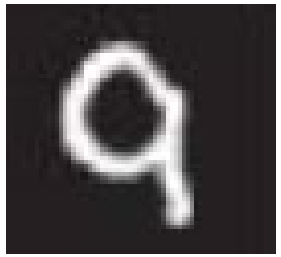
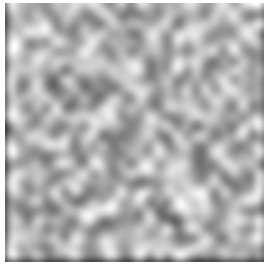
Deep Learning Tutorial

李宏毅

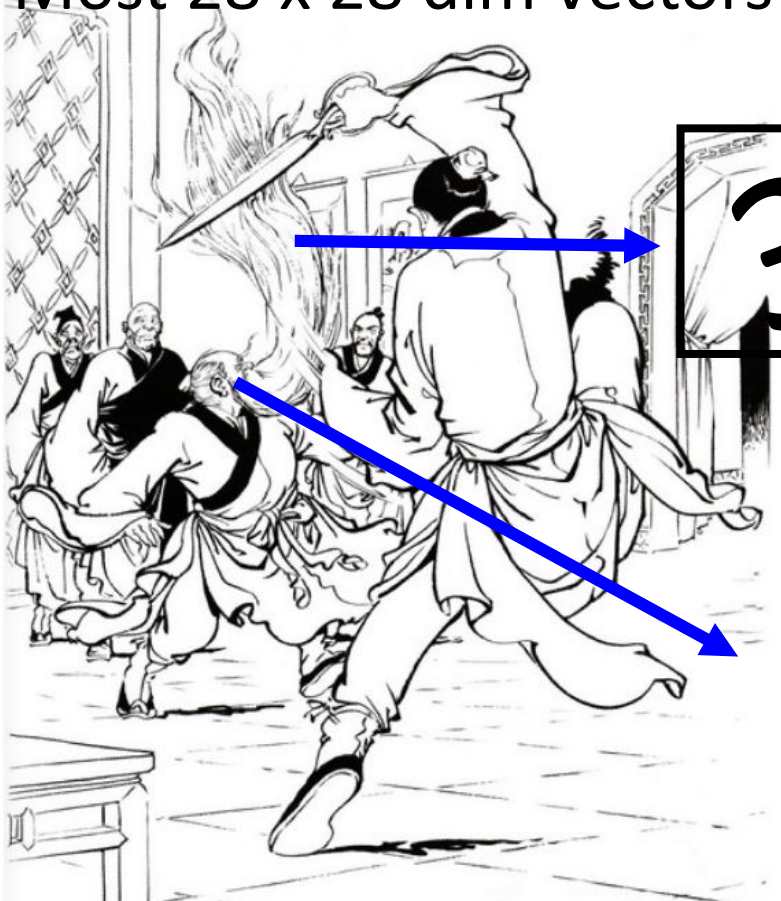
Hung-yi Lee

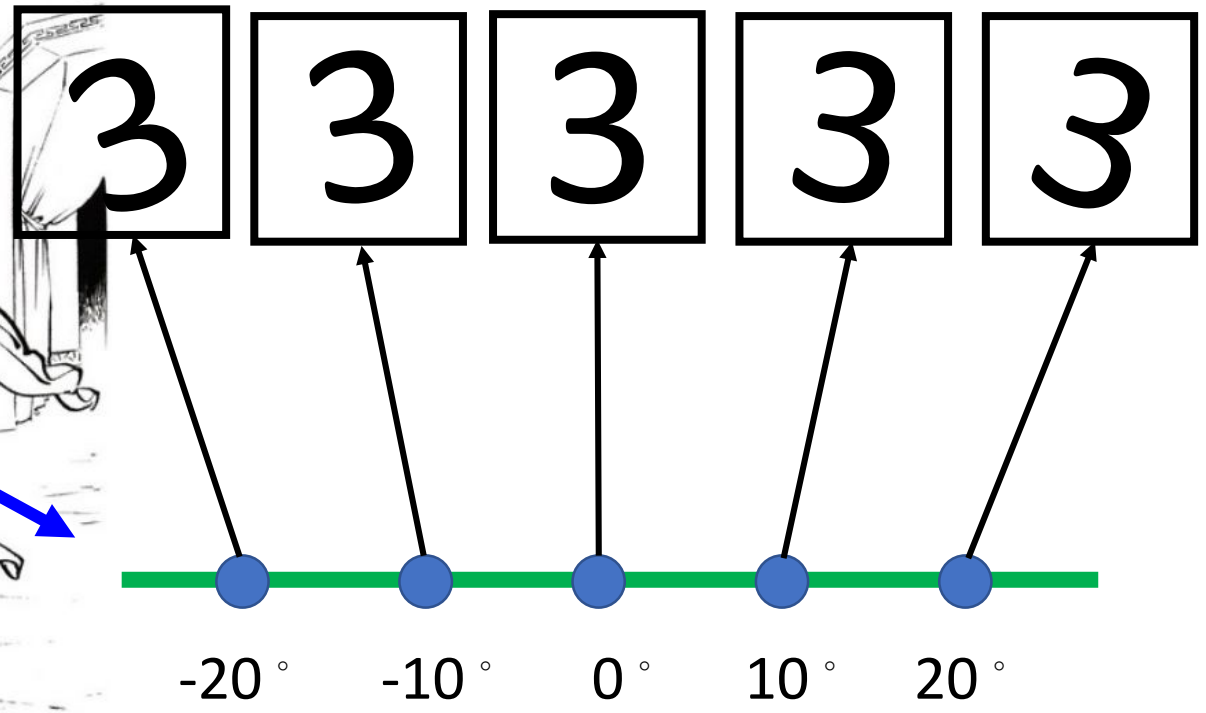
Motivation



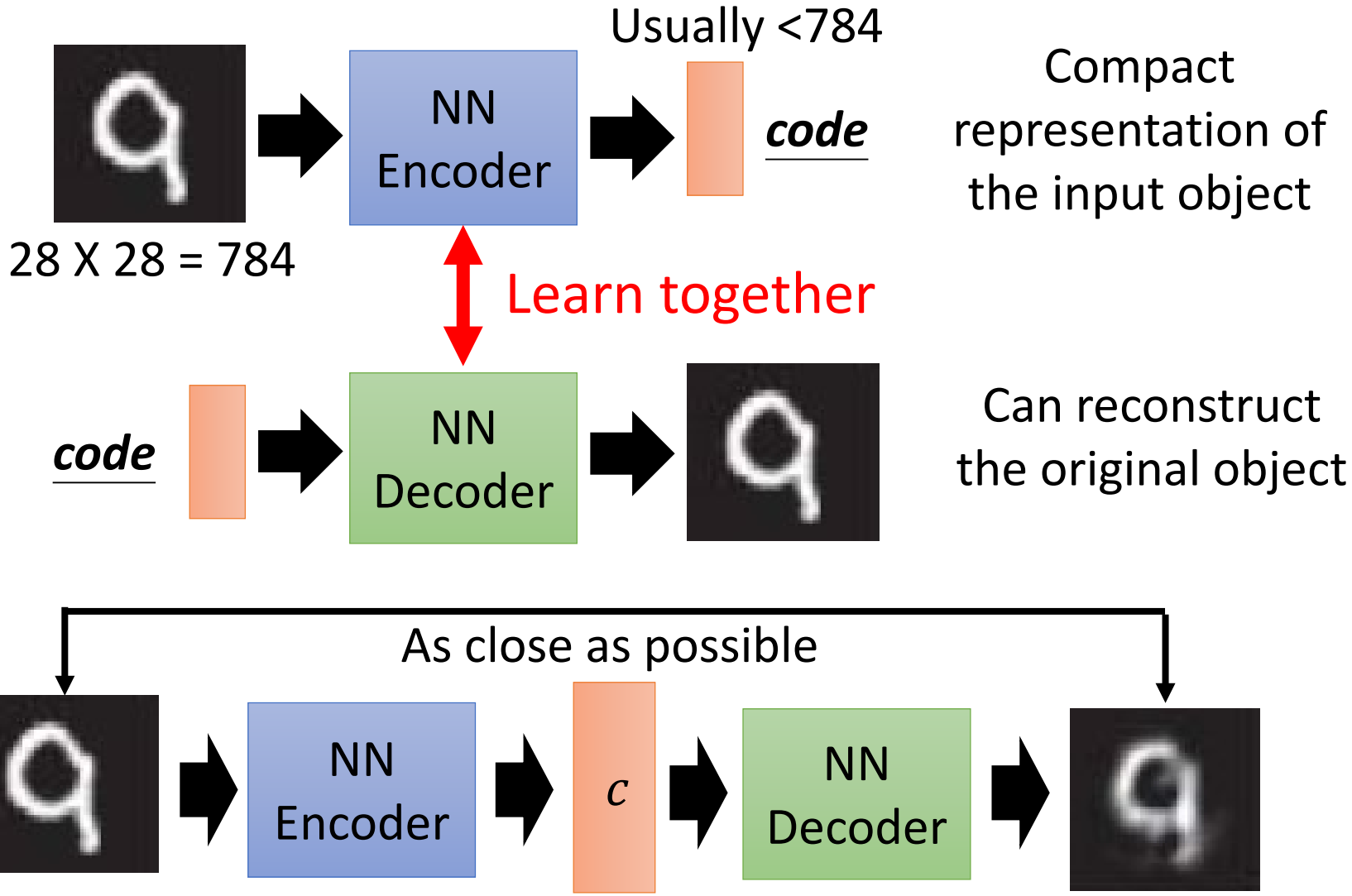


- In MNIST, a digit is 28 x 28 dims.
  - Most 28 x 28 dim vectors are not digits



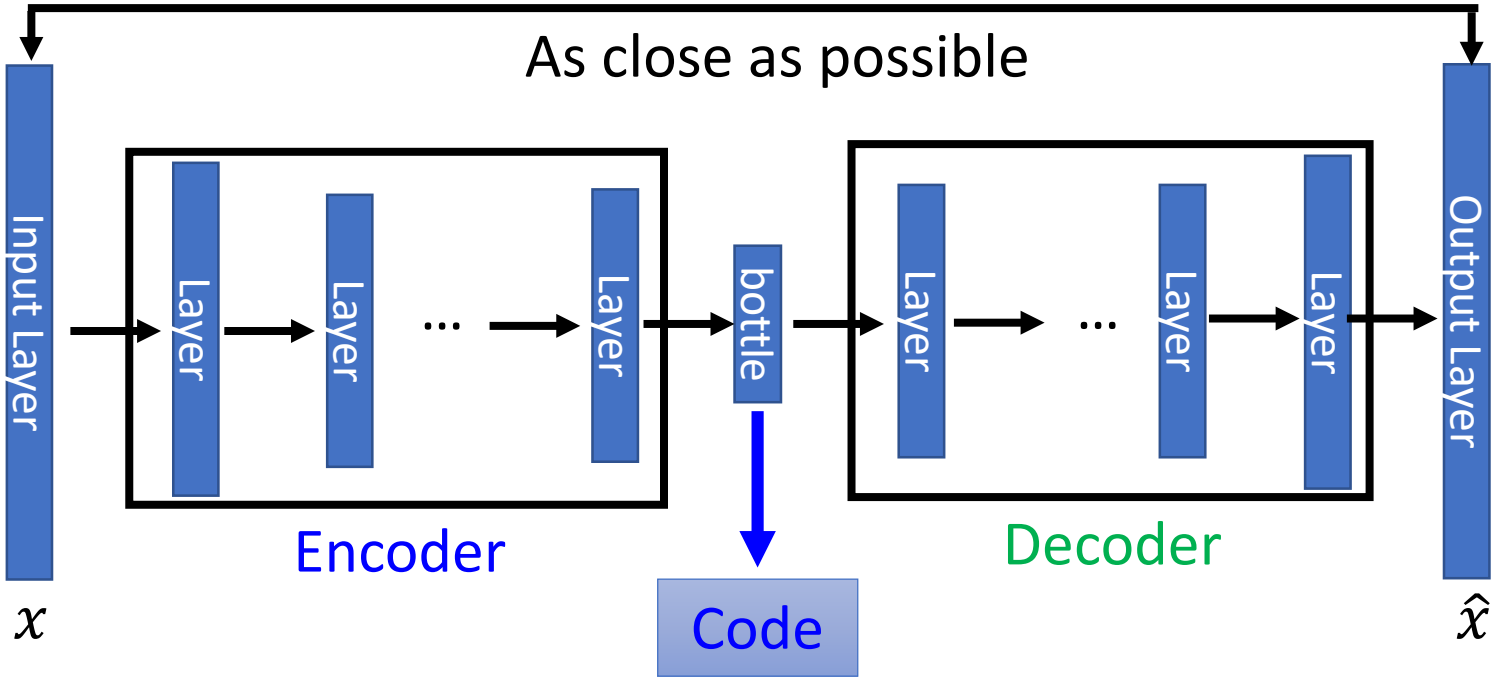


Auto-encoder



Deep Auto-encoder

- NN encoder + NN decoder = a deep network



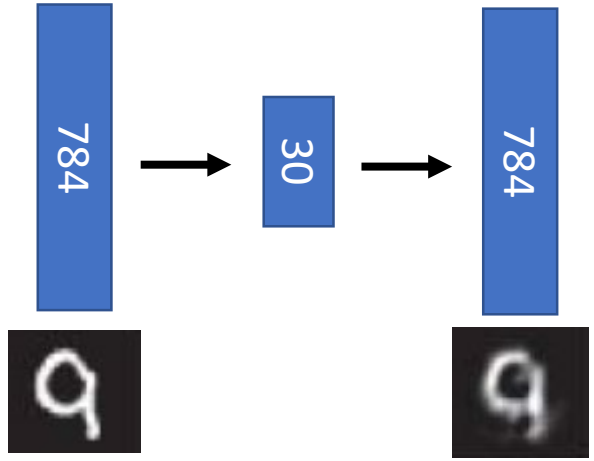
Reference: Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." *Science* 313.5786 (2006): 504-507

