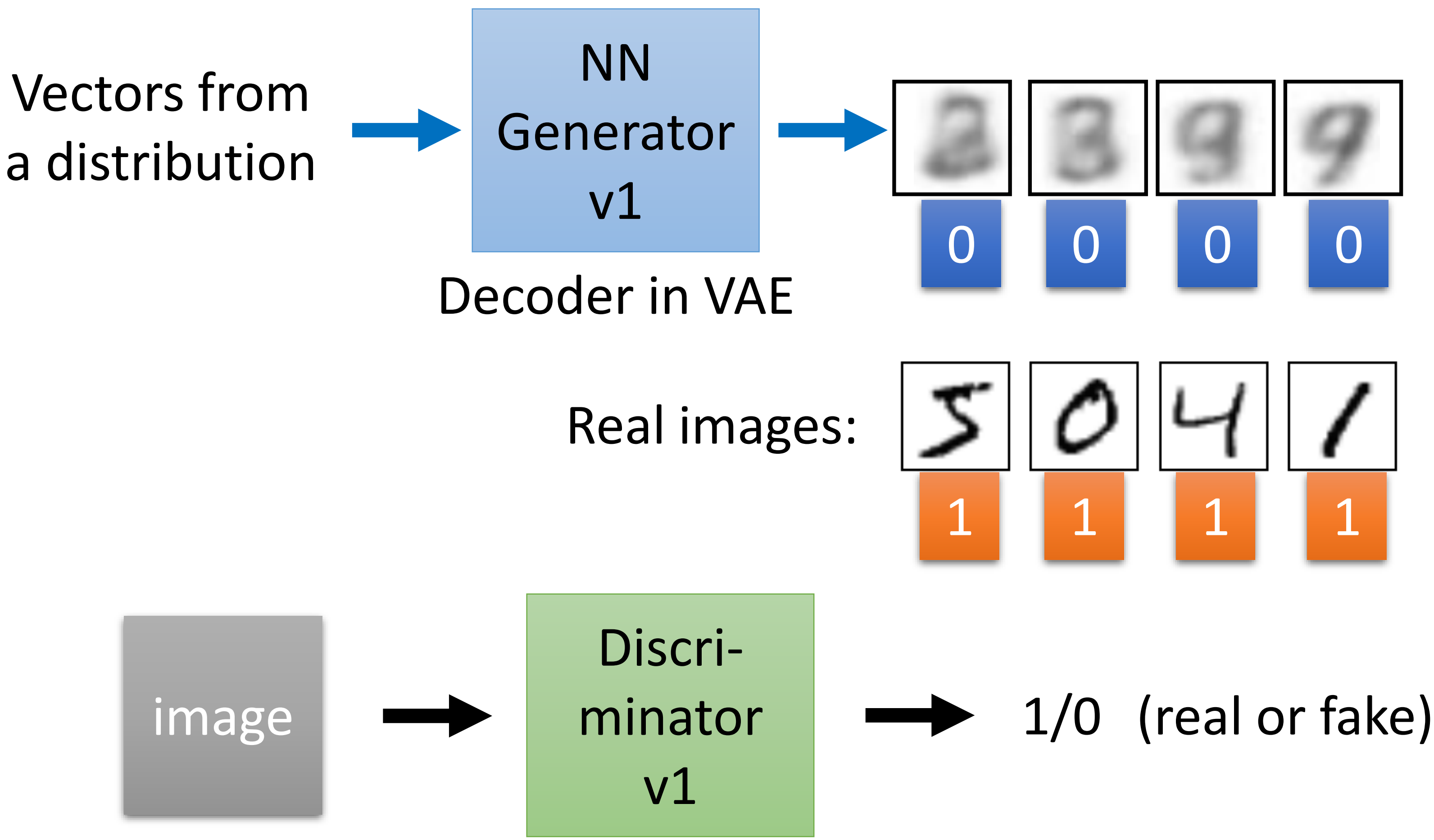
蝴蝶沒有葉脈

………

The evolution of generation



GAN - Discriminator



GAN - Generator

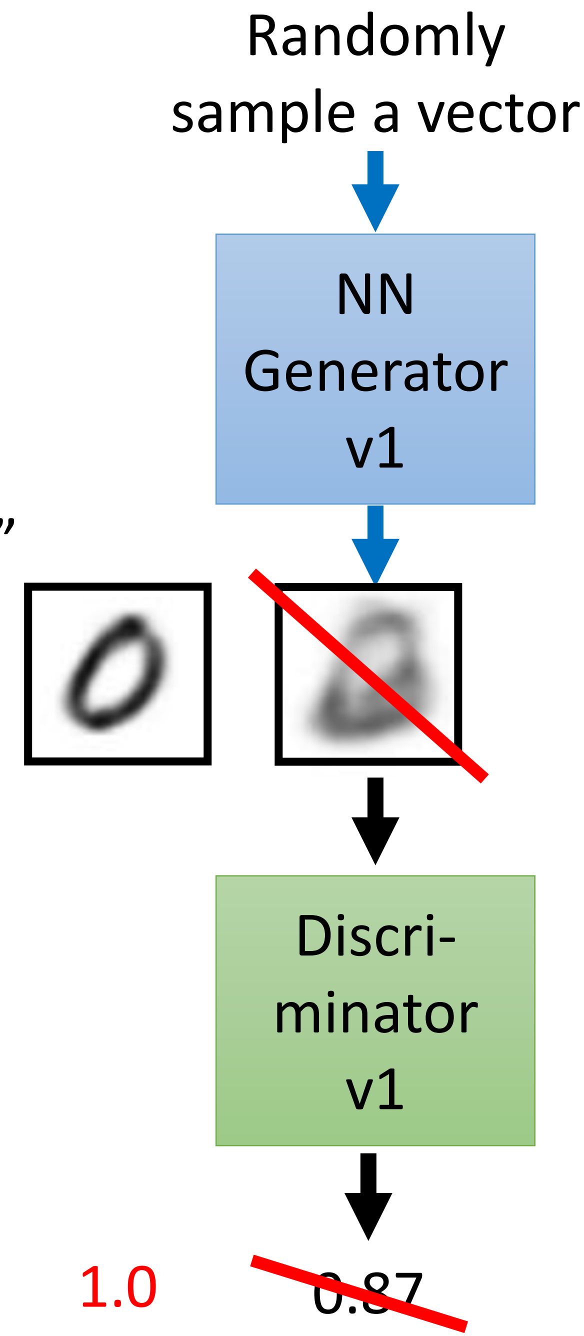
"Tuning" the parameters of generator

