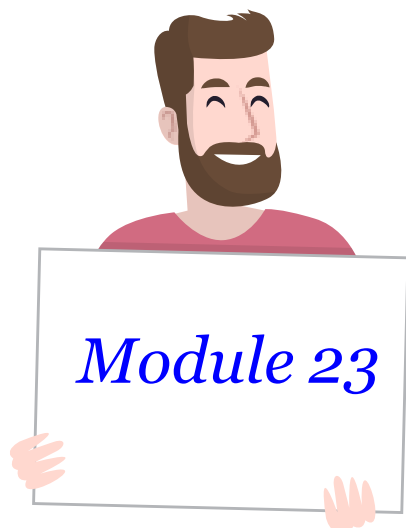




Autoencoder介紹



designed by  freepik

Estimated time:
45 min.

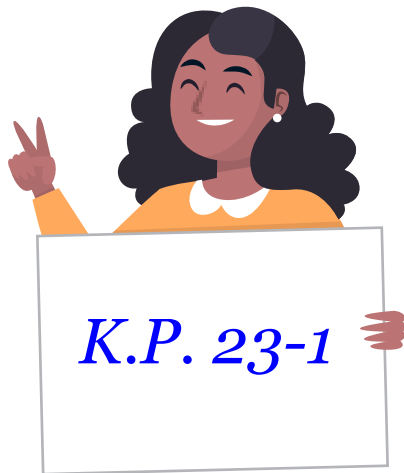
學習目標

- 23-1: Autoencoder介紹
- 23-2: Sparse/Denoise autoencoder
- 23-3: VAE



23-1: Autoencoder 介紹

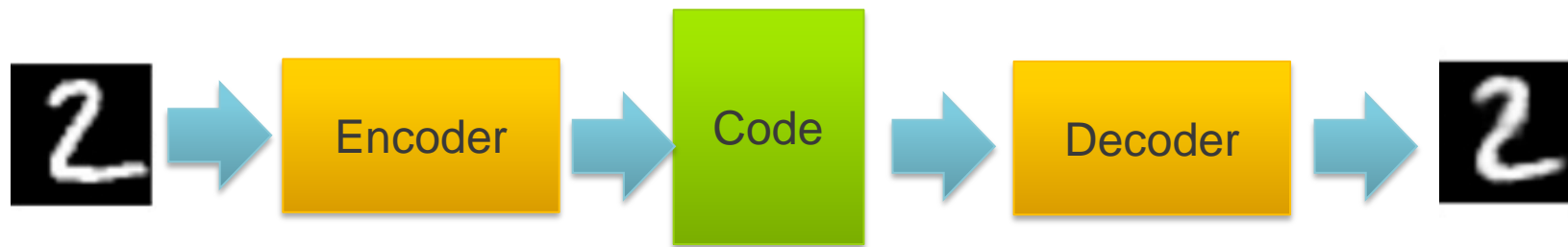
- 什麼是Autoencoder
- Autoencoder原理



designed by freepik

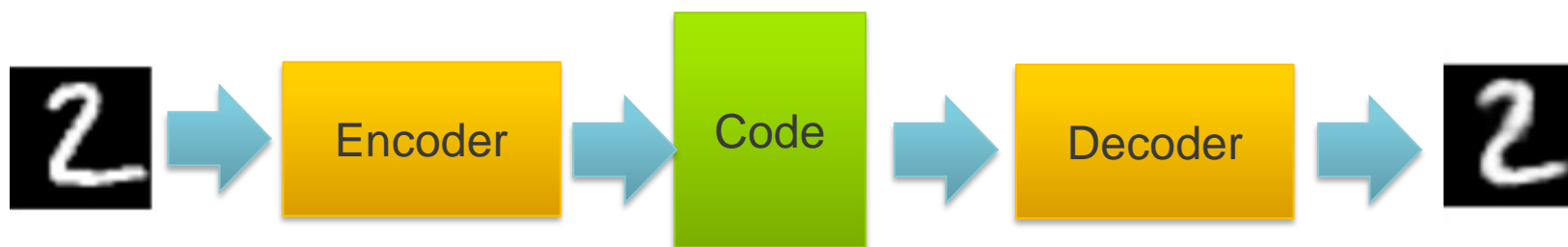
什麼是Autoencoder

- **Autoencoder是一種神經網路**
 - 其目標是要學習出原始資料更好的表示方法
 - 可以把它當作降維演算法的一種
- **Autoencoder可以把原始資料編碼成壓縮的形式，也能將壓縮的形式還原回來**