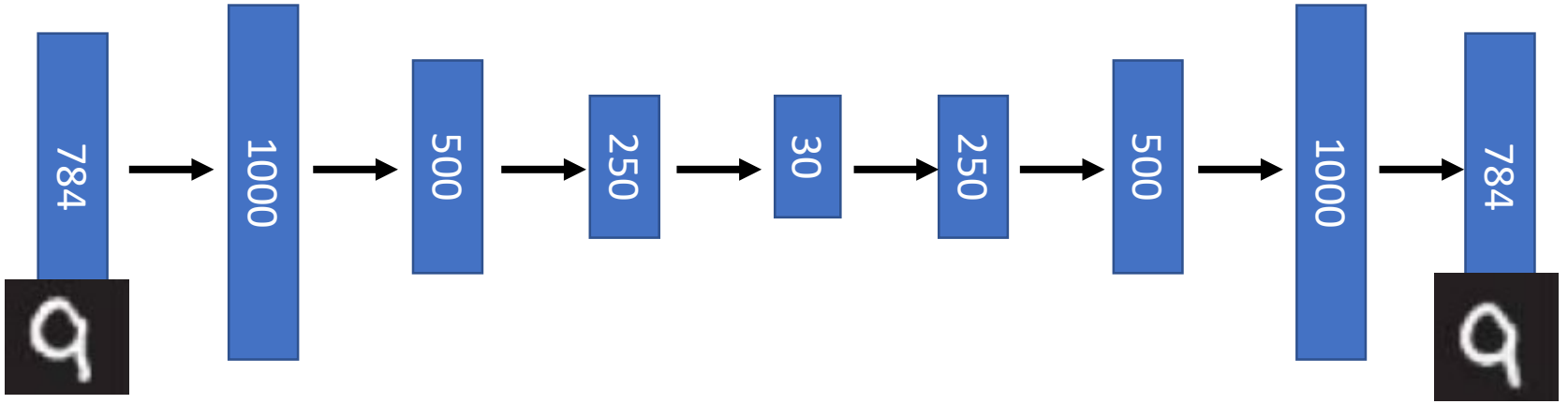
Deep Auto-encoder

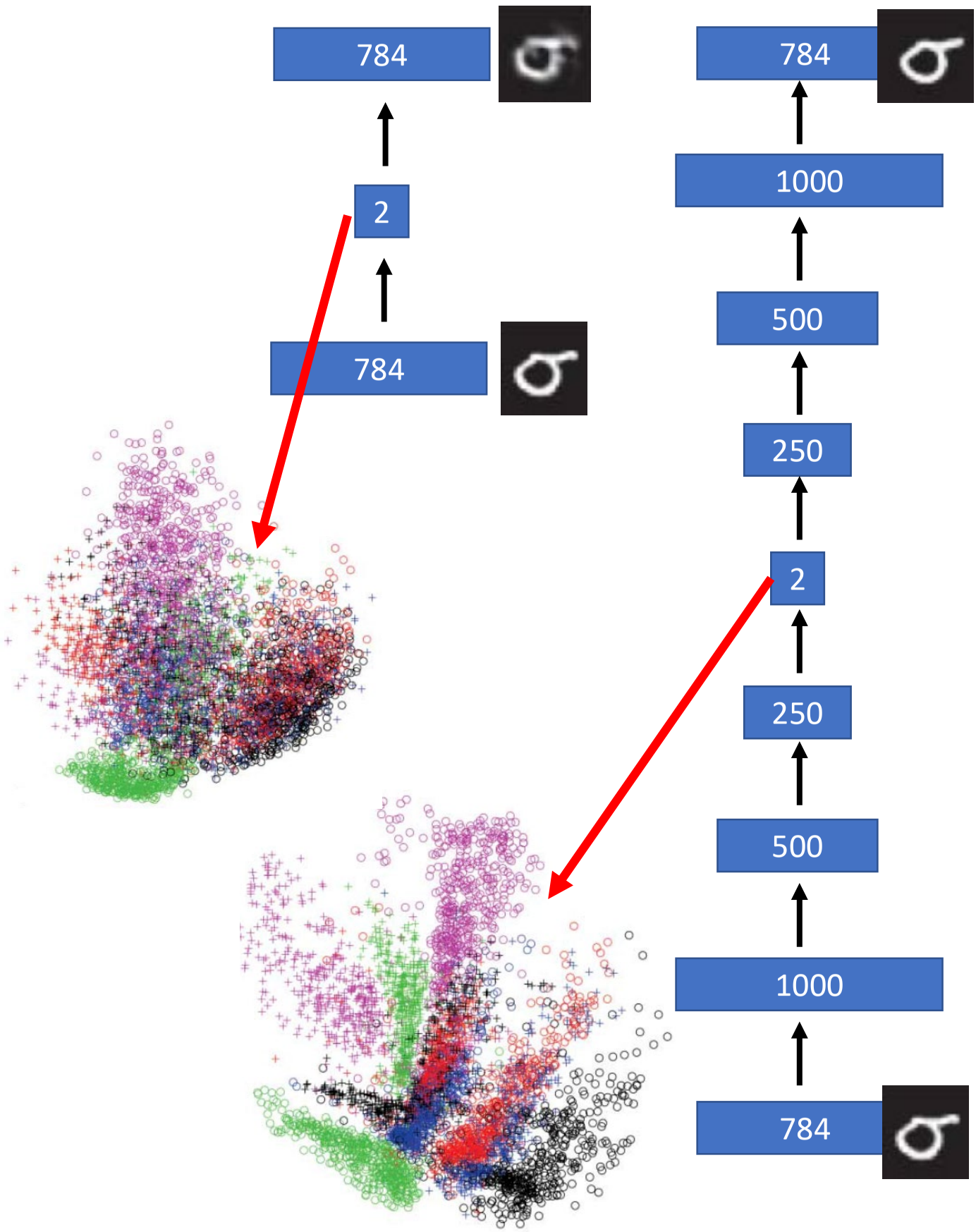
Original Image

PCA

Deep Auto-encoder





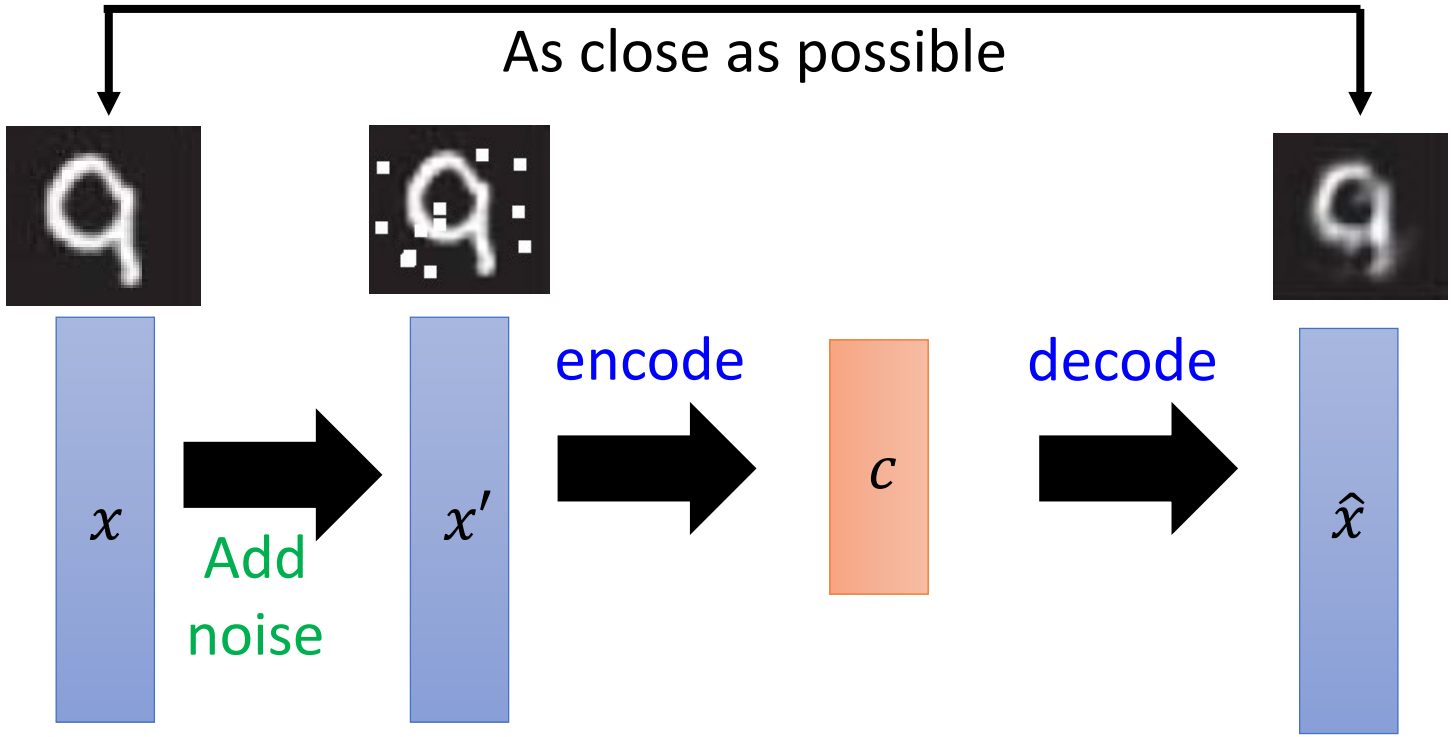

784
2
784
784
1000
500
250
2
250
500
1000
784

Auto-encoder

More: Contractive auto-encoder

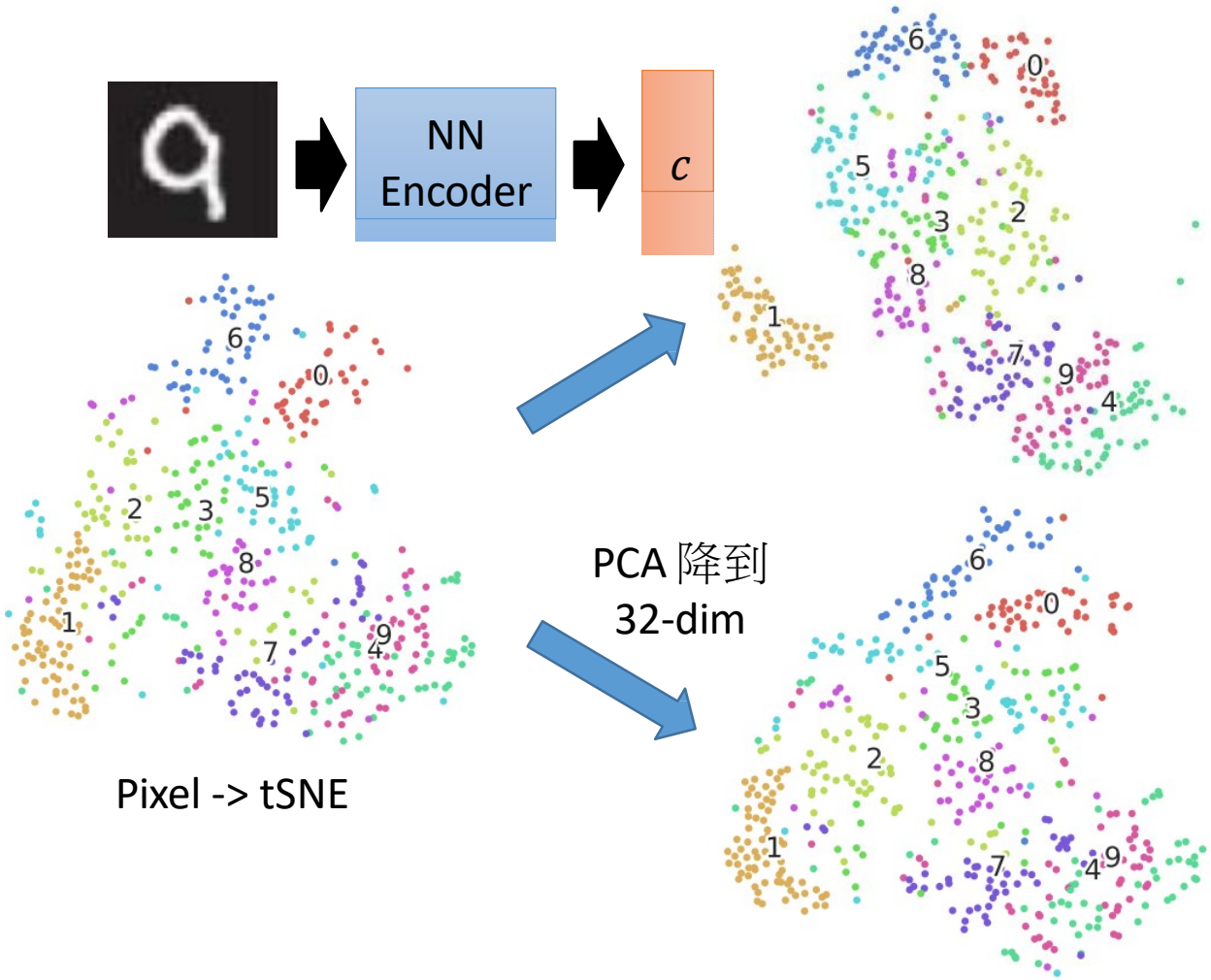
Ref: Rifai, Salah, et al. "Contractive auto-encoders: Explicit invariance during feature extraction." *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 2011.

- De-noising auto-encoder



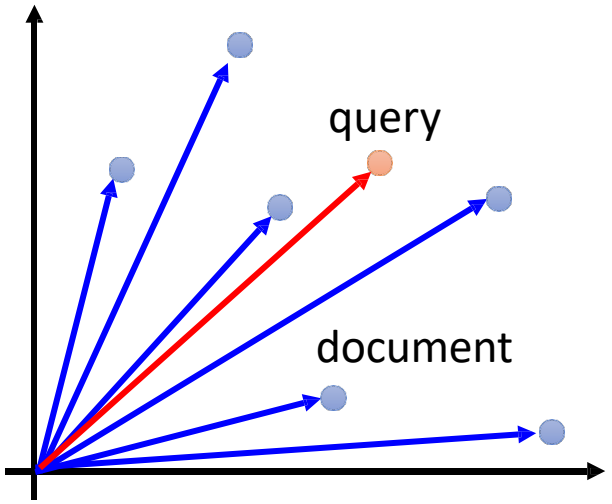
Vincent, Pascal, et al. "Extracting and composing robust features with denoising autoencoders." *ICML,* 2008.

Deep Auto-encoder - Example



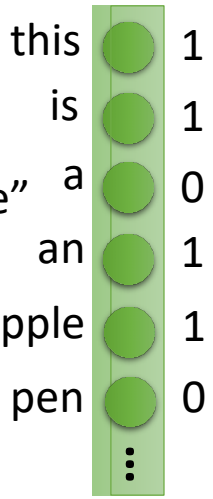
Auto-encoder – Text Retrieval

Vector Space Model



Bag-of-word

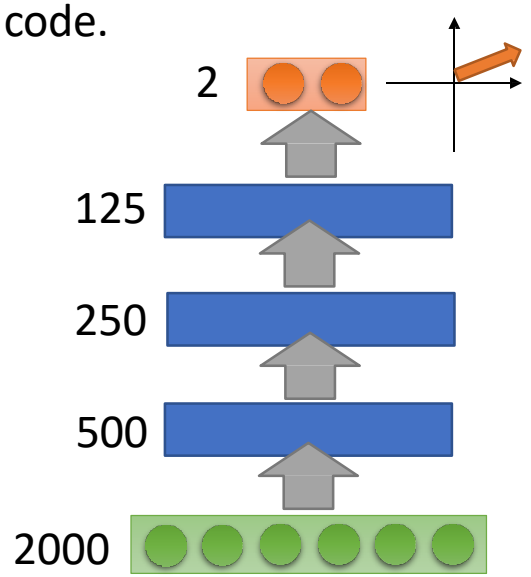
word string:
"This is an apple"



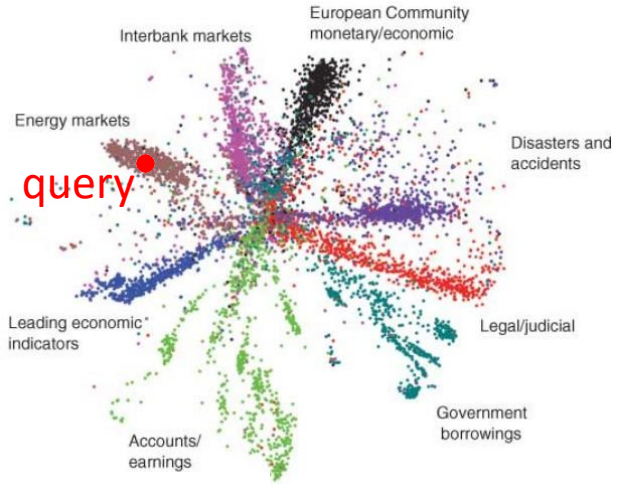
Semantics are not considered.

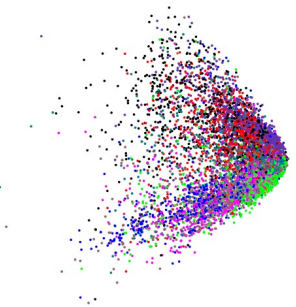
Auto-encoder – Text Retrieval

The documents talking about the same thing will have close code.