➡ The output be classified as "real" (as close to 1 as possible)

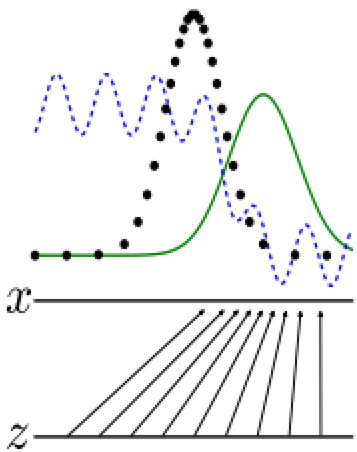
Generator + Discriminator = a network

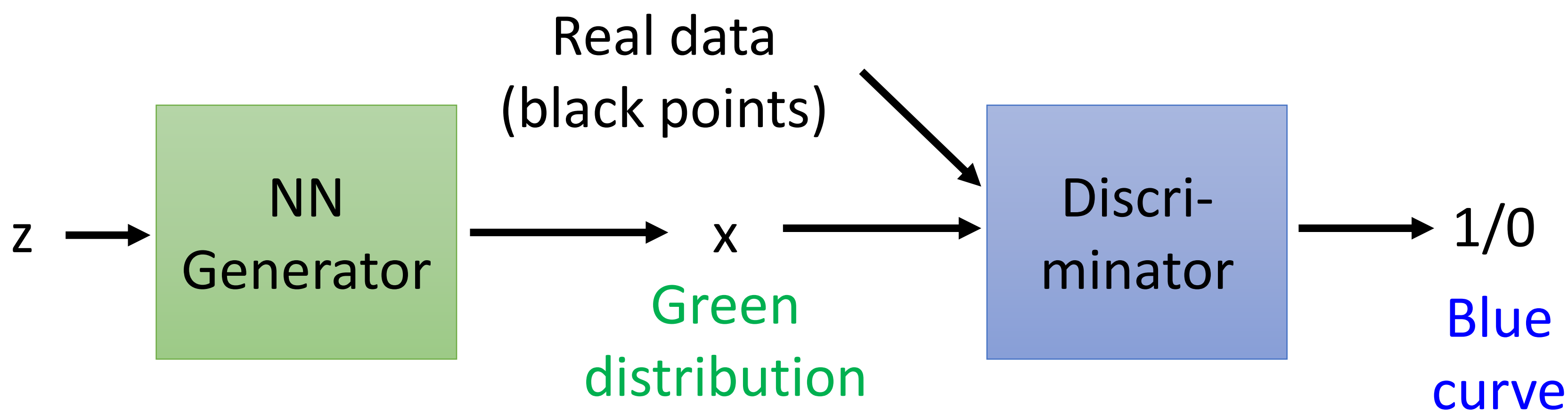
Using gradient descent to find the parameters of generator