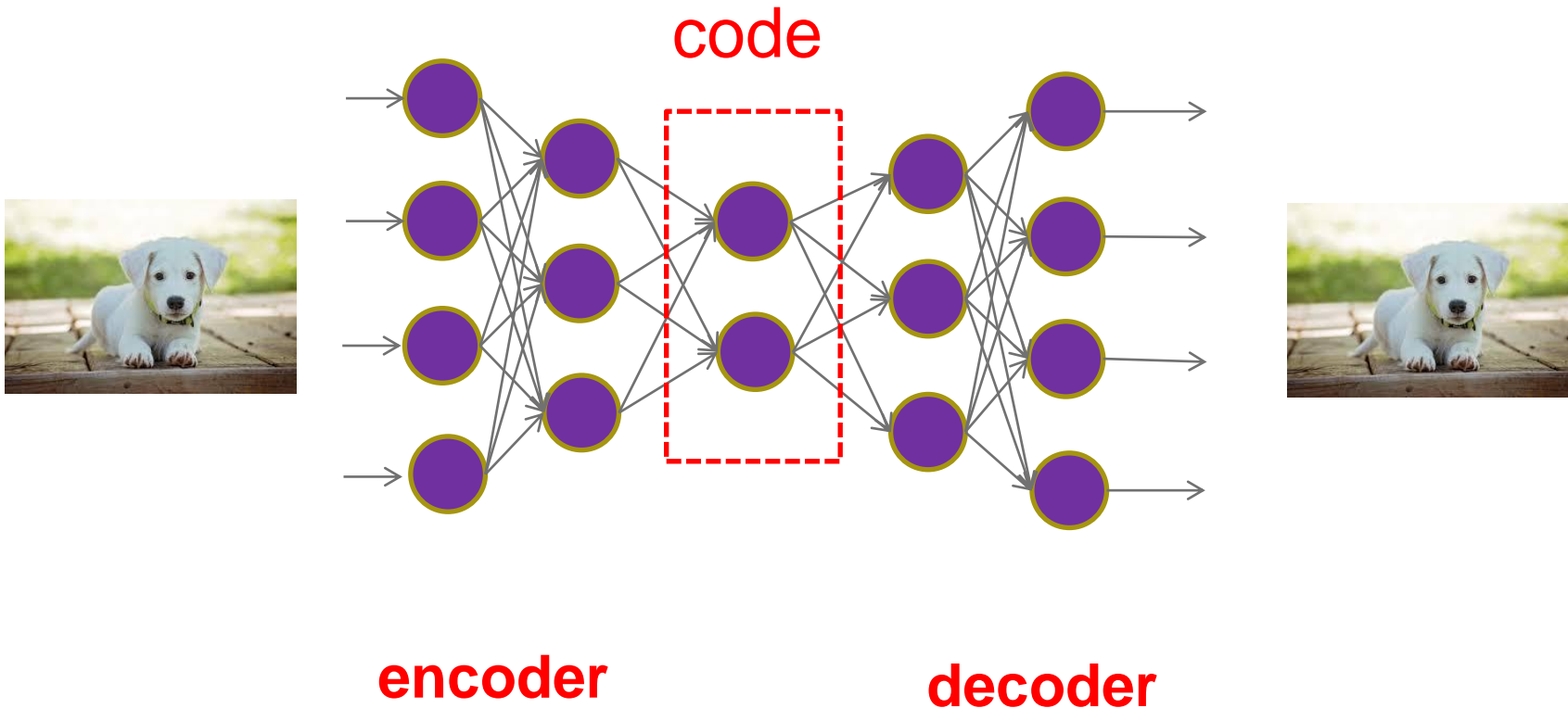
什麼是Autoencoder

- 一個Autoencoder通常包含了encoder以及decoder
 - encoder是用來將原始資料做壓縮
 - decoder是用來將壓縮資料還原



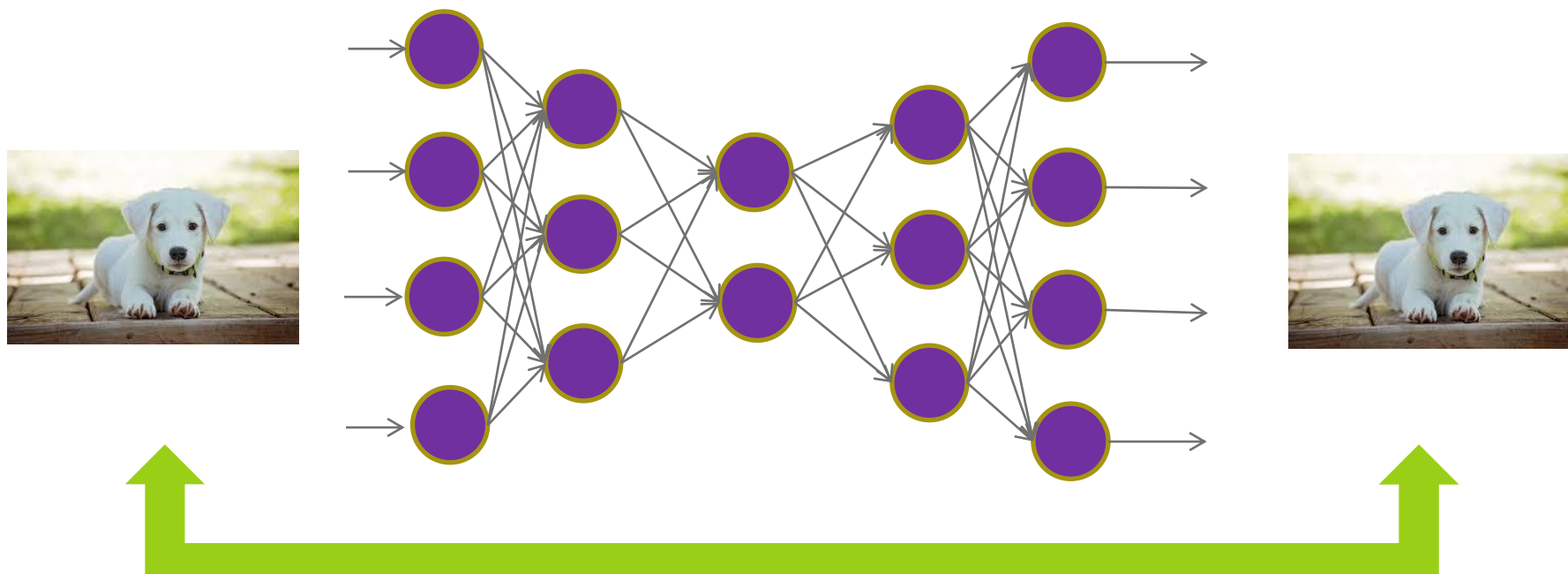
Autoencoder 原理

- 假設我們要將一張狗的照片做編碼，我們可以將狗照片丟入一個 autoencoder
 - 此 autoencoder 的 encoder 及 decoder 我們假設為 DNN 神經網路