Bag-of-word
(document or query)

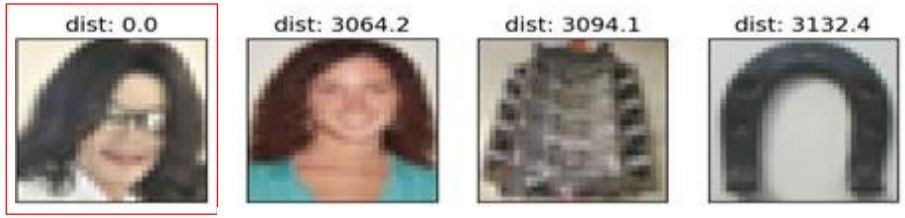




LSA: project documents to 2 latent topics

Auto-encoder – Similar Image Search

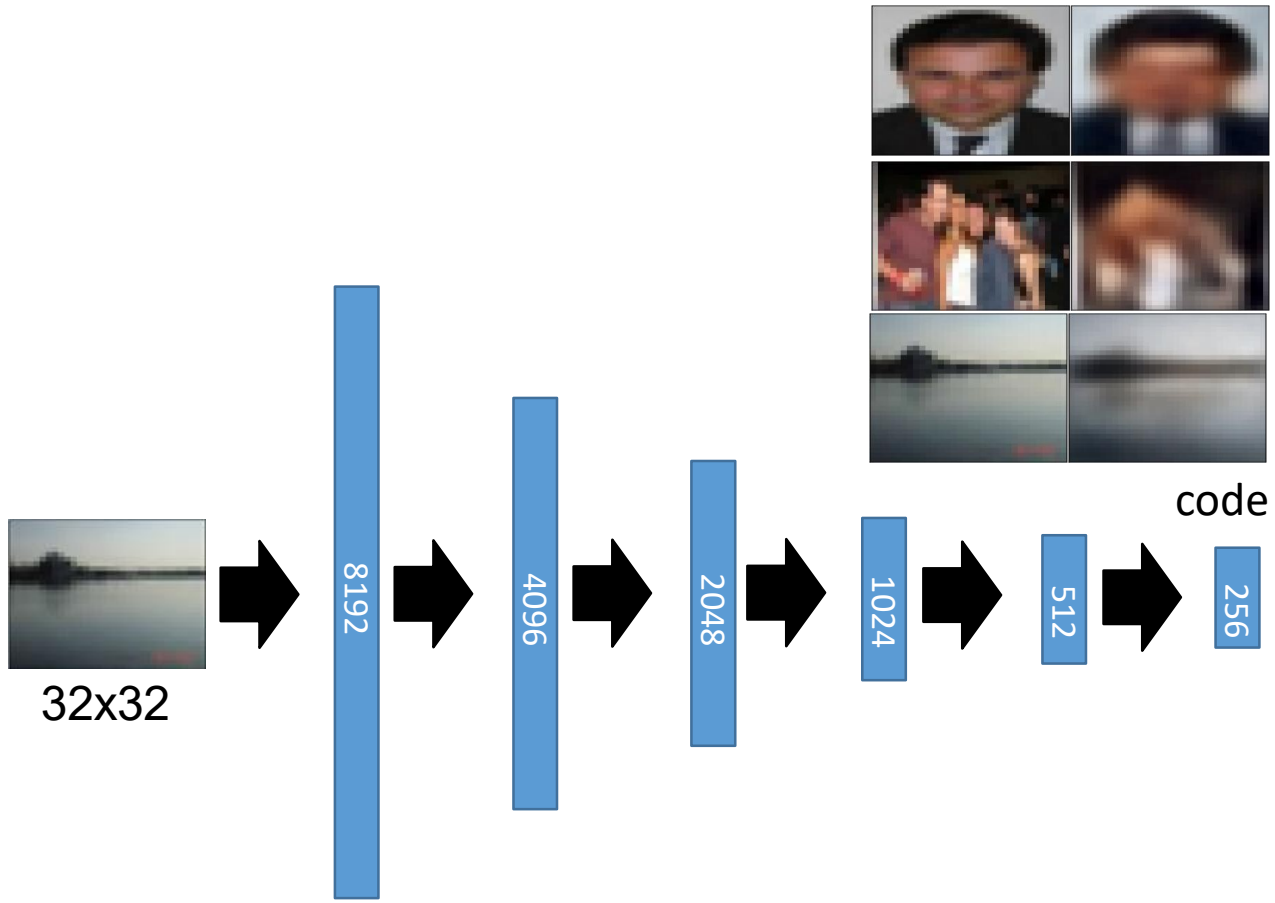
Retrieved using Euclidean distance in pixel intensity space



(Images from Hinton's slides on Coursera)

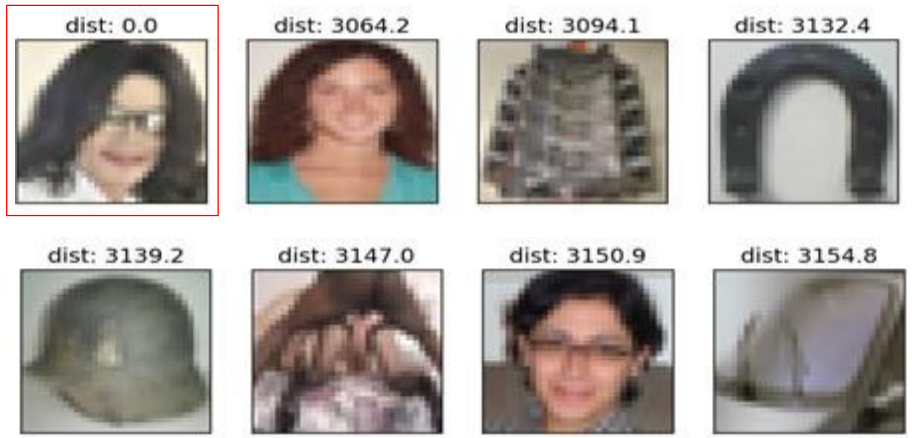
Reference: Krizhevsky, Alex, and Geoffrey E. Hinton. "Using very deep autoencoders for content-based image retrieval." *ESANN*. 2011.

Auto-encoder – Similar Image Search

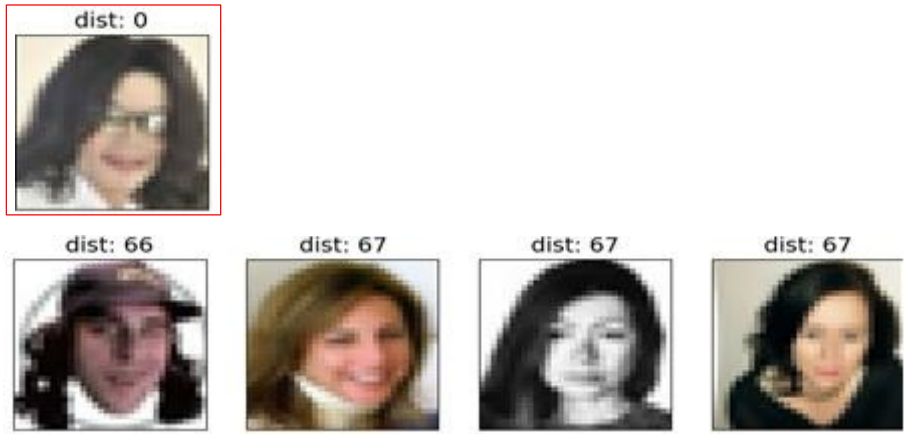


(crawl millions of images from the Internet)

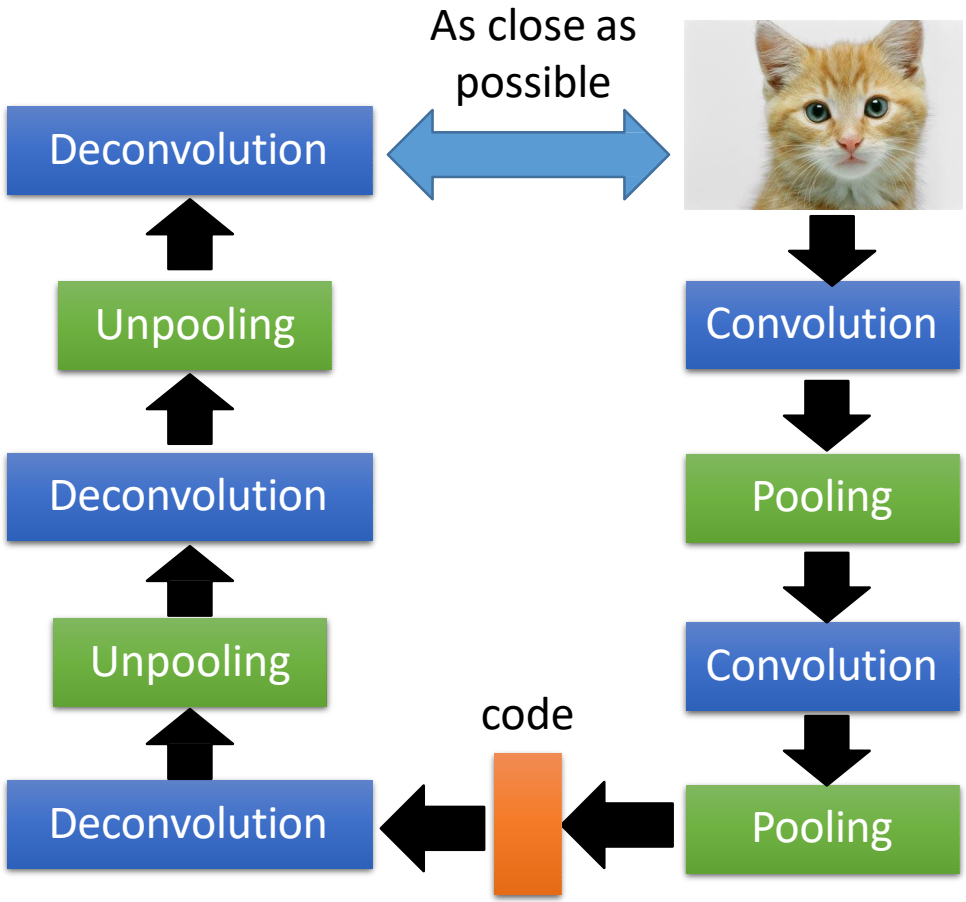
Retrieved using Euclidean distance in pixel intensity space



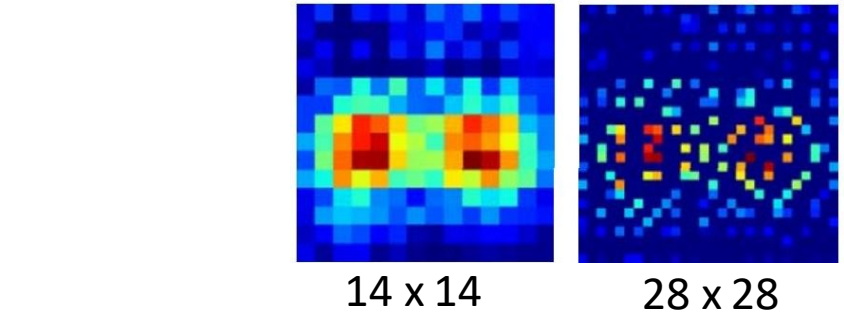
retrieved using 256 codes

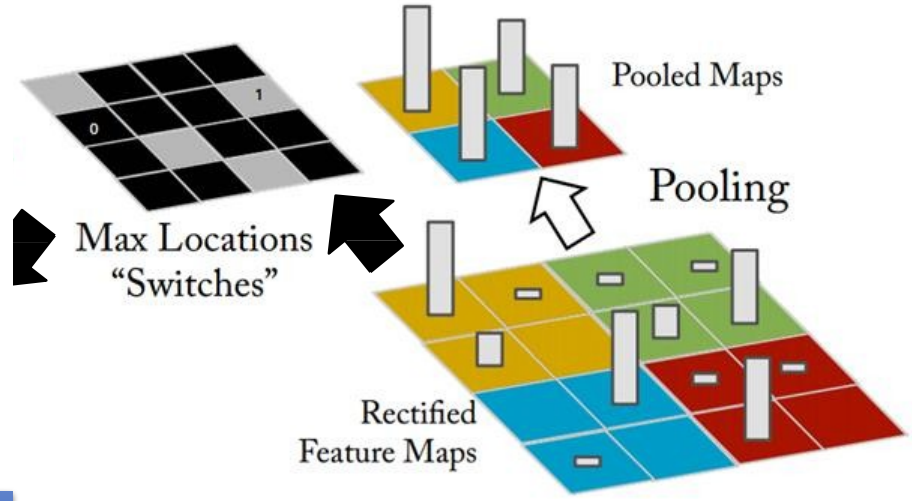


Auto- encoder for CNN



CNN -Unpooling



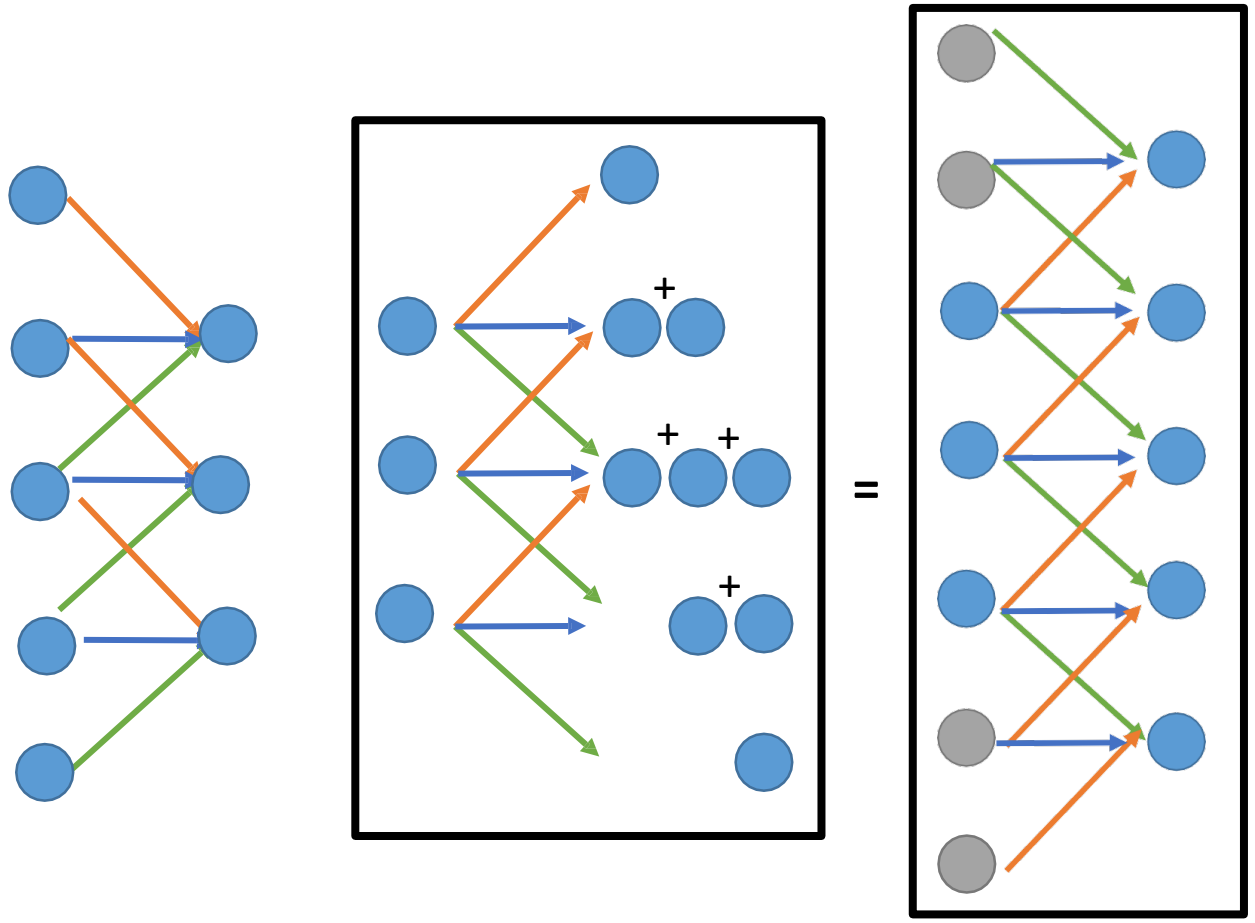


Alternative: simply repeat the values

Source of image : https://leonardoaraujosantos.gitbooks.io/artificial-inteligence/content/image_segmentation.html