Fix the discriminator



GAN – Toy Example

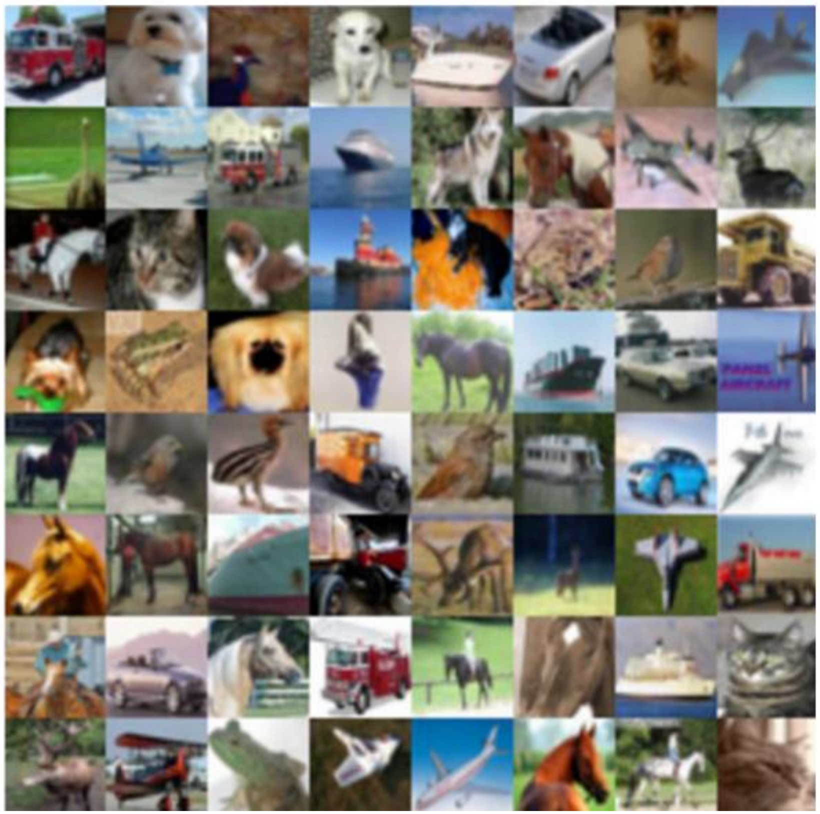


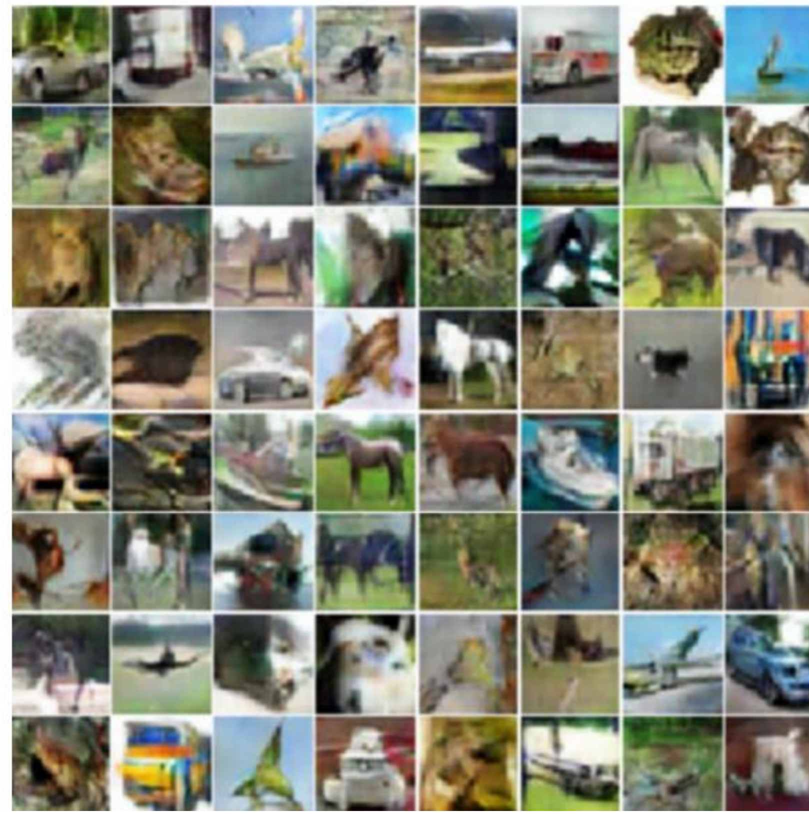


Demo: http://cs.stanford.edu/people/karpathy/gan/