Autoencoder 原理

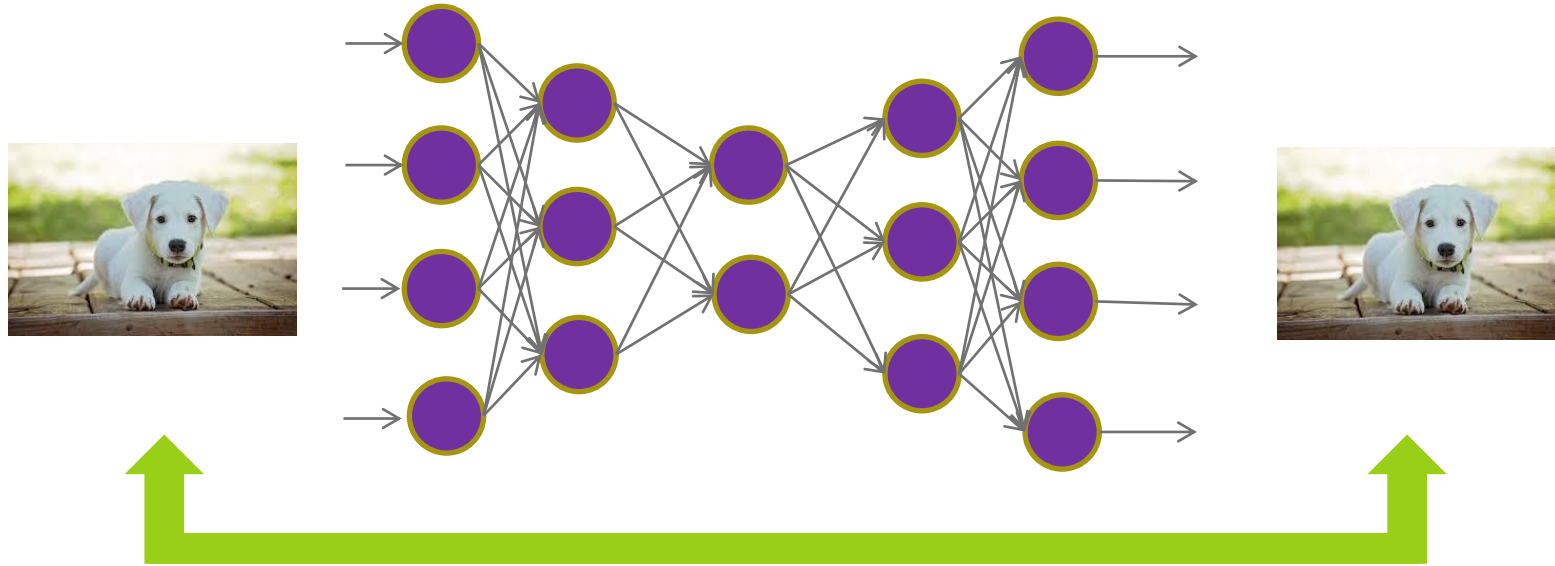
- 我們會期望此autoencoder的輸出會與原始照片一樣



期望輸出與原始照片一樣

Autoencoder原理

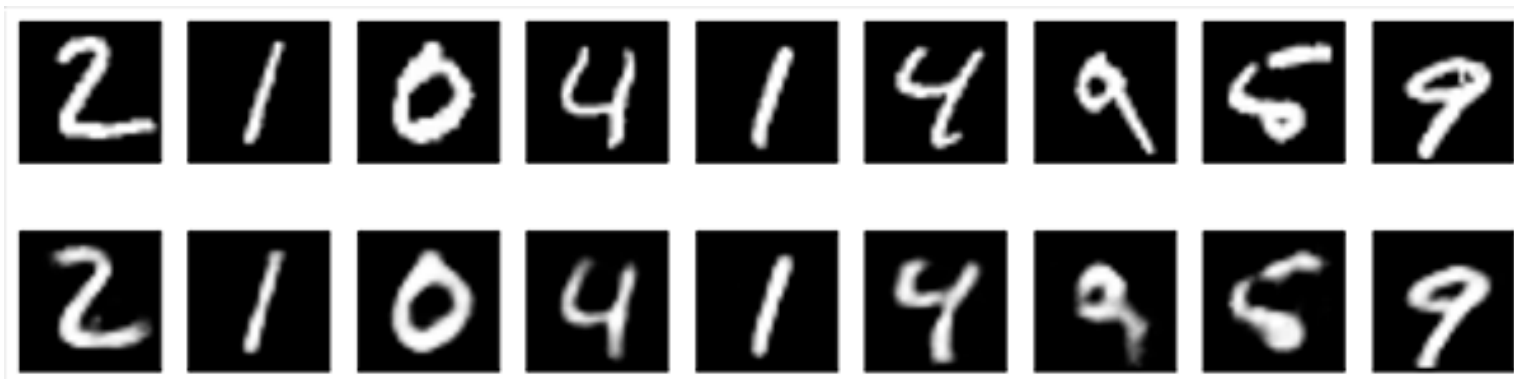
- 我們可以定義一個損失函數去衡量輸入與輸出照片的差值
 - 一般來說我們喜歡用**MSE**來當損失函數
 - 在**autoencoder**裡面，損失函數常常也被稱為**reconstruction error**