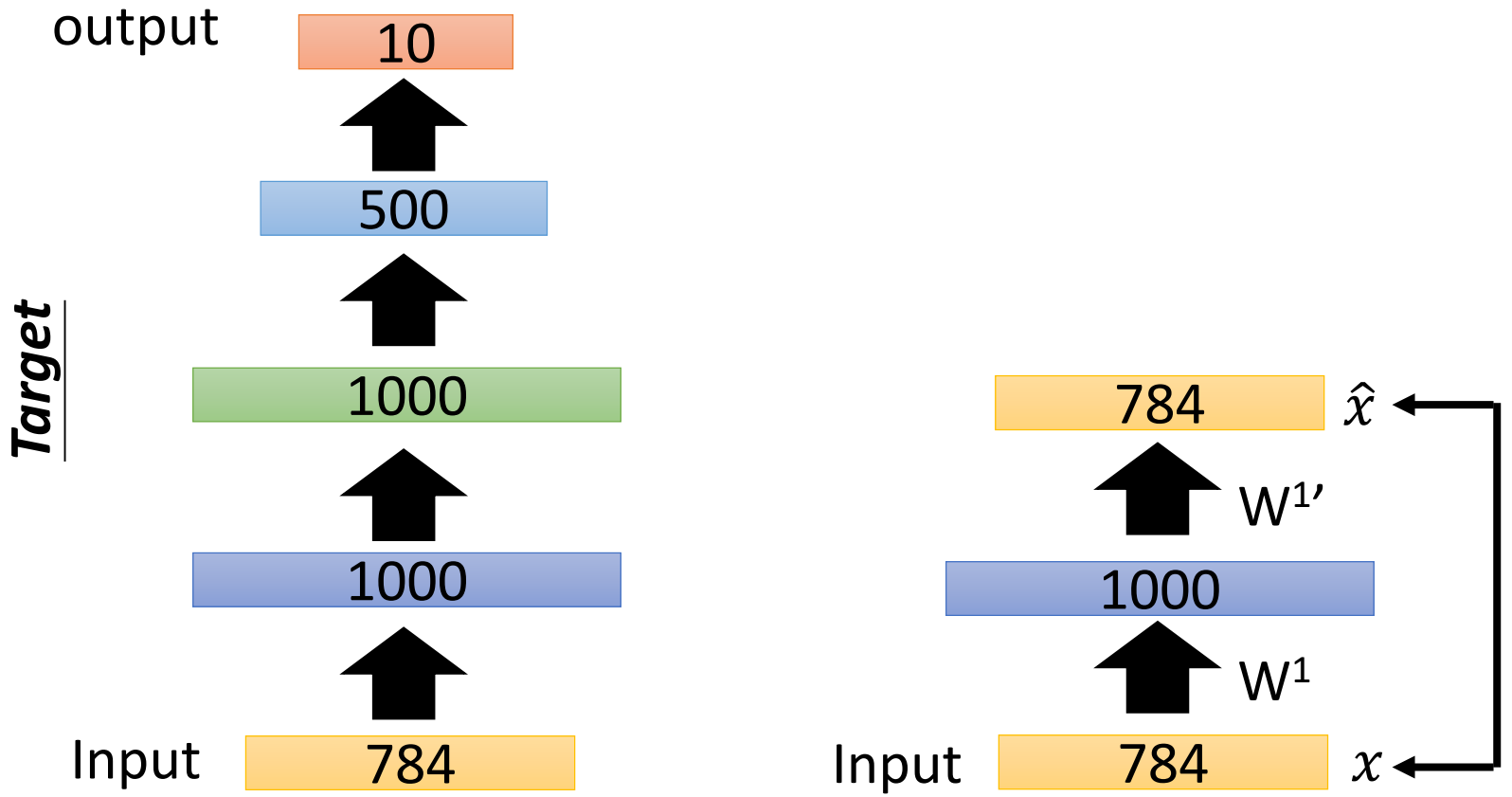
CNN- Deconvolution

Actually, deconvolution is convolution.



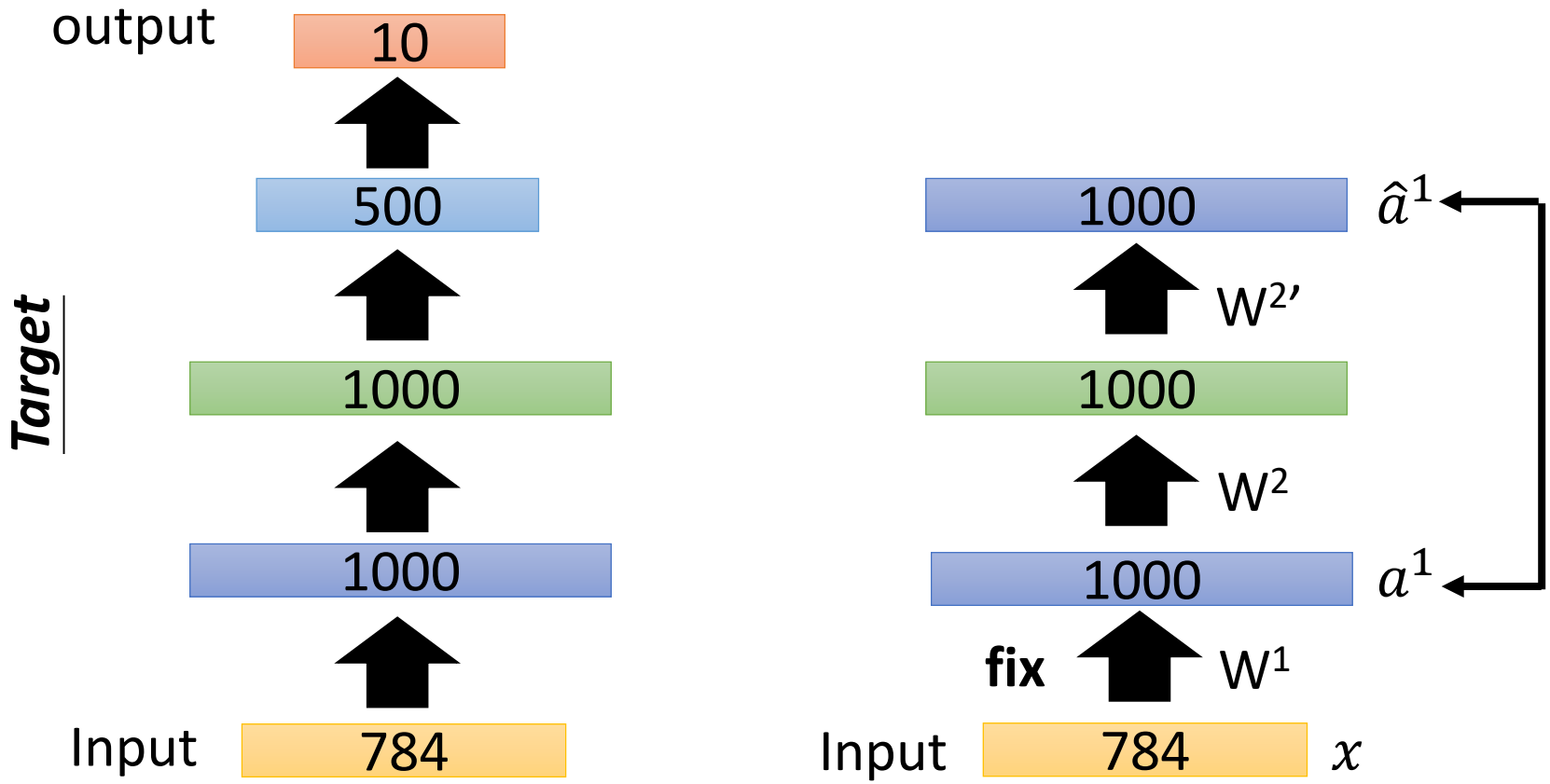
Auto-encoder – Pre-training DNN

- Greedy Layer-wise Pre-training *again*



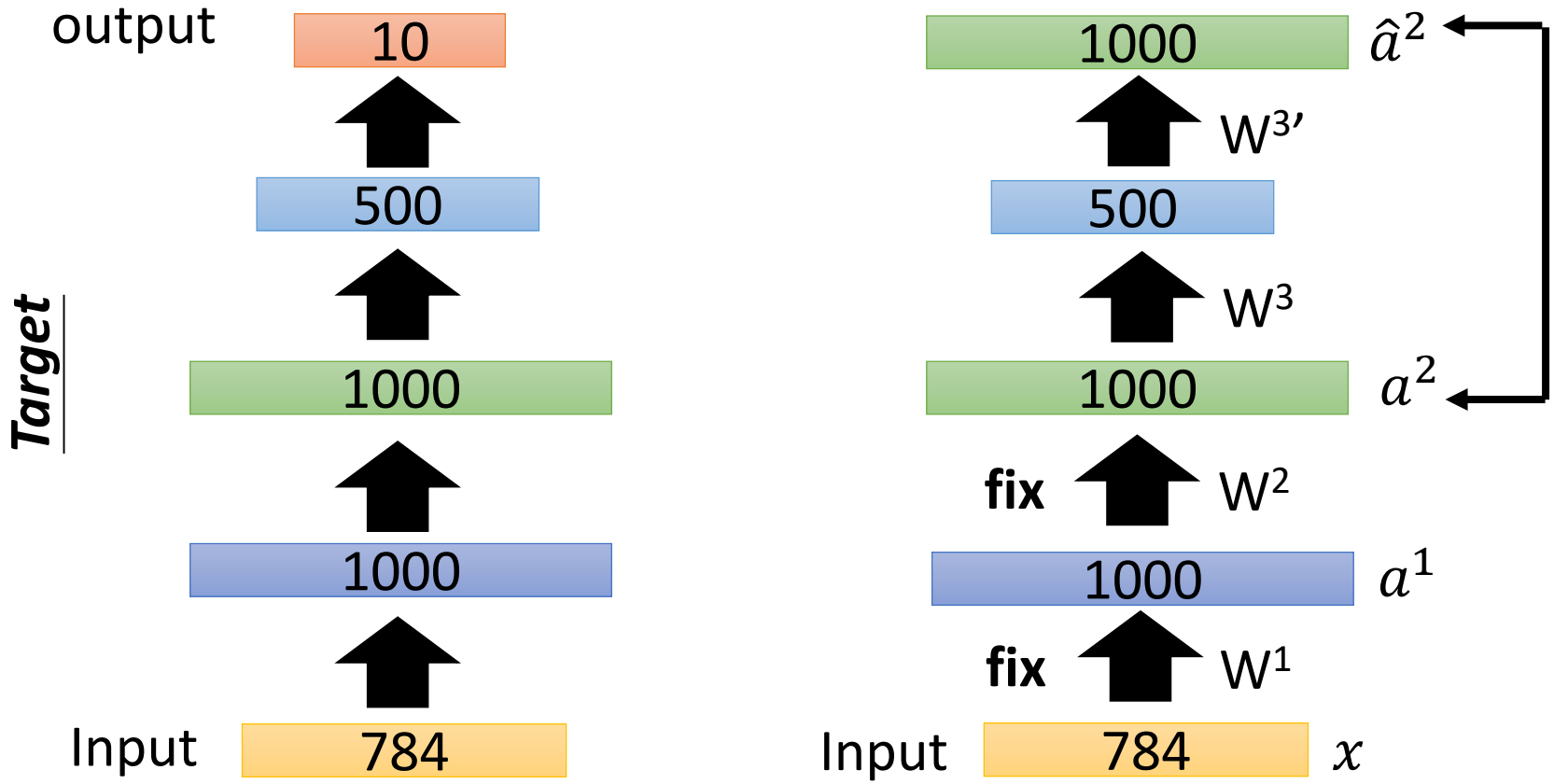
Auto-encoder – Pre-training DNN

- Greedy Layer-wise Pre-training *again*



Auto-encoder – Pre-training DNN

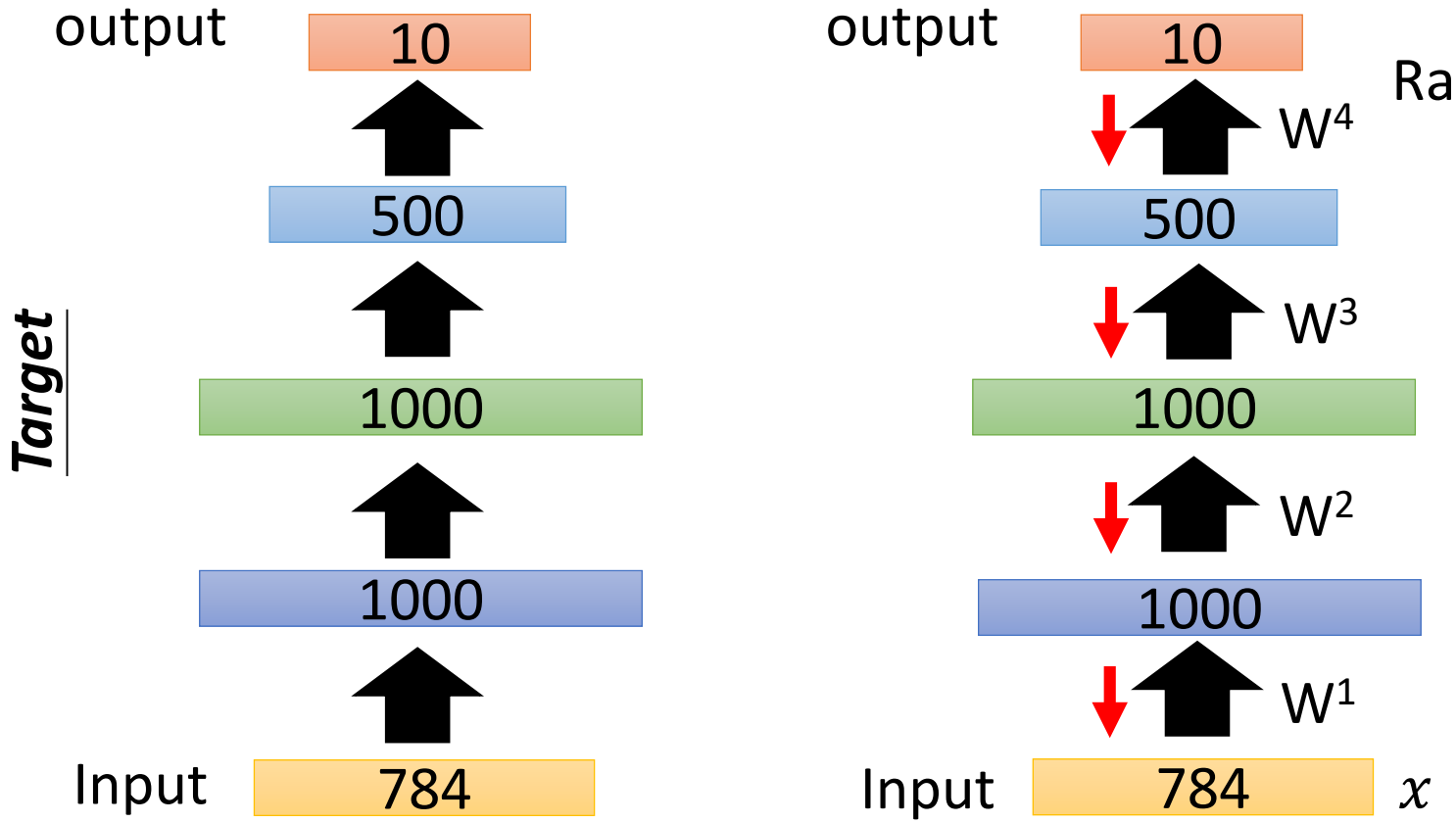
- Greedy Layer-wise Pre-training *again*



