

Chapter15 Unsupervised Learning – Generation

1. Pixel RNN
2. Variational Auto-Encoder
 - a) Encoder's output: $c_i = \exp(\sigma_i) * e_i + m_i$
 - b) Minimize: Reconstruction Error & $\sum(\exp(\sigma_i) - (1 + \sigma_i) + (m_i)^2)$
 - i) m_i means original code
 - ii) c_i means code with noisy
 - iii) $(m_i)^2$ means L2 Regularization
 - c) VAE may just memorize the existing images, instead of generating new images
 - d) Conditional VAE
3. Generative Adversarial Network
 - a) A generator G is a network. The network defines a probability distribution
 - b) The loss of discriminator is related to JS divergence
 - c) Using gradient descent to update the parameters in g, but fix the d
 - d) GANs are difficult to optimize
 - e) No explicit signal about how good the generator is
 - f) When d fails, it doesn't guarantee that g generates realistic images, sometimes g finds a specific example to fail the d

Chapter16 Unsupervised Learning – Linear Dimension Reduction

1. Unsupervised Learning is all about reduction and generation
2. Clustering
 - a) K-means
 - b) Hierarchical Agglomerative Clustering HAC
 - i) Build a tree
 - ii) Pick a threshold
3. Distributed Representation
4. Principle Component Analysis
 - a) $z = Wx$
 - i) Project all the data points x onto W, and obtain a set of z
 - ii) We want the variance of z as large as possible
 - b) $\|w^1\|_2 = 1 \quad w^1 \cdot w^2 = 0$
 - c) $x - \bar{x} \approx c_1 u^1 + c_2 u^2 + \dots + c_K u^K = \hat{x}$
 - i) Reconstruction Error: $\|(x - \bar{x}) - \hat{x}\|_2$
 - ii) $\{w^1, w^2, \dots, w^K\}$ is the component $\{u^1, u^2, \dots, u^K\}$ minimizing L
 - iii)
$$L = \min \sum \|(x - \bar{x}) - (\sum_{k=1}^K c_k u^k)\|_2$$
 - d) PCA looks like neural network with one hidden layer (linear activation function)
 - e) Weakness of PCA
 - Can not handle non-linear dimension reduction
 - f) PCA involves adding up and subtracting some components
 - NMF: Non-negative Matrix Factorization
 - Forcing a and W to be non-negative