$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2 \quad (\text{reconstruction error})$$

Autoencoder 原理

- 當我們訓練完一個autoencoder後
 - 我們可以把一張照片輸入並觀察它回復照片的結果(又叫做生成照片)

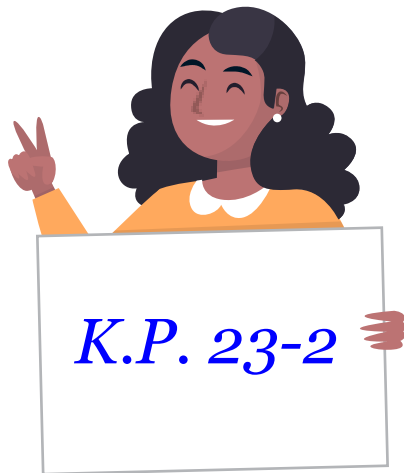


原始照片

生成照片

23-2: Sparse/Denoise autoencoder

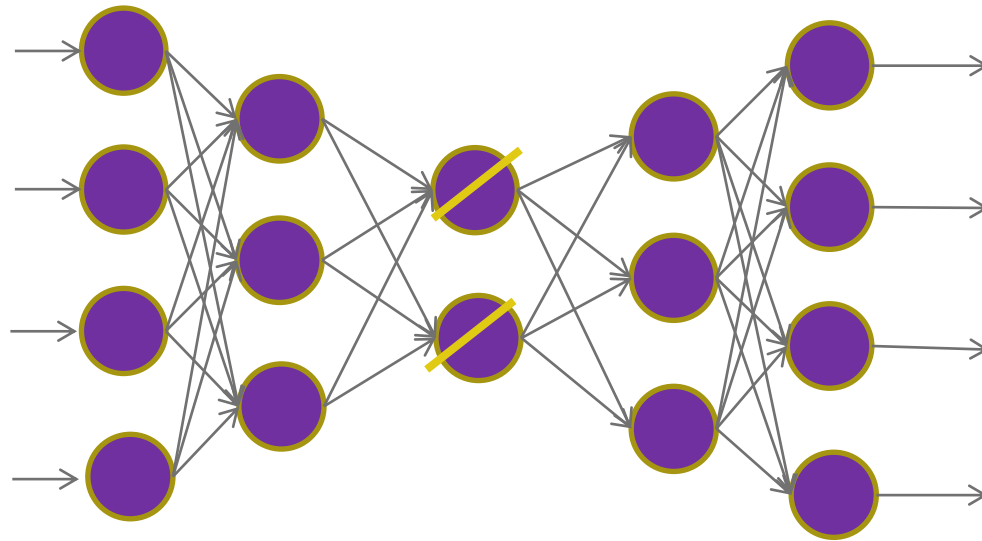
- Sparse autoencoder
- Denoise autoencoder



designed by freepik

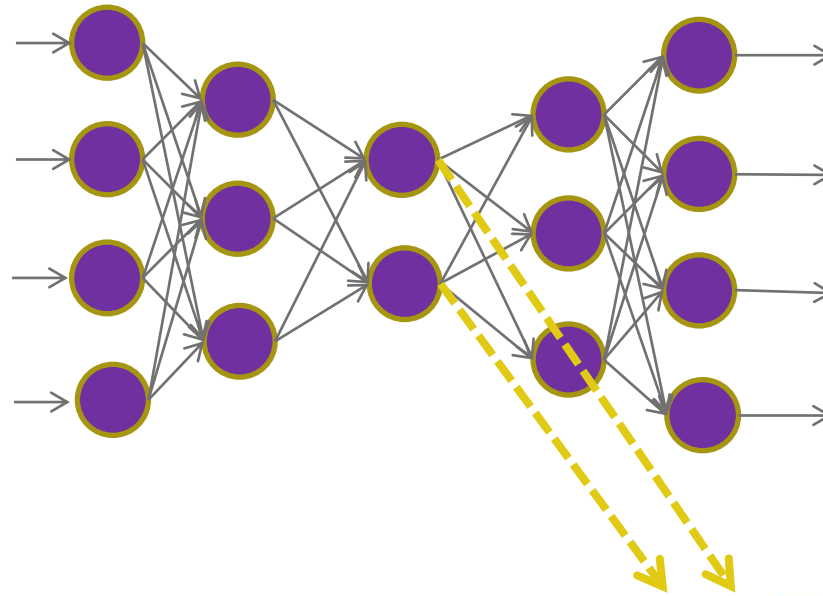
Sparse autoencoder

- 一般Autoencoder有個問題是，每次當輸入資料時，幾乎所有隱藏層的神經元都是處於激活狀態
 - 我們會希望針對不同資料，激活部分的神經元就好，因為我們會希望對於每個隱藏層的神經元來說，它們學習到的東西差異化越大越好，而非幾乎所有情況大家都被激活



Sparse autoencoder

- Sparse autoencoder使用sigmoid為隱藏層的激活函數
- Sparse autoencoder定義了稀疏參數
 - 此參數使用者可以自己設定，並控制隱藏層同時激活神經元的比例