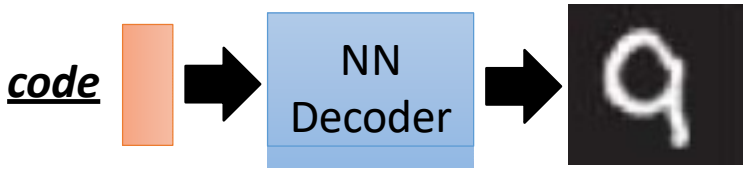
Auto-encoder – Pre-training DNN

- Greedy Layer-wise Pre-training *again*

Find-tune by backpropagation

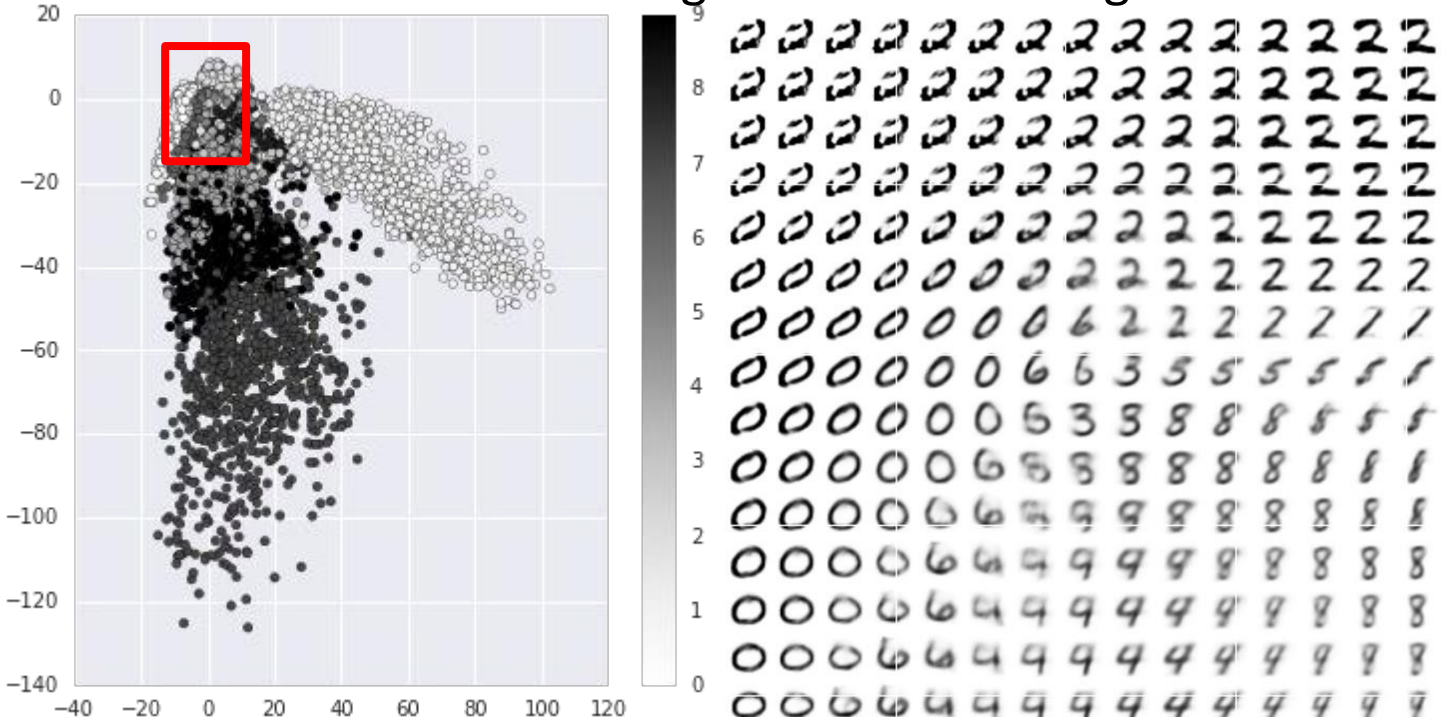


Next

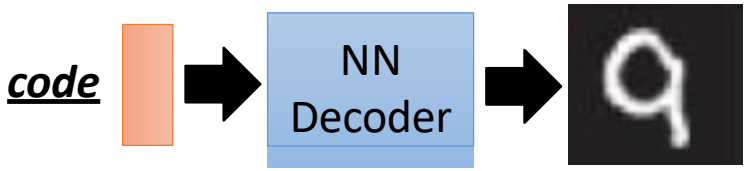


.....

- Can we use decoder to generate something?

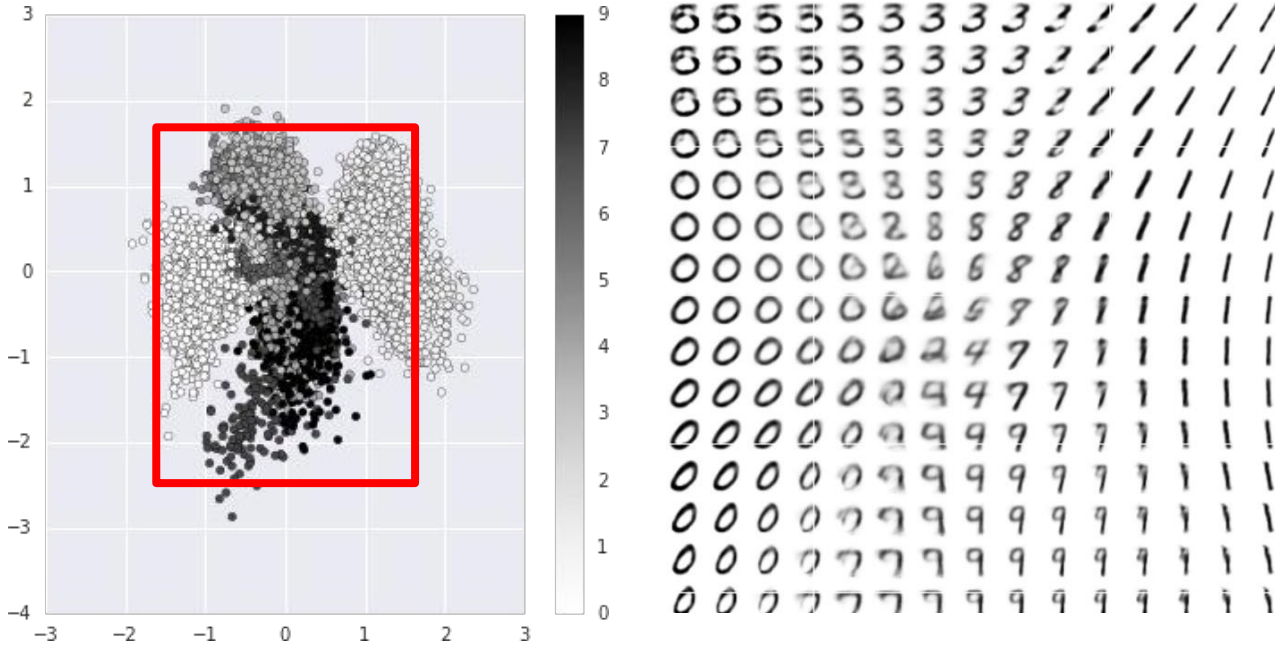


Next



.....

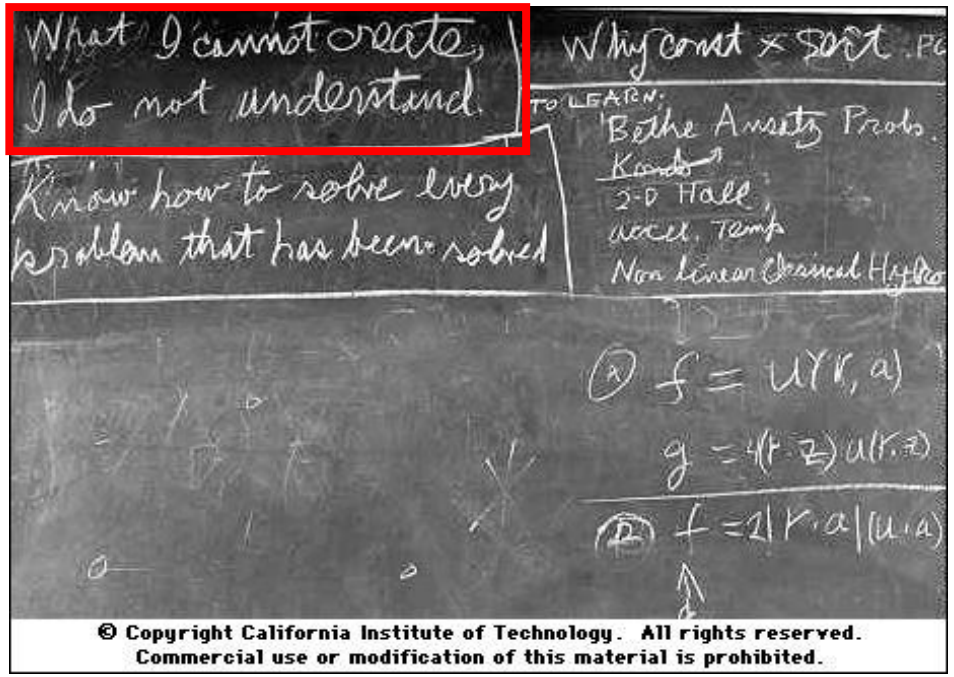
- Can we use decoder to generate something?



Creation

Creation

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

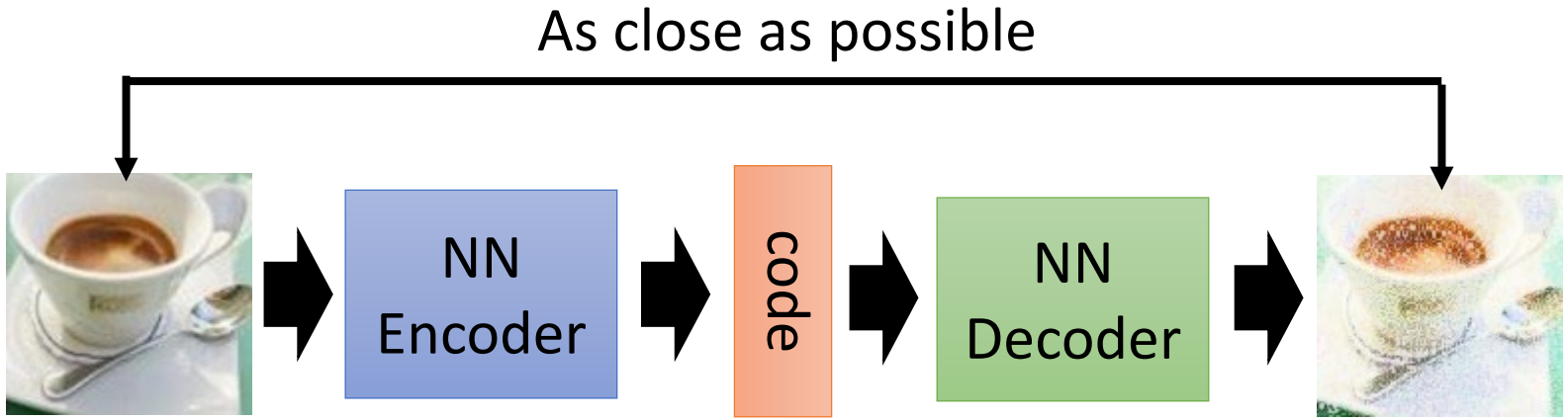


What I cannot create,
I do not understand.

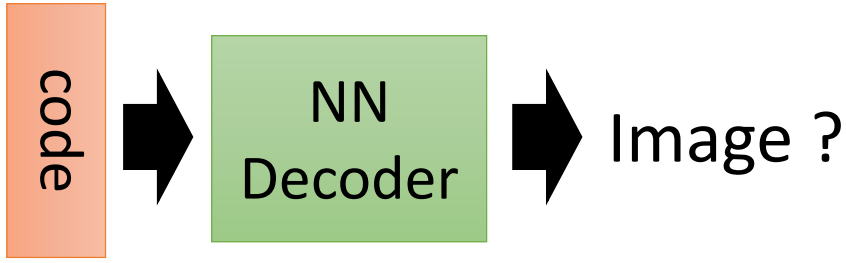
Richard Feynman

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

Auto-encoder



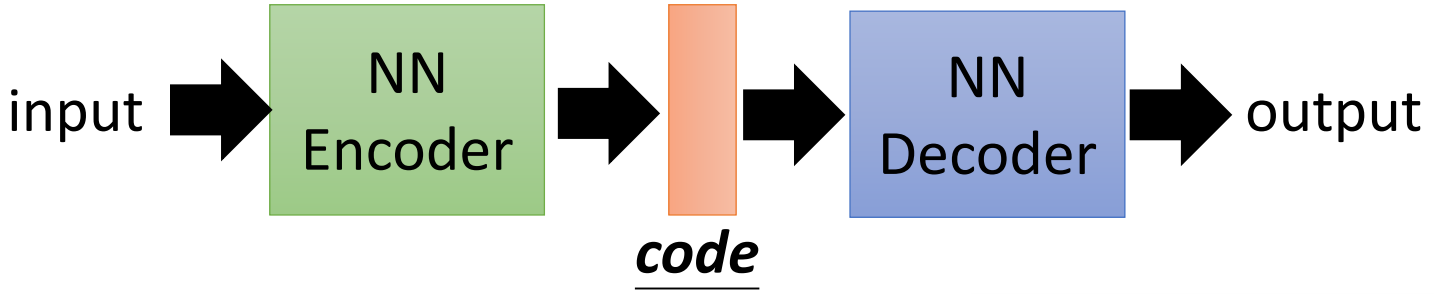
Randomly generate a vector as code



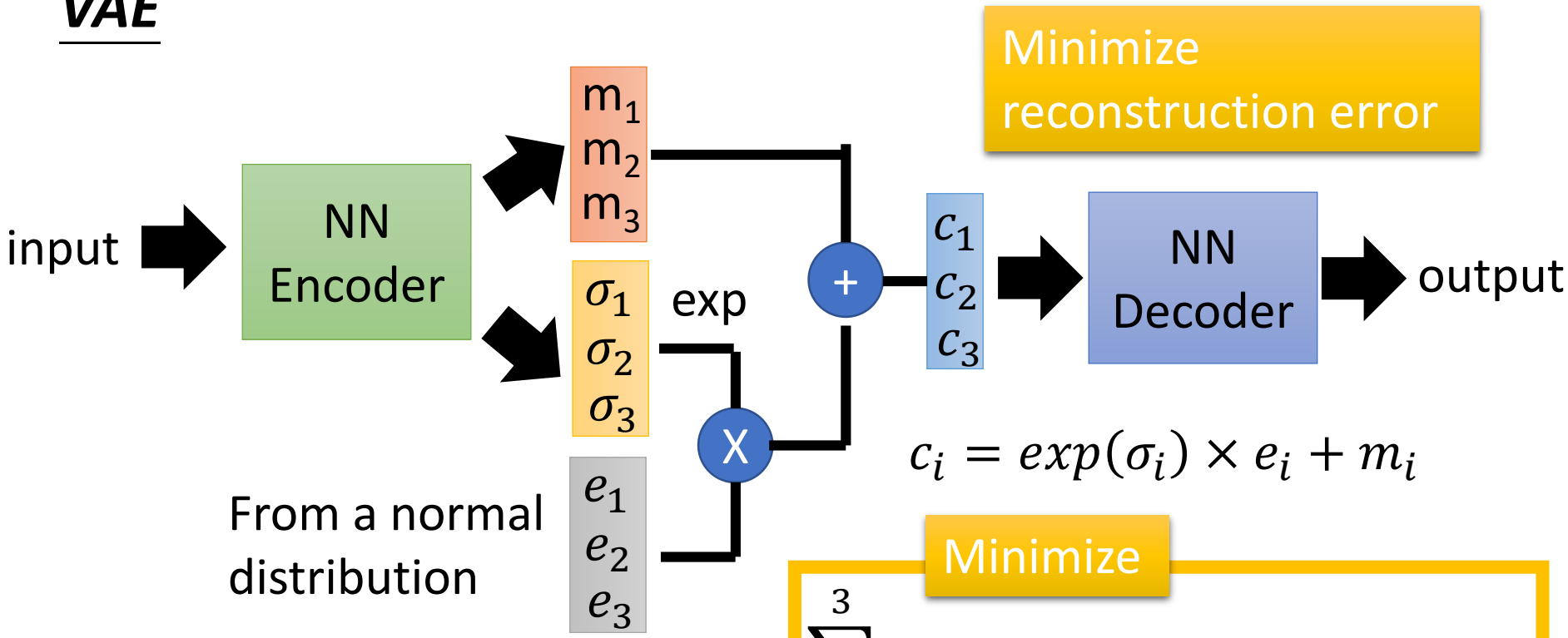
Variation Auto-encoder (VAE)

