

* 기계학습 Information Theory & Clustering (Page 2) 2016104109 김민석 김성우

• Self Information : 사건 x_i 의 정보량

$$I(x_i) = \log_b \frac{1}{P(x_i)} = -\log_b P(x_i).$$

• Entropy : 평균 정보량

$$H(X) = E[I(x_i)] = \sum_{i=1}^m P(x_i) I(x_i) \\ = - \sum_{i=1}^m P(x_i) \log_2 P(x_i) \text{ b/symbol}$$

• Property

$$0 \leq H(X) \leq \log_2 m \quad m: \text{the number of symbols of the source } X$$

• Cross Entropy

$$H(p, q) = E_p[I_q(X)] = E_p[-\log(q(X))] \\ \text{true guess} \\ = - \sum_{i=1}^m p(x_i) \log(q(x_i))$$

• KL Divergence

$$D_{KL}(P||Q) = \sum_{i=1}^m p(x_i) \log\left(\frac{P(x_i)}{Q(x_i)}\right) = H(P, Q) - H(P, P) \geq 0$$

* ML Estimation

• Likelihood of θ

θ 가 정답이라 가정했을 때, 샘플 x 의 발생 확률

$$\hat{\theta} = \arg\max_{\theta} P(X|\theta)$$

$$L(\theta) = \log P(X|\theta) = \sum_{i=1}^n \log P(x_i|\theta)$$

$$\hat{\theta} = \arg\max_{\theta} L(\theta)$$

* MAP Estimation

• Prior & Posterior Probability of θ

• 사전 확률: θ 가 정답일 확률

• 사후 확률: 샘플 x 가 발생했다고 할 때, θ 가 정답일 확률

$$\hat{\theta} = \arg\max_{\theta} P(\theta|X) = \arg\max_{\theta} \underbrace{P(X|\theta)}_{\text{likelihood}} \underbrace{P(\theta)}_{\text{Prior}}$$

$$\log P(X|\theta)P(\theta) = \log P(X|\theta) + \log P(\theta)$$

$$= \sum_{i=1}^n \log P(x_i|\theta) + \log P(\theta)$$

* 3- Tasks of Unsupervised Learning

① Clustering

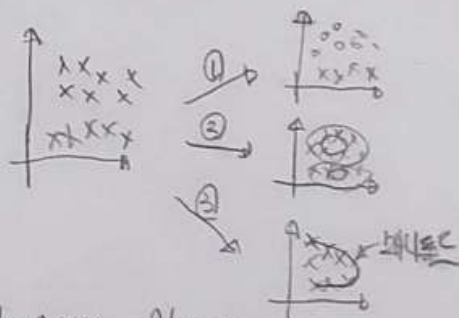
→ 유사 샘플을 모아 같은 그룹으로 묶는 일

② Density Estimation

→ 데이터로부터 확률 분포를 추정하는 일

③ Feature Space Conversion

→ 원래 특징을 저차원 or 고차원으로 변환하는 일



* k-means Algorithm

1) k개의 중심 중심을 초기화

2) 지를 가장 가까운 중심 중심에 매칭 ($i \rightarrow 1 \sim n$)

3) k개 매칭된 샘플의 평균의 위치를 대체 ($j \rightarrow 1 \sim k$)