Cifar-10

- Which one is machine-generated?





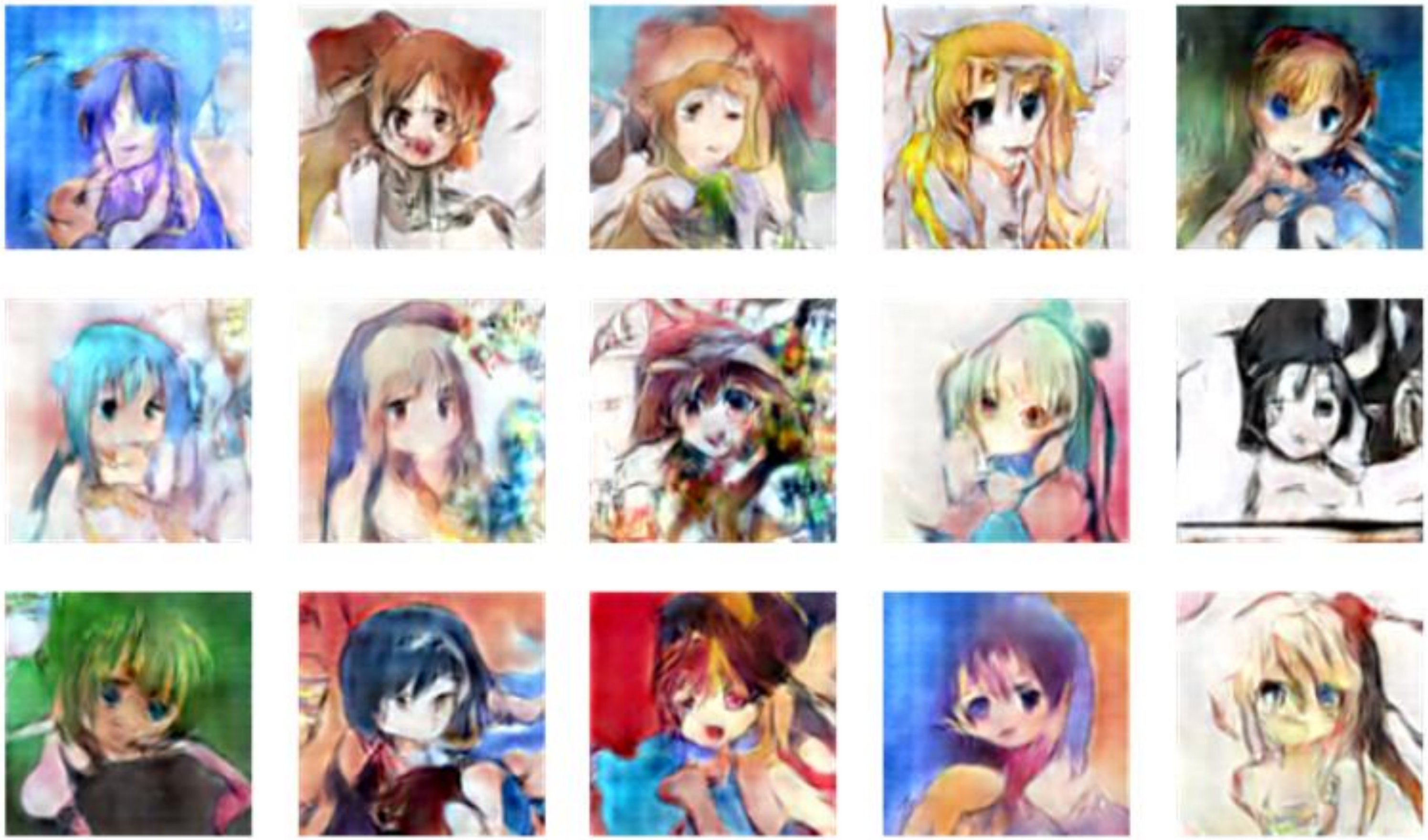
Ref: https://openai.com/blog/generative-models/