Ref: Auto-Encoding Variational Bayes, https://arxiv.org/abs/1312.6114

Auto-encoder



VAE

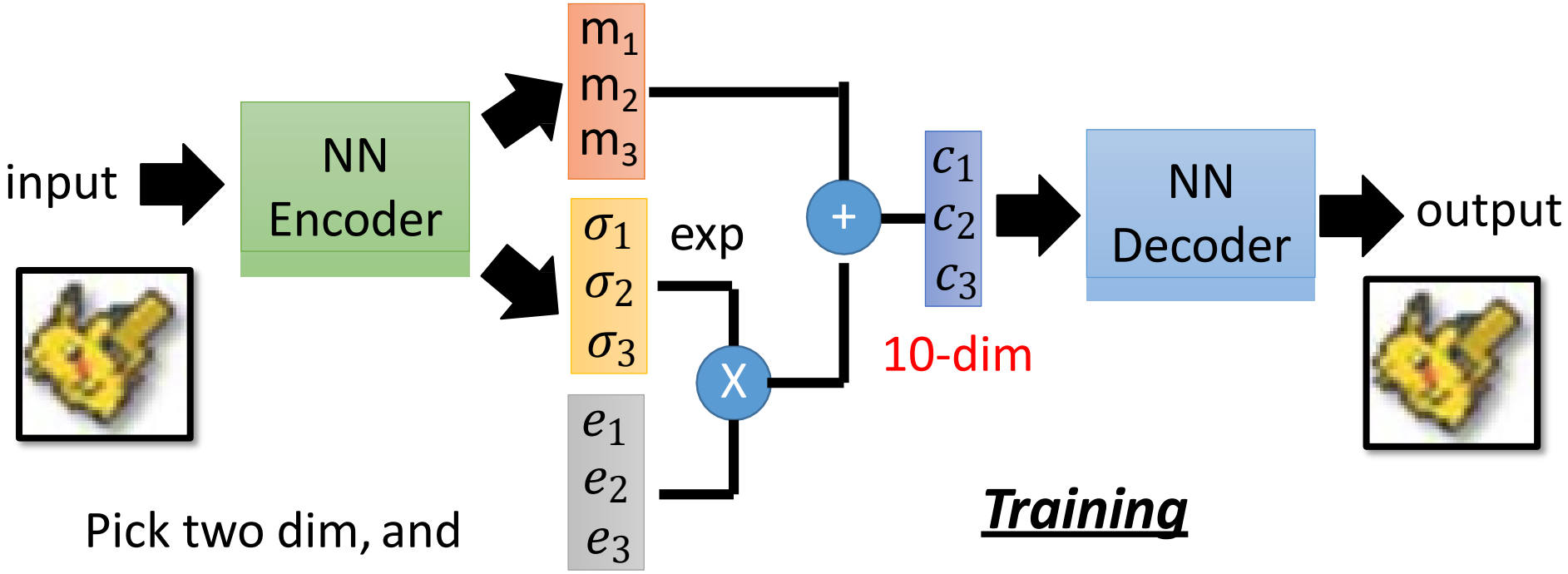


Cifar-10

https://github.com/openai/iaf

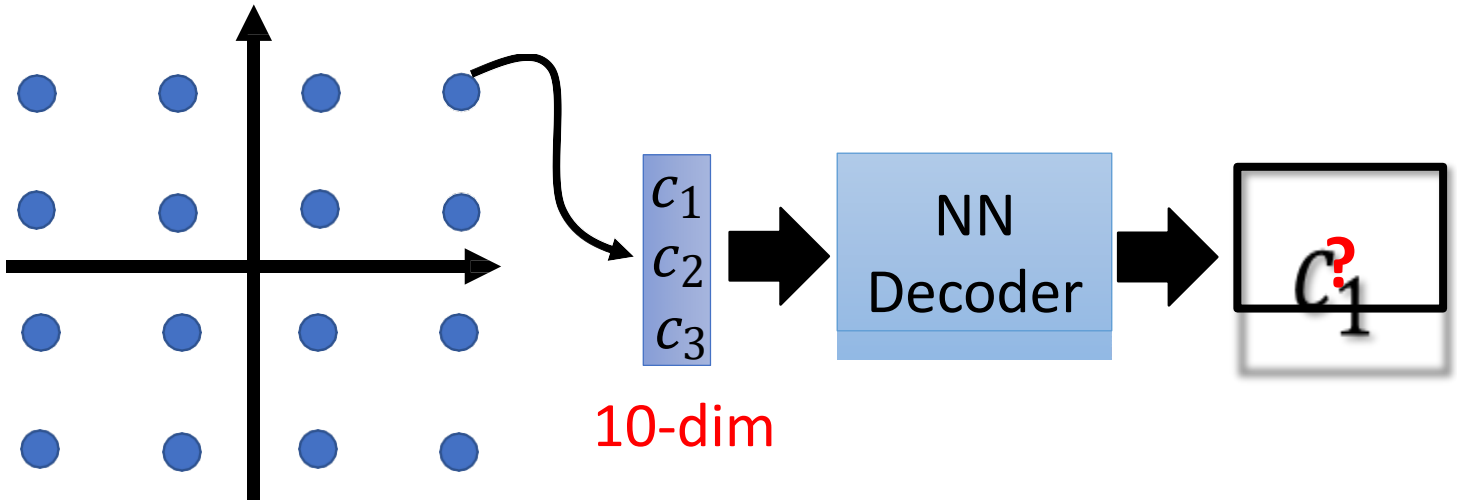
Source of image: https://arxiv.org/pdf/1606.04934v1.pdf