$\hat{\rho}_j = \rho$ sparsity parameter

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^m [a_j^{(2)}(x^{(i)})]$$

Sparse autoencoder

- Sparse autoencoder最終的損失函數如下
 - 其將原始的損失函數加上一個稀疏參數的KL divergence

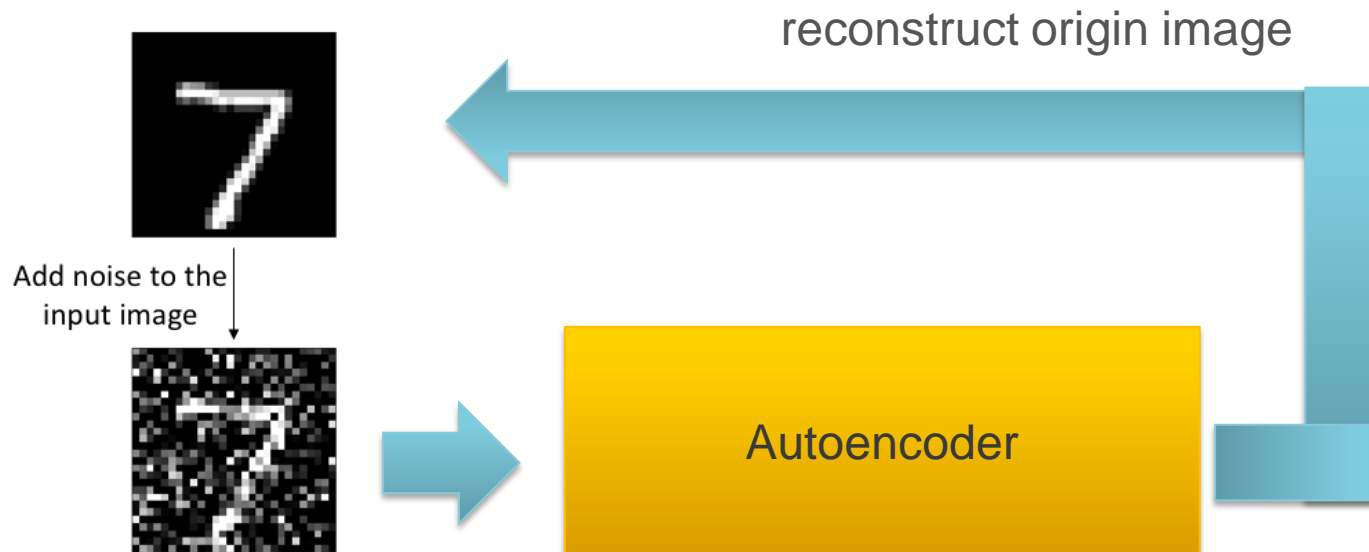
$$J_{\text{sparse}}(W, b) = J(W, b) + \beta \sum_{j=1}^{s_2} \text{KL}(\rho || \hat{\rho}_j)$$



add sparsity constraint

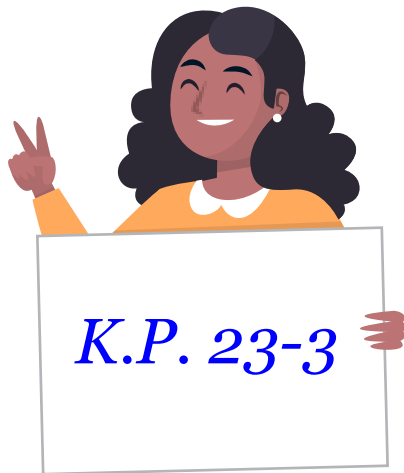
Denoise autoencoder

- Denoise autoencoder的觀念是為了避免autoencoder學習過後變成一個identity matrix(這樣沒有意義)
 - 因此其在輸入資料前，會先將資料加上noise在輸入autoencoder