Moving on the code space

Alec Radford, Luke Metz, Soumith Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, ICLR, 2016

畫漫畫

- Ref: https://github.com/mattya/chainer-DCGAN

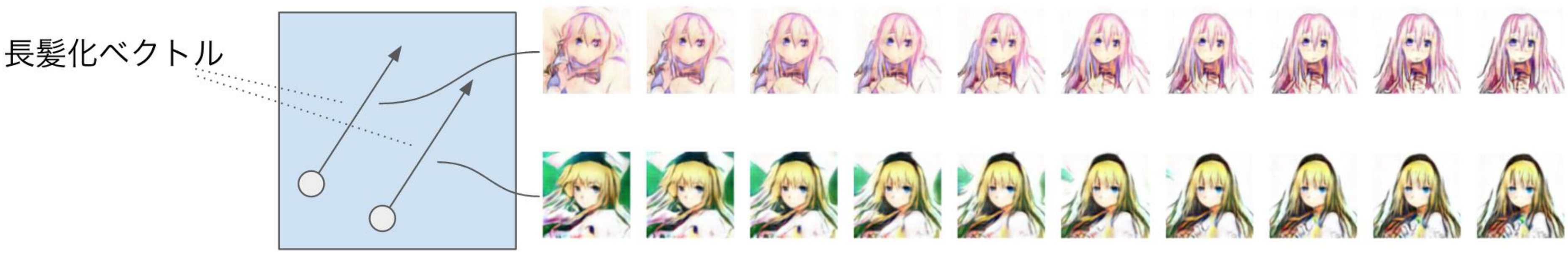


畫漫畫

Web demo: http://mattya.github.io/chainer-DCGAN/

- Ref: http://qiita.com/mattya/items/e5bfe5e04b9d2f0bbd47

元画像　－赤髪＋金髪　－赤目＋青目　＋制服＋セーラー　＋笑顔＋口開き　＋青背景



一番左のキャラクターが元画像で、
右に行くほど長髪化ベクトルを強く足している

In practical ……

- GANs are difficult to optimize.
- No explicit signal about how good the generator is
  - In standard NNs, we monitor loss
  - In GANs, we have to keep “well-matched in a contest”
- When discriminator fails, it does not guarantee that generator generates realistic images
  - Just because discriminator is stupid
  - Sometimes generator find a specific example that can fail the discriminator
  - Making discriminator more robust may be helpful.

To learn more …

- “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks”
- “Improved Techniques for Training GANs”
- “Autoencoding beyond pixels using a learned similarity metric”
- “Deep Generative Image Models using a Laplacian Pyramid of Adversarial Network”
- “Super Resolution using GANs”
- “Generative Adversarial Text to Image Synthesis”

To learn more …

- Basic tutorial:
  - http://blog.aylien.com/introduction-generative-adversarial-networks-code-tensorflow/
  - https://bamos.github.io/2016/08/09/deep-completion/
  - http://blog.evjang.com/2016/06/generative-adversarial-nets-in.html

Acknowledgement

- 感謝 Ryan Sun 來信指出投影片上的錯字