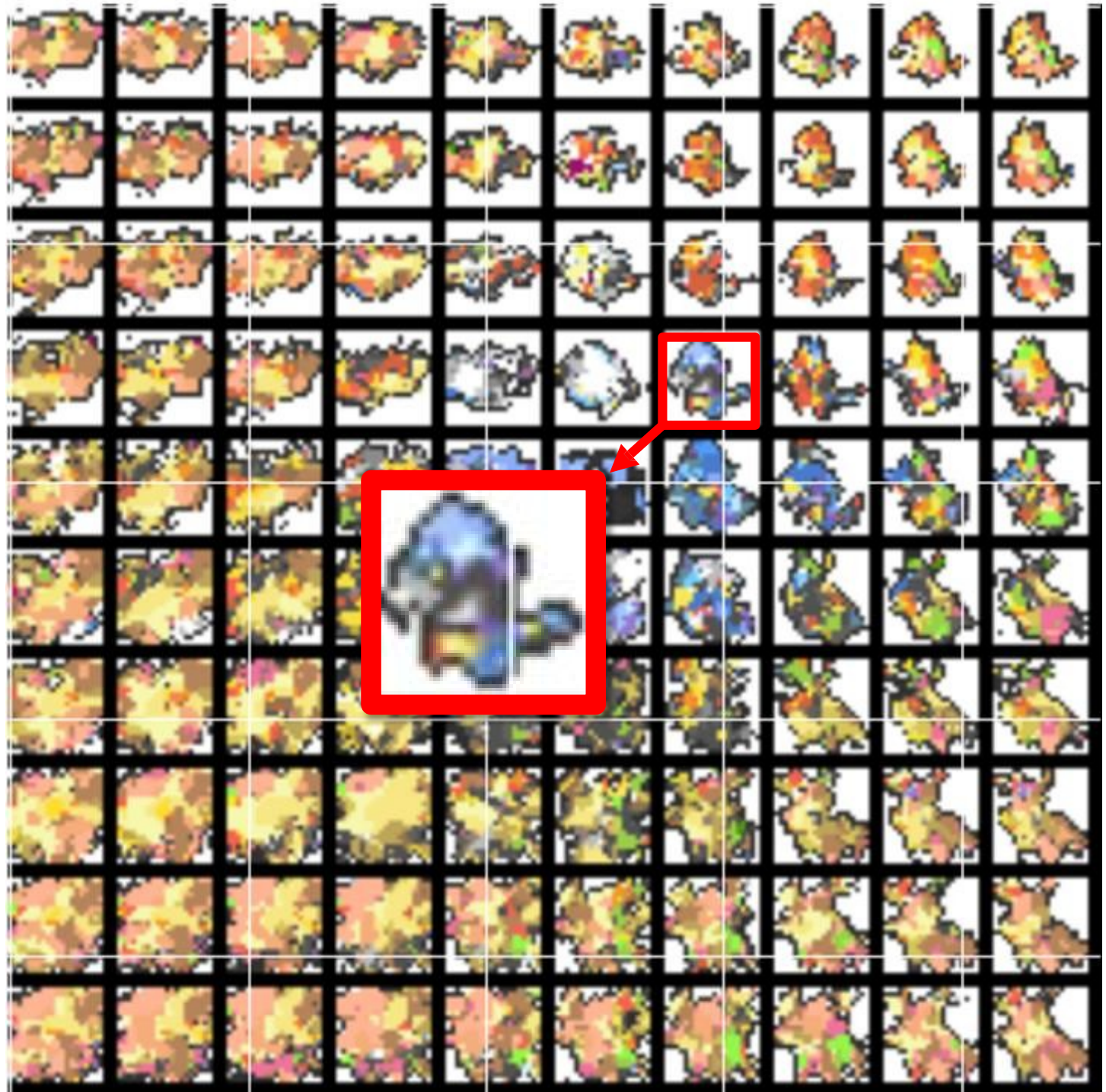
Pokémon Creation



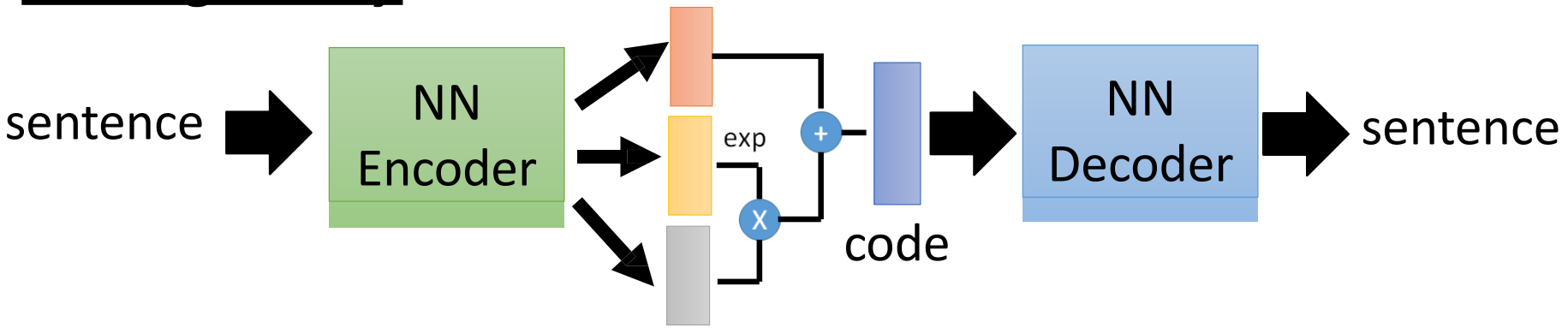
Training

Pick two dim, and fix the rest eight

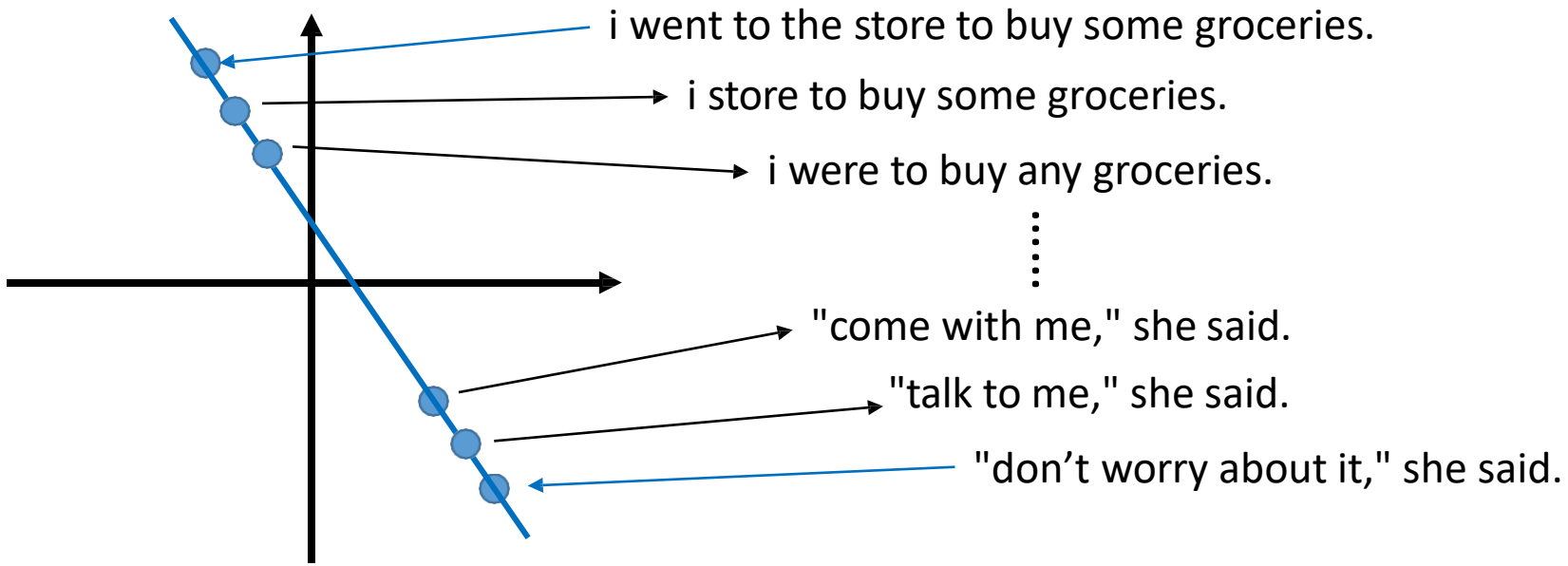




Writing Poetry



Code Space

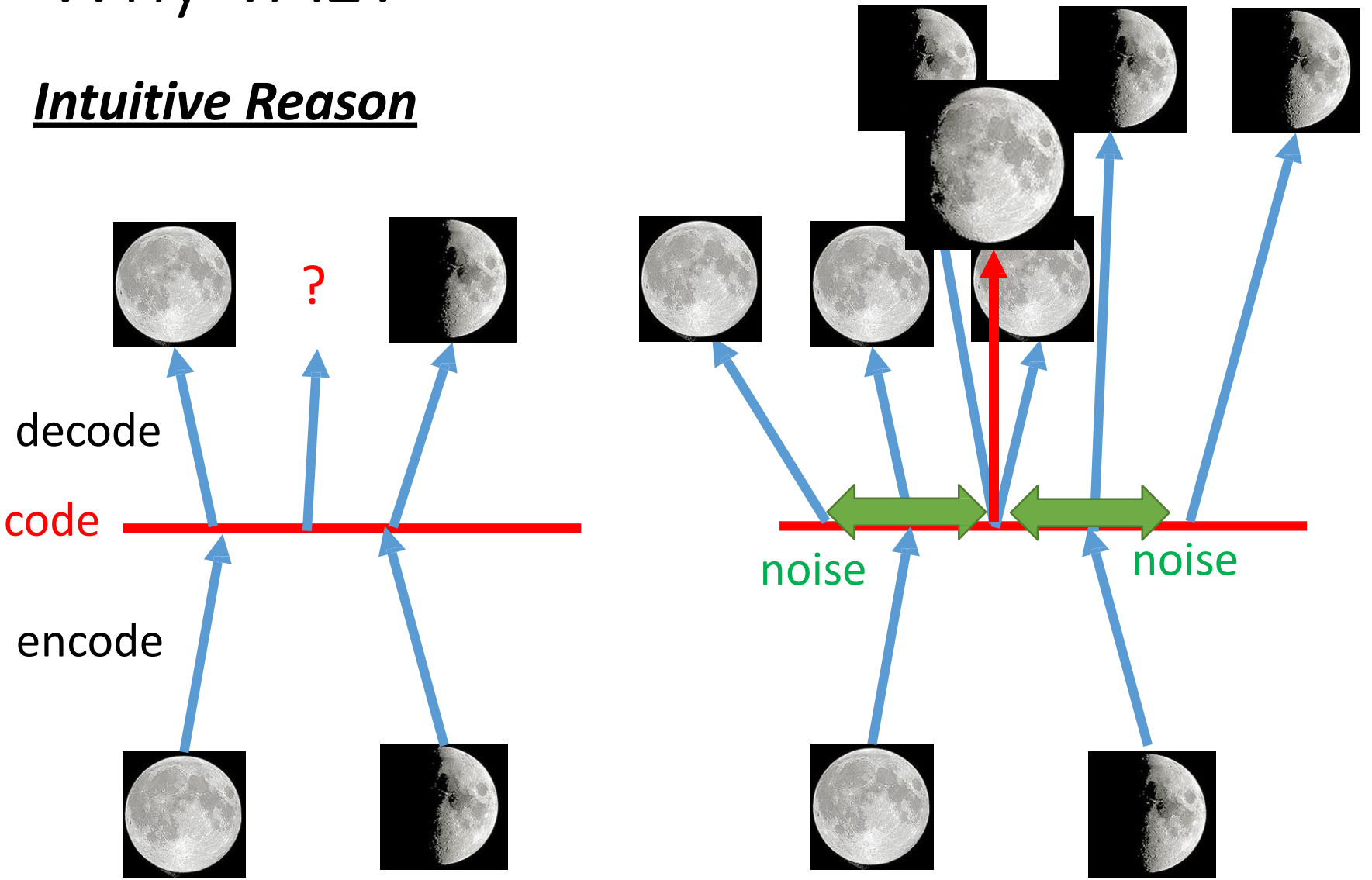


Ref: http://www.wired.co.uk/article/google-artificial-intelligence-poetry

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?

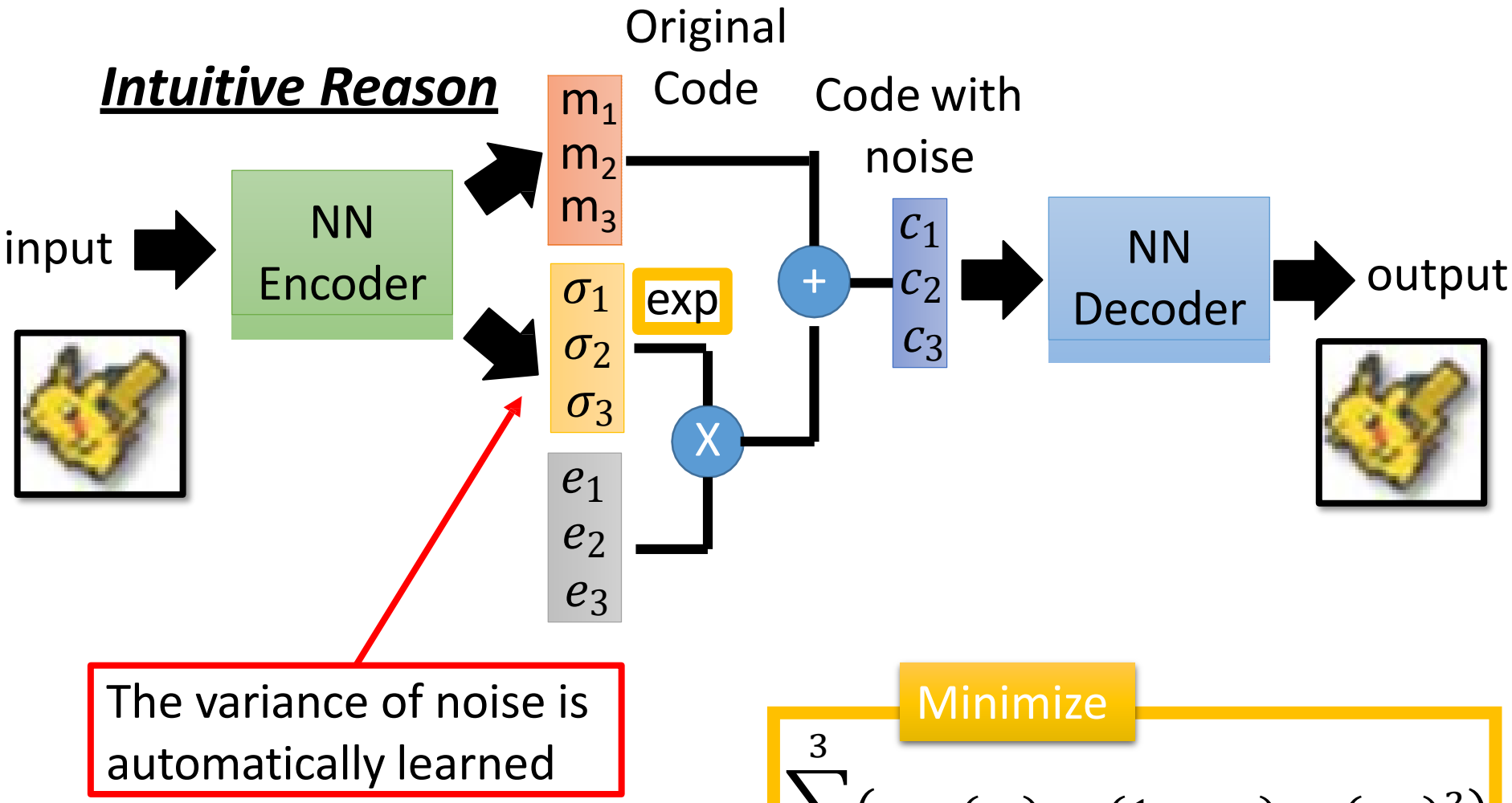
Intuitive Reason



Why VAE?

What will happen if we only minimize reconstruction error?

Intuitive Reason



The variance of noise is automatically learned

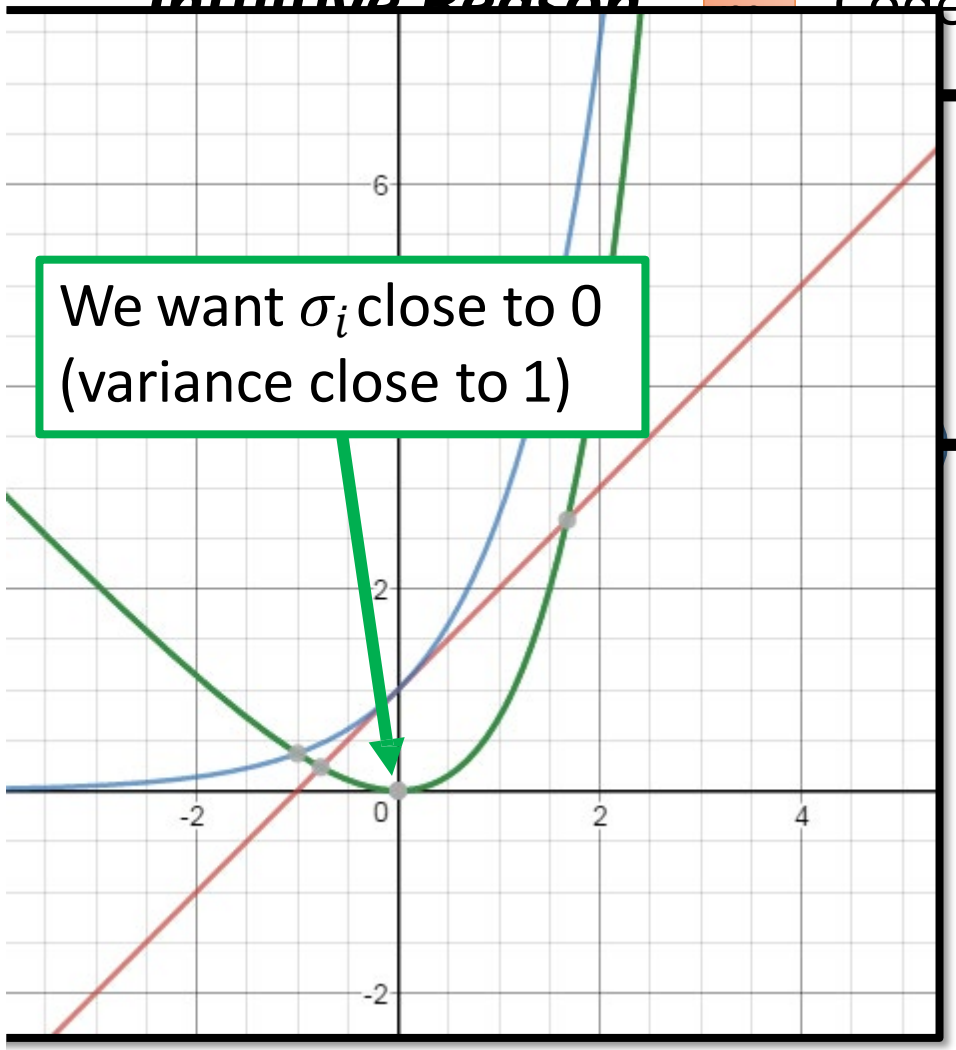
Minimize

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

Why VAE?

What will happen if we only minimize reconstruction error?

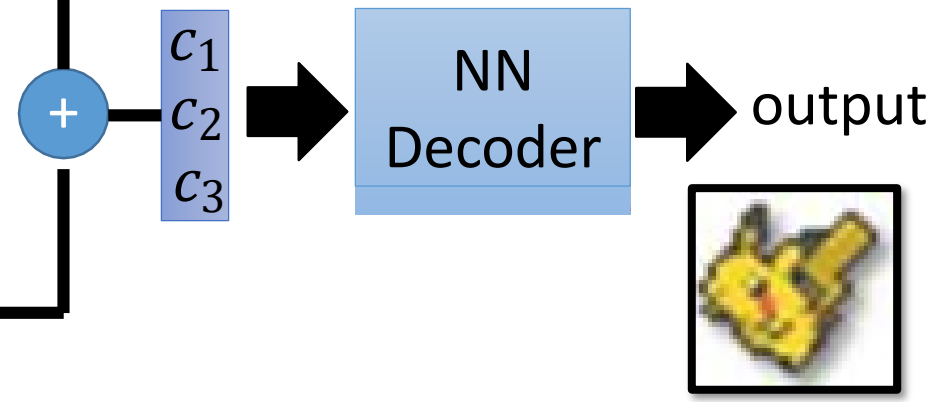
Intuitive Reason



We want σ_i close to 0 (variance close to 1)

Original Code

Code with noisy



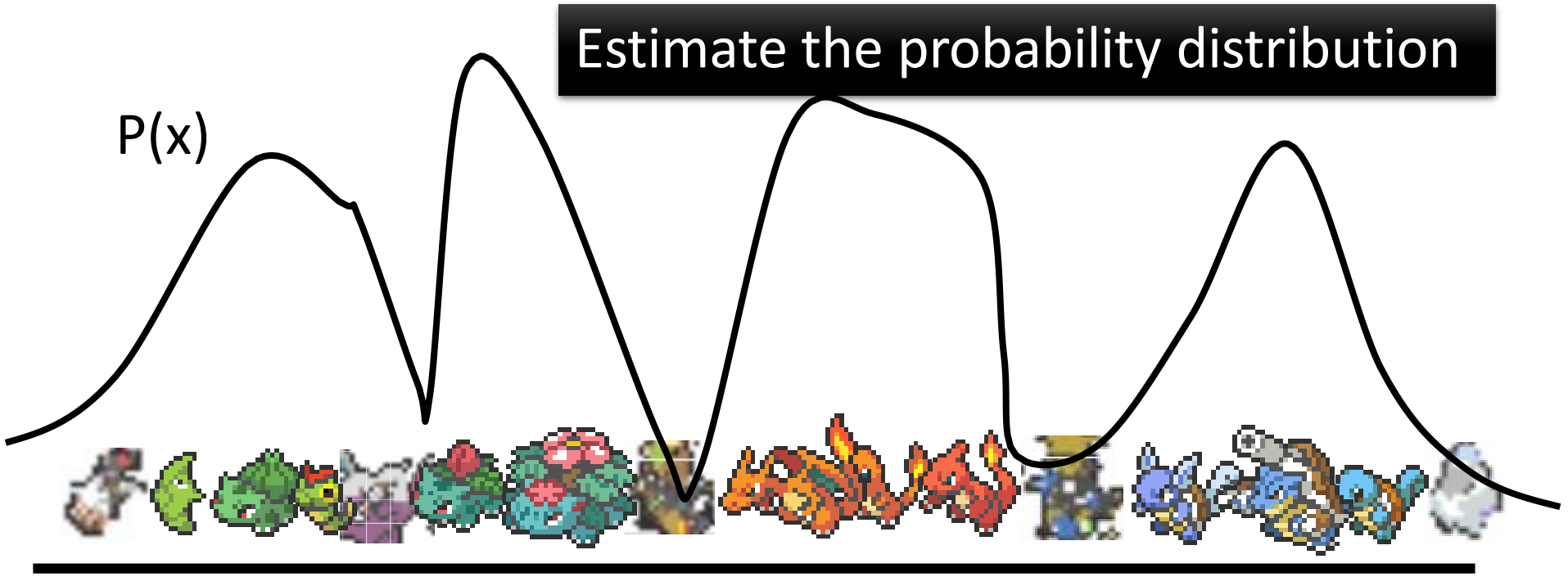
Minimize

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

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?