23-3: VAE

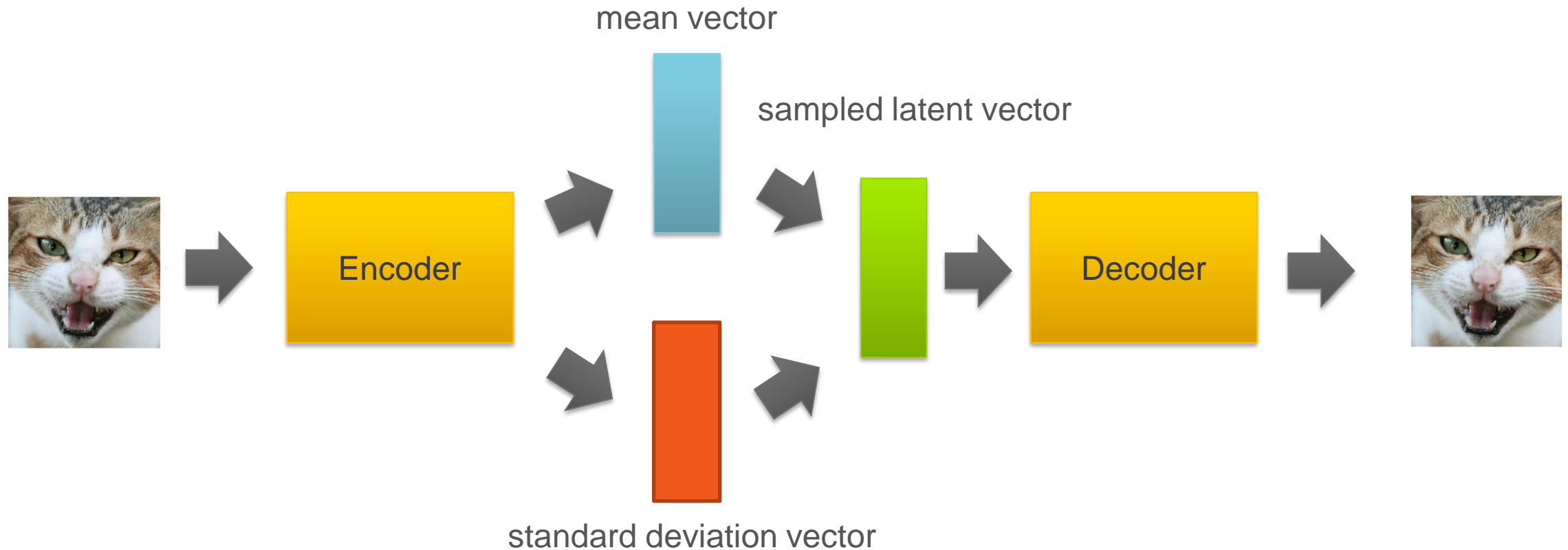
- VAE介紹
- VAE原理



designed by freepik

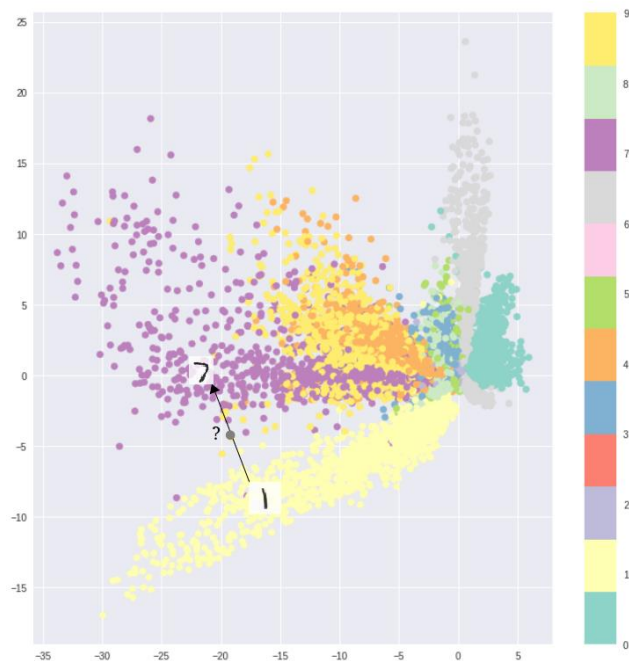
VAE介紹

- VAE是一個近期受歡迎的autoencoder
 - 它的latent space是連續的



VAE介紹

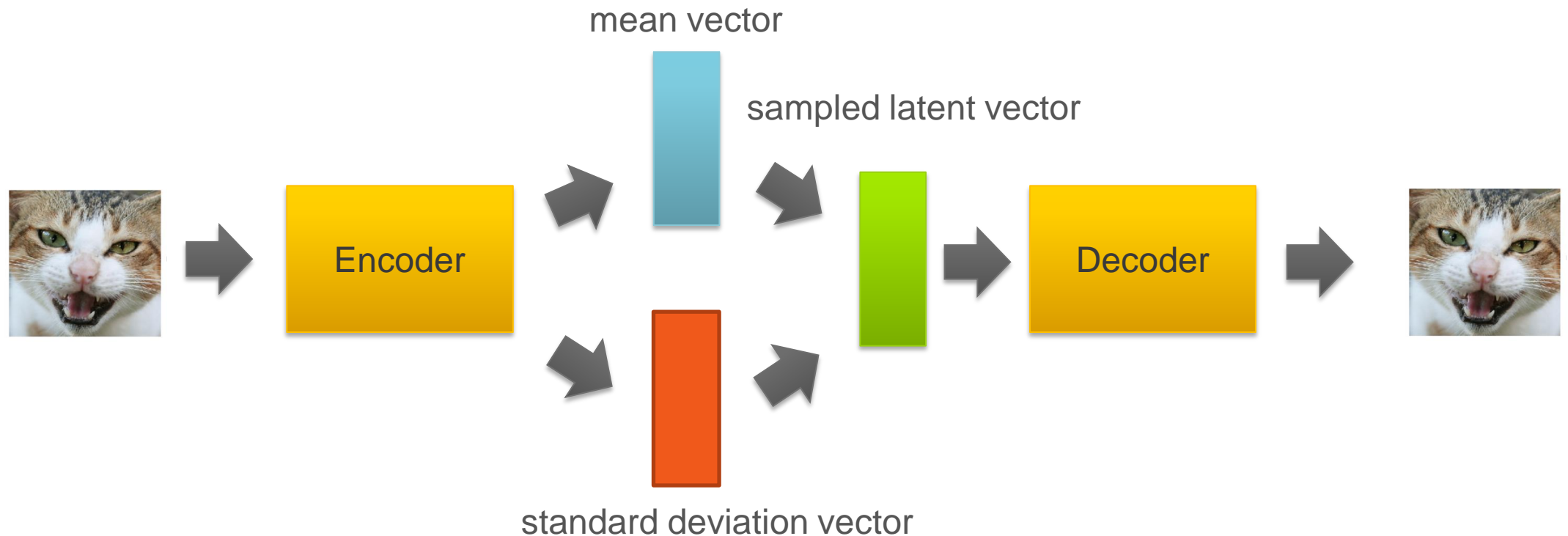
- **Latent space**指的是資料壓在一個低維度所形成空間
 - 可以發現大部分autoencoder的latent space會有間隙
 - 或者可以說有些資料點無法很明確看到資料之間如何慢慢演化



