

Generative Models 1

- **Generation (sampling)**

sample $x_{new} \sim p(x)$

implicit model: VAE, GAN처럼 생성만 하는 모델

- **Density estimation (anomaly detection)**

$p(x)$ 높아야함

explicit model: 확률값까지도 얻어낼 수 있는 모델

- **Unsupervised representation learning (feature learning)**

이미지의 공통적 특징 학습

- **Basic Discrete Distributions**

- **Bernoulli distribution:** (biased) coin flip (1,0)

$$D = \{\text{Heads}, \text{Tails}\}$$

Specify $P(X = \text{Heads}) = p$, then $P(X = \text{tails}) = 1 - p$

Write: $X \sim \text{Ber}(p)$

- **Categorical distribution:** (biased) m-sided dice

$$D = \{1, \dots, m\}$$

Specify $P(Y = i) = p_i$, such that $\sum_{i=1}^m p_i = 1$

Write: $Y \sim \text{Cat}(p_1, \dots, p_m)$

- **Chain rule**

$$p(x_1, \dots, x_n) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2) \dots p(x_n|x_1, \dots, x_{n-1})$$

- **Bayes' rule**

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(y|x)p(x)}{p(y)}$$

- **Conditional independence**

If $x \perp y|z$, then $p(x|y, z) = p(x|z)$

- Markov assumption 적용 시, 획기적으로 감소

- **parameters**

- Chain rule 이용 시,

$$p(x_1, \dots, x_n) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2) \dots p(x_n|x_1, \dots, x_{n-1})$$

- $p(x_1)$: 1
- $p(x_2|x_1)$: 2 $\Rightarrow p(x_2|x_1 = 0), p(x_2|x_1 = 1)$
- $p(x_3|x_1, x_2)$: 4
- total: $1 + 2 + 2^2 + \dots + 2^{n-1} = 2^n - 1$

n개의 independent binary pixel 가진 이미지의 파라미터 수로도 해석 가능

- **Markov assumption** 적용 시, 획기적으로 감소

suppose $X_{i+1} \perp X_1, \dots, X_{i-1} | X_i$

- X_{i+1} (미래)의 조건부확률분포는 X_i (현재)가 주어졌을 시, 이전의 모든 과거 X_1, X_2, \dots, X_{i-1} 와 독립
 - $p(x_1, \dots, x_n) = p(x_1)p(x_2|x_1)p(x_3|x_2) \dots p(x_n|x_{n-1})$
 - total: $2n - 1$
- exponential reduction

• Auto-regressive Model

Markov chain 활용 => 이전의 n개(AR-n model) / 1개(AR-1 model) 고려

chain rule로 joint distribution 나눔

랜덤 변수들의 순서(order) 필요

- 이미지(2차원)의 순서(1차원)을 어떻게 매기는가 => 성능 달라짐
- 어떤 식으로 conditional independence 주는가 => 전체 모델의 structure 달라짐

• NADE(Neural Autoregressive Density Estimator)

i 번째 픽셀을 $1, 2, \dots, i$ 번째 픽셀까지에 dependent

매번 뉴럴넷 층으로 보냄 => 층의 사이즈 점차 커짐

explicit model로, generation뿐만 아니라 확률(density) 계산 가능

- joint dist 이용해 $p(x_i|x_{1:i-1})$ 계산 가능

continuous variables 모델링 경우, 마지막 layer에 gaussian mixture model 활용, 연속 dist 생성 가능

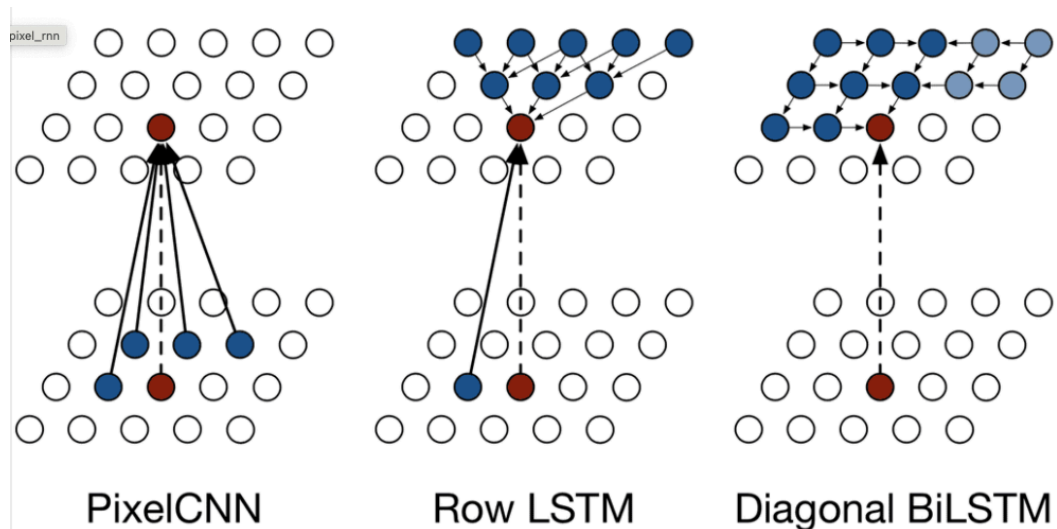
• Pixel RNN

use RNNs to define an auto-regressive model

$$p(x) = \prod_{i=1}^{n^2} (\text{prob } i\text{번째 R}) (\text{prob } i\text{번째 G}) (\text{prob } i\text{번째 B})$$

$$p(x) = \prod_{i=1}^{n^2} p(x_{i,R}|x_{<i})p(x_{i,G}|x_{<i}, x_{i,R})p(x_{i,B}|x_{<i}, x_{i,R}, x_{i,G})$$

- ordering 따른 종류



- Row LSTM
- Diagonal BiLSTM