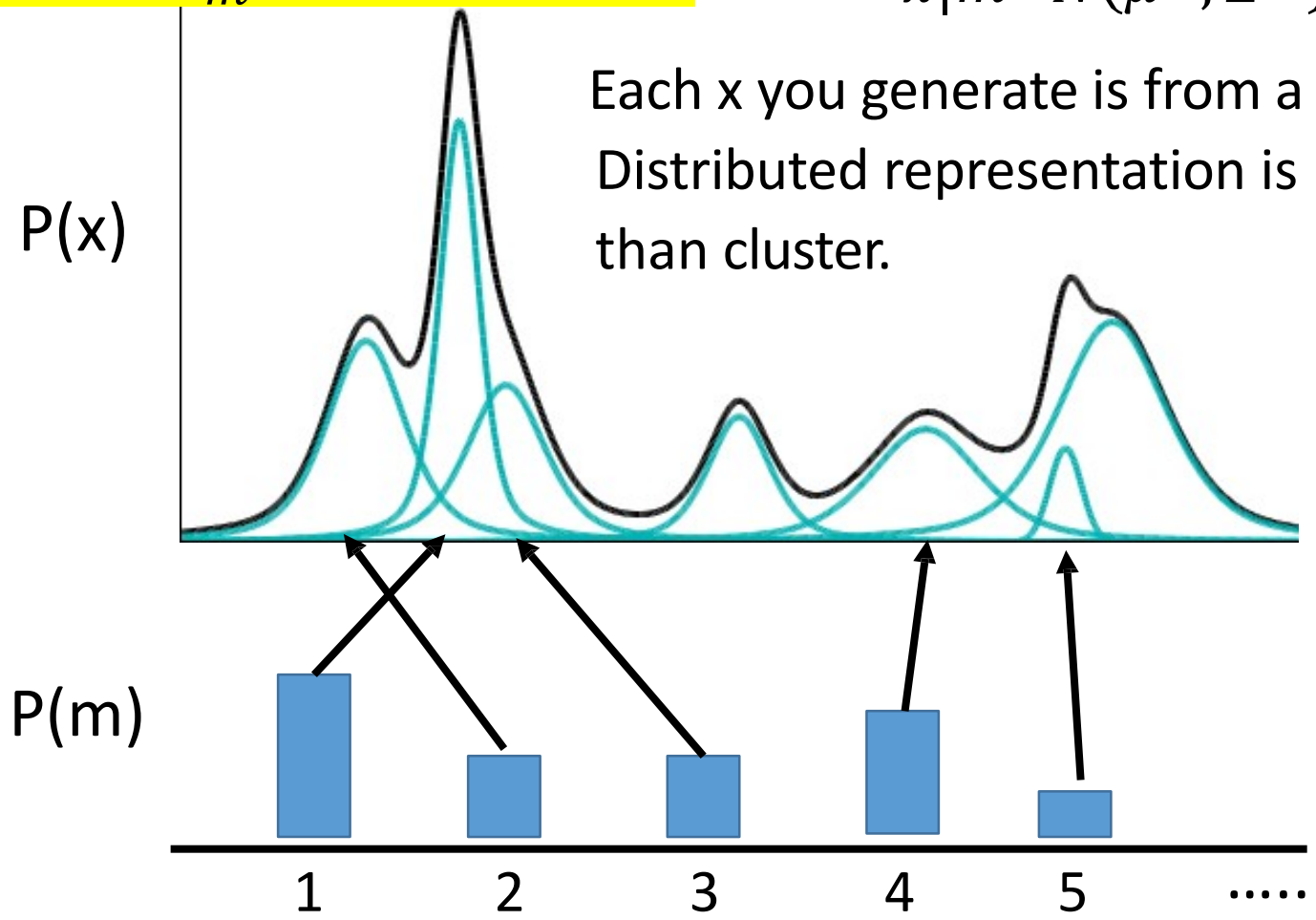
$m \sim P(m)$ (multinomial)

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.



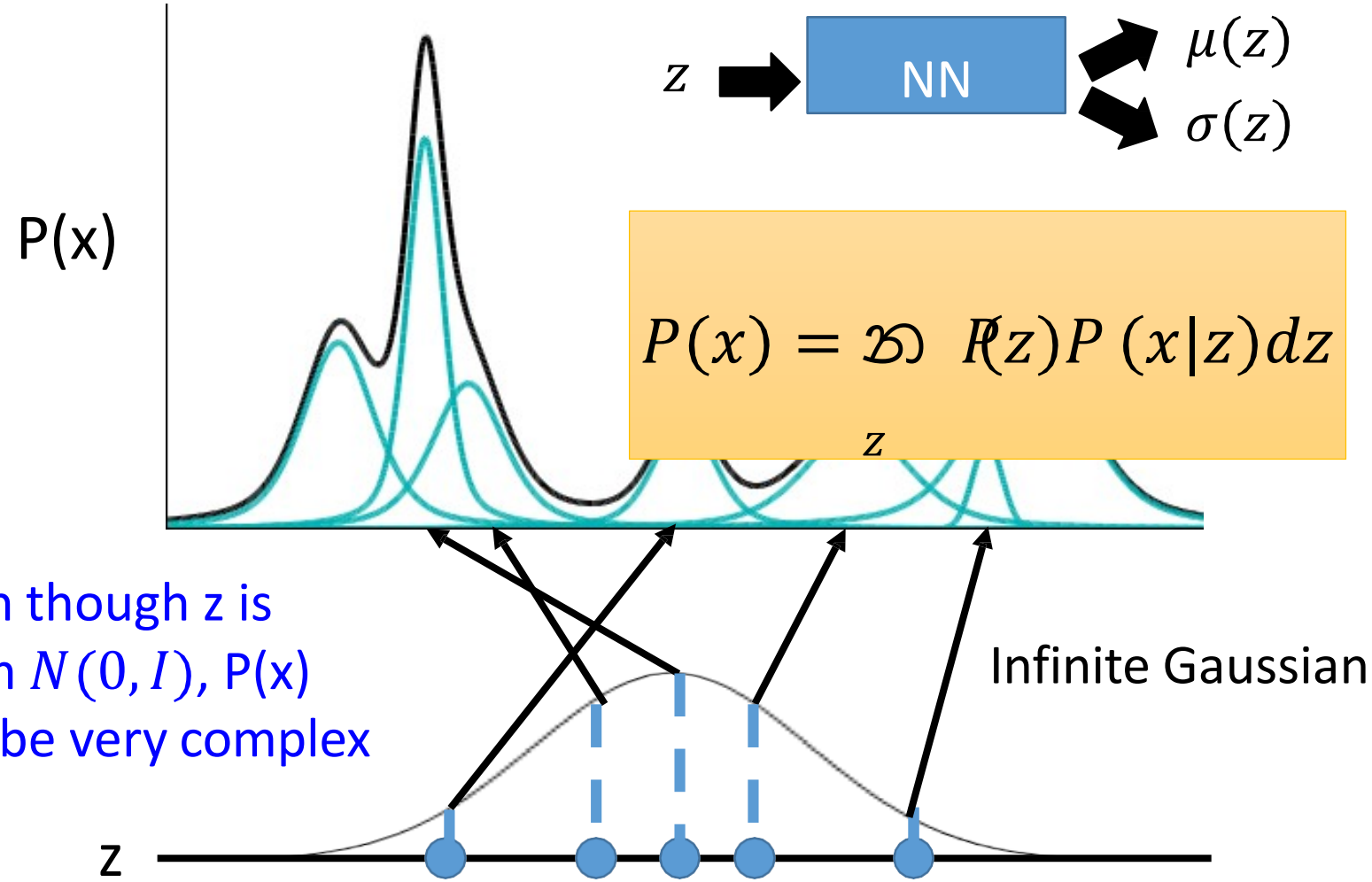
VAE

$z \sim N(0, I)$

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

z is a vector from normal distribution

Each dimension of z represents an attribute



Maximizing Likelihood

P(z) is normal distribution

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

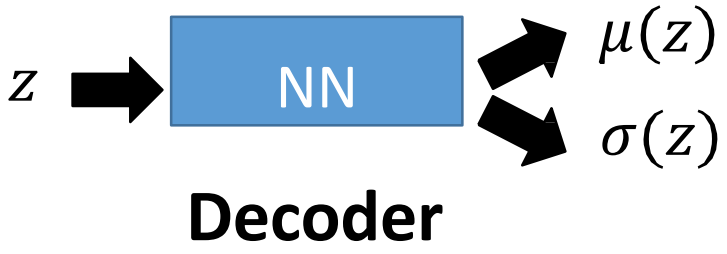
$$x|z \sim N\big(\mu(z), \sigma(z)\big)$$

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

$$L = \sum_x \log P(x)$$

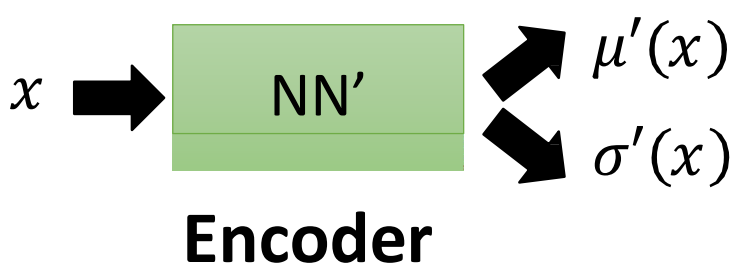
Maximizing the likelihood of the observed x

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

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

$\mu(z), \sigma(z)$ 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

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

q(z|x) can be any distribution

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

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

≥ 0

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

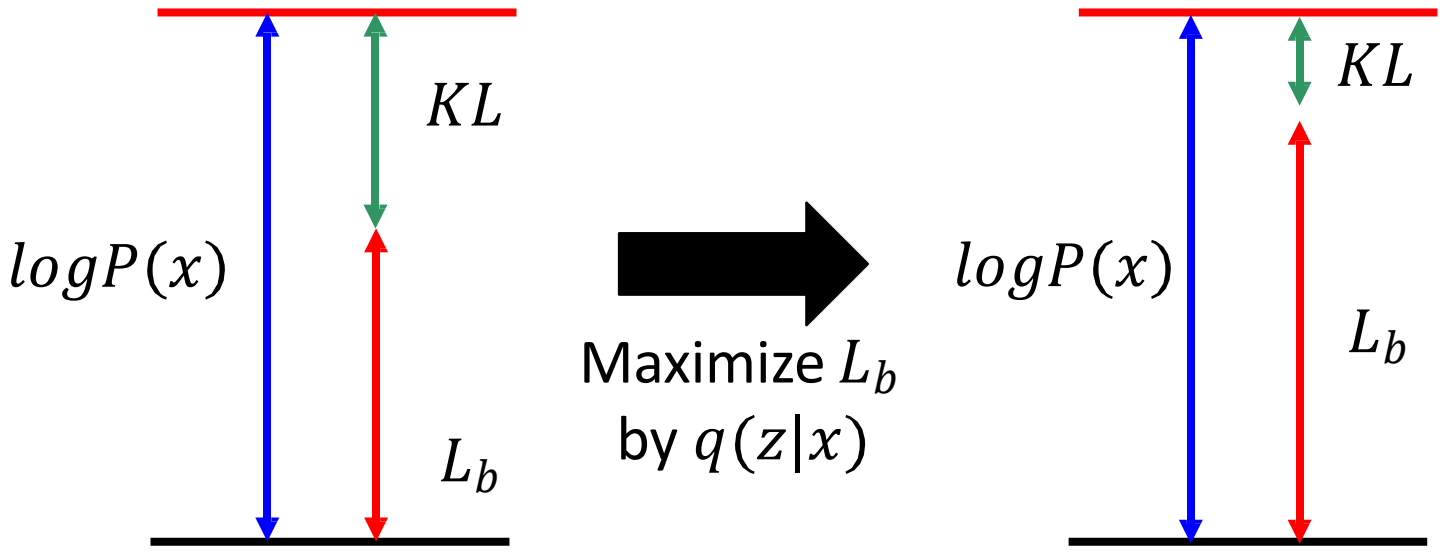
$lower\ bound\ L_b$

Maximizing Likelihood

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

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

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(z) is normal distribution

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

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

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

$$L = \sum_x \log P(x)$$

Maximizing the likelihood of the observed x

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

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

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

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

Connection with Network

Minimize

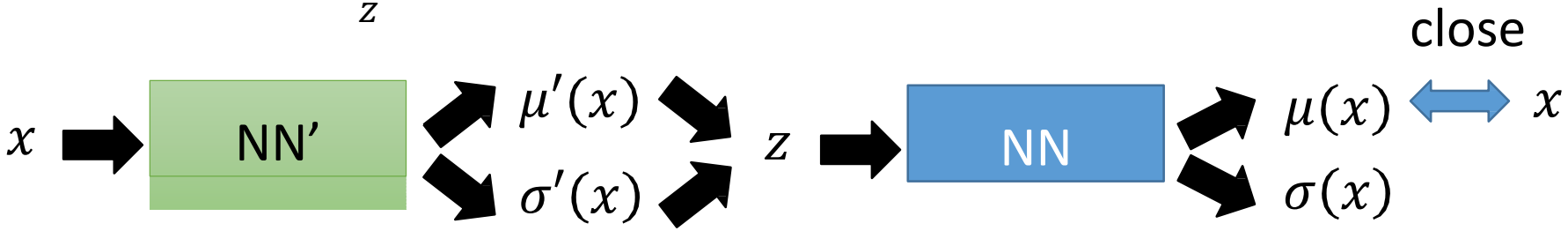
$$\sum_{i=1}^{3} \left(exp(\sigma_i) - (1 + \sigma_i) + (m_i)^2 \right)$$

Minimizing $KL\left(q(z|x)||P(z)\right)$

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

Maximizing

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



This is the auto-encoder

Conditional VAE

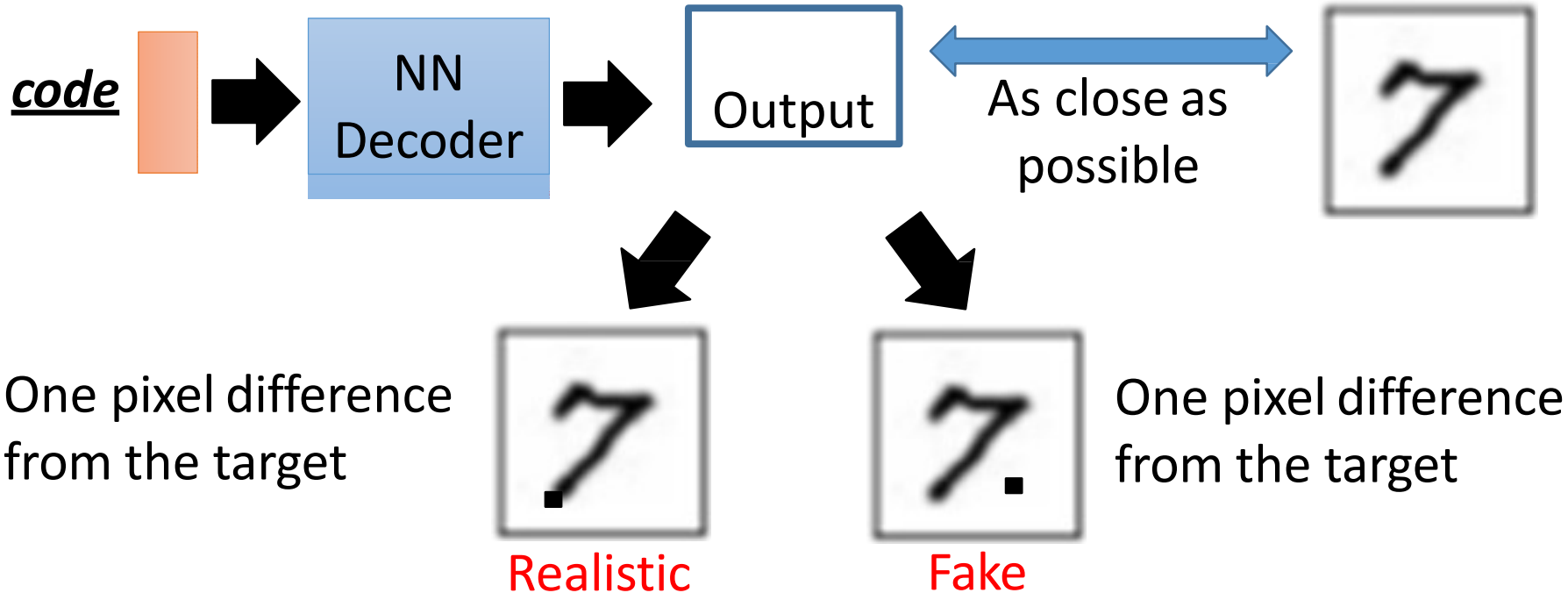
https://arxiv.org/pdf/1406.5298v2.pdf

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