Autoencoder在MNIST上的結果(投影在2D座標上)

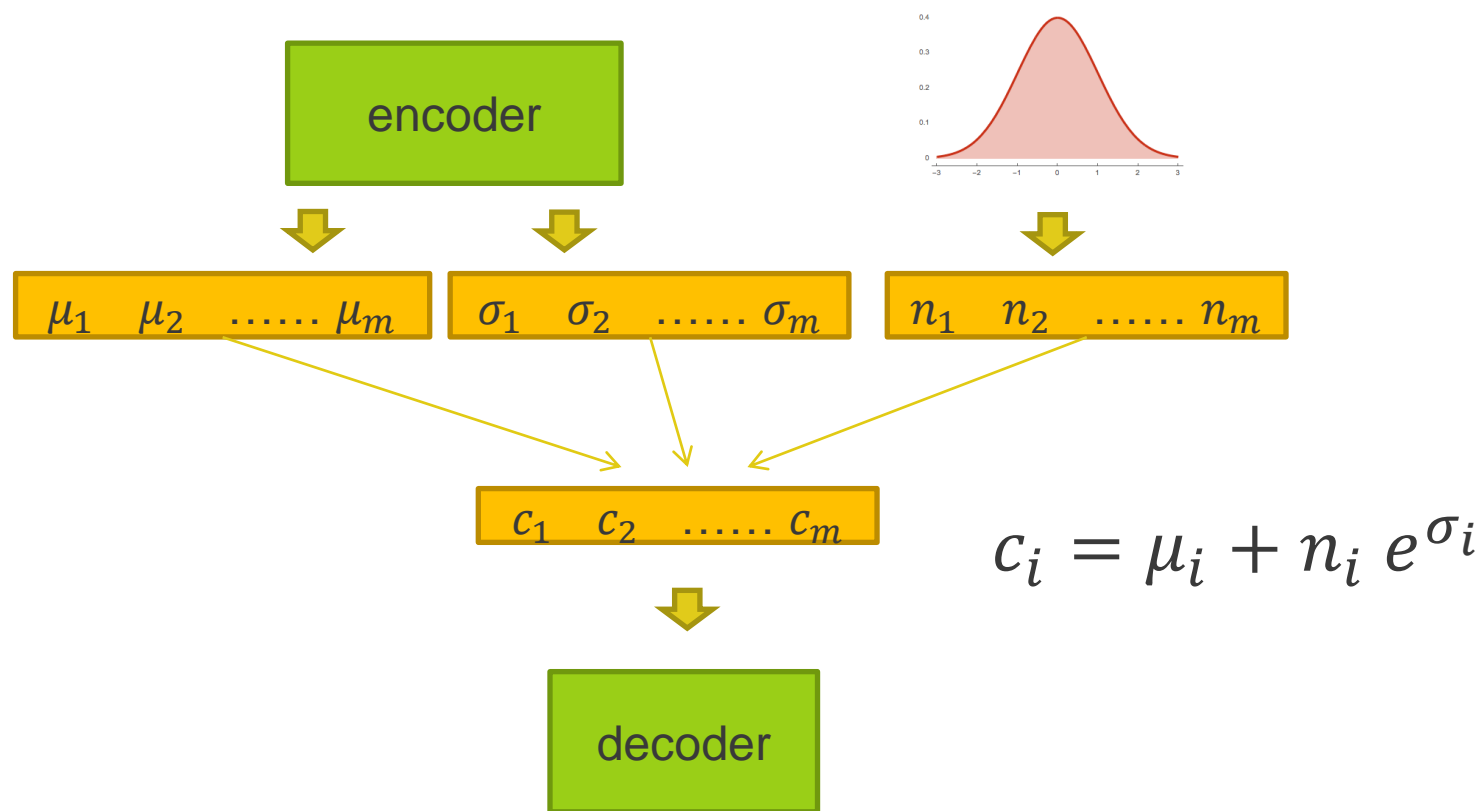
VAE原理

- VAE的encoder會先去預測平均向量以及標準差向量
 - 在藉由這兩個向量去產生latent vector
- 之後decoder會將latent vector解碼成原始照片



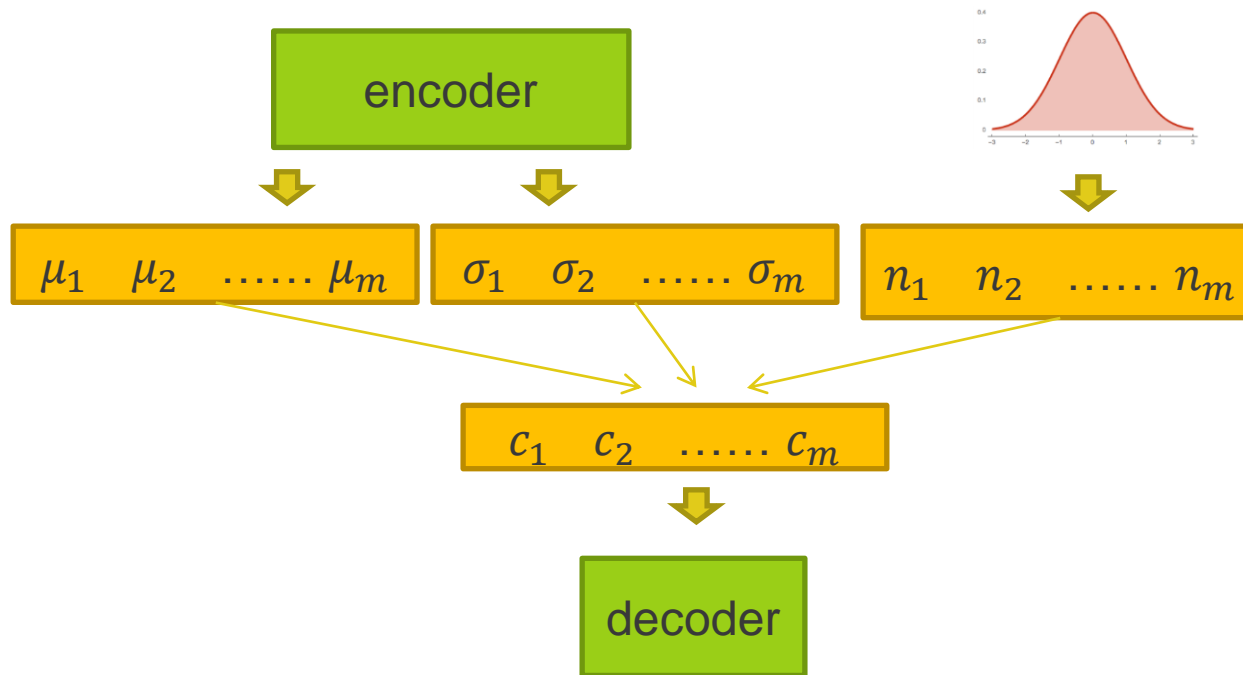
VAE原理

- 平均向量以及標準差向量如何計算成latent vector如下
 - 會使用一個常態分布去隨機產生一個noise向量並結合平均向量、標準差向量去計算latent vector



VAE原理

- VAE的損失函數除了原本的reconstruction error外，後面還增加了一項
 - 這一項是VAE paper根據推導而來的，有興趣的同學可以深入研究



Cost function



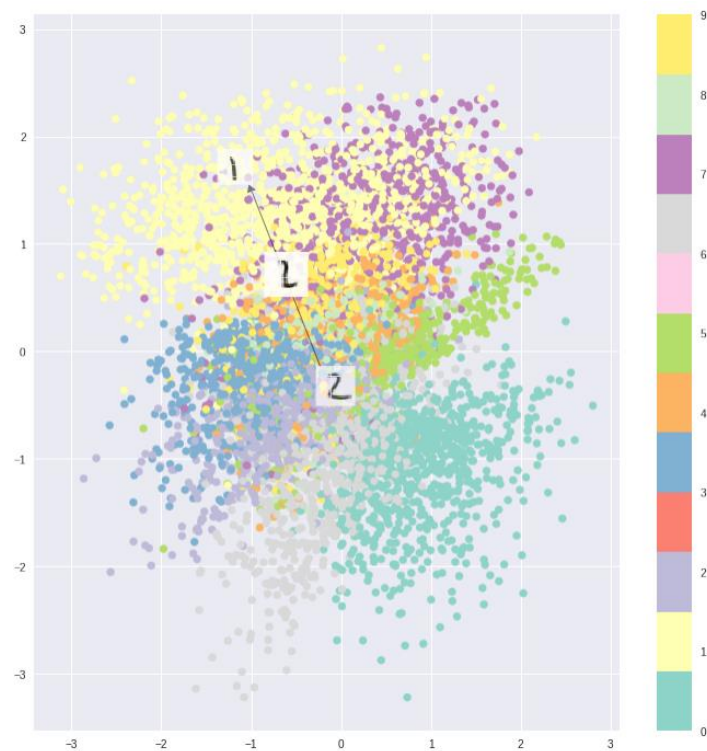
reconstruction
error



$$\sum_{i=1}^n \sigma_i^2 + \mu_i^2 - \log(\sigma_i) - 1$$

VAE原理

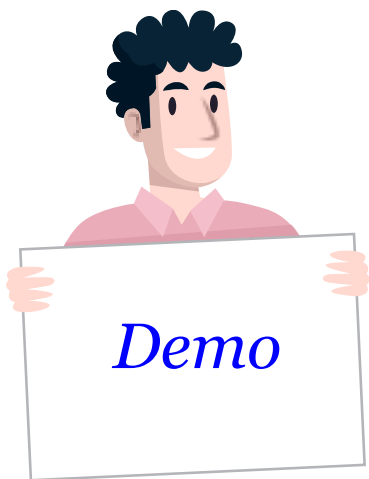
- 可以看到VAE所產生出來的結果不會有間隙



VAE result

Demo 23-3

- 將MNIST資料集輸入Autoencoder
- MNIST資料集Autoencoder可視化
- Convolutional Autoencoder



designed by freepik

線上Corelab

- 題目1：完成notMNIST Autoencoder網路訓練
 - 給予notMNIST資料集，完整整個Autoencoder網路部分
- 題目2：完成notMNIST Autoencoder網路訓練
 - 給予notMNIST資料集，完整整個Autoencoder網路以及訓練部分
- 題目3：完成Cifar10 Convolution Autoencoder
 - 給予Cifar10資料集，完整整個Convolution Autoencoder網路以及訓練部分

本章重點精華回顧

- Autoencoder介紹
- Sparse/Denoise autoencoder
- VAE



Lab: Autoencoder 實作

- **Lab01: 將MNIST資料集輸入Autoencoder**
- **Lab02: MNIST資料集Autoencoder可視化**
- **Lab03: Convolutional Autoencoder**

Estimated time:
20 minutes

