Introduction of Unsupervised Learning
Part I

Namwoo Kang

Smart Design Lab
CCS Graduate School of Green Transportation
KAIST

Reference

❑ 강의 슬라이드 및 실습코드는 아래의 링크에서 받으실 수 있습니다

- http://www.smartdesignlab.org/dl_aischool_2021.html
- Contributors: 김성신, 유소영, 이성희, 김은지

❑ 강의 소스

- Andrew Ng의 ML Class (www.holehouse.org/mlclass/)
- Fei-Fei Li & Justin Johnson & Serena Yeung, CS231n: Convolutional Neural Networks for Visual Recognition, Stanford (http://cs231n.stanford.edu/)
- Stefano Ermon & Aditya Grover, CS 236: Deep Generative Models , Stanford (https://deepgenerativemodels.github.io/)
- 모두를 위한 딥러닝 (https://hunkim.github.io/ml/)
- 모두를 위한 딥러닝 시즌 2 (https://deeplearningzerotoall.github.io/season2/lec_tensorflow.html)
- 이활석, Autoencoders (https://www.slideshare.net/NaverEngineering/ss-96581209)
- 최윤제, 1시간만에 GAN(Generative Adversarial Network) 완전 정복하기 (https://www.slideshare.net/NaverEngineering/1-gangenerative-adversarial-network?qid=c53ce33f-6643-4437-8e93-88776c9cebb1&v=&b=&from_search=5)
- 김성범, [핵심 머신러닝] Principal Component Analysis (PCA, 주성분 분석) (https://youtu.be/FhQm2Tc8Kic)

Contents

Basic Probability

- **조건부 확률**

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

$$p(x,y) = p(y|x)p(x) = p(x|y)p(y)$$

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

- **베이즈 정리(Bayes Rule)**

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

- **전체 확률의 법칙 (Law of Total Probability)**

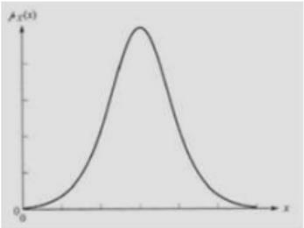
$$p(y) = \textstyle\sum_x p(y|x)p(x)$$

전체 확률은 조건부 확률의 합으로 표현할 수 있다.

- **Marginalization**

$$p(y) = \int_{-\infty}^{\infty} p(x,y)dx = \int_{-\infty}^{\infty} p(y|x)p(x)dx$$

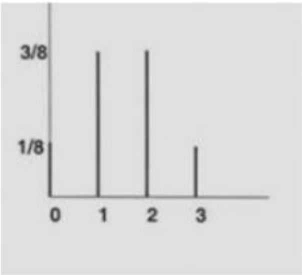
- **pdf** (probability density function)



$$\int_{-\infty}^{\infty} p(x)dx = 1$$

$$P(a < X < b) = \int_a^b p(t)dt$$

- **pmf** (probability mass function)



$$\sum_x p(x) = 1$$

Basic Probability

- **Gaussian Distribution (Normal Distribution)**

$$p(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

d차원의 경우:

$$p(X) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp[-\frac{1}{2}(X-\mu)^t \Sigma^{-1}(X-\mu)]$$

- **Bernoulli Distribution**

$$p(x;p) = \begin{cases} p & if\ x = 1, \\ 1-p & if\ x = 0 \end{cases}$$

$$p(x;p) = p^x(1-p)^{(1-x)}$$

- **Cross-entropy**

$$H(P,Q) = -\sum_x P(x)\log Q(x)$$

두 확률분포 P와 Q 사이의 차이를 계산.
P와 Q가 같을때 최소값

- **Kullback–Leibler divergence (KLD)**

$$KL(P||Q) = -\sum_x P(x)\log\frac{Q(x)}{P(x)} = \sum_x P(x)\log\frac{P(x)}{Q(x)}$$

- **Jenson-Shannon divergence (JSD)**

$$JSD(P||Q) = \frac{1}{2}KL(P||M) + \frac{1}{2}KL(Q||M)$$

where $M = \frac{1}{2}(P+Q)$

Basic Probability

- **Expectation**

x의 확률

이산 랜덤변수 기대값: $$E[X] = \sum_{\forall k} x_k p_X(x_k)$$

연속 랜덤변수 기대값: $$E[X] = \int_{-\infty}^{\infty} x\, p_X(x) dx$$

이산 랜덤변수 함수 기대값: $$E[f(X)] = \sum_{\forall k} f(x_k) p_X(x_k)$$

연속 랜덤변수 함수 기대값: $$E[f(X)] = \int_{-\infty}^{\infty} f(x) p_X(x)\, dx$$

$$E_{x \sim \boxed{p(x)}}[\boxed{f(X)}] = \int_x \boxed{f(x)}\boxed{p(x)} dx$$

x의 확률분포

ex) 주사위의 기대값 

$$1\times\frac{1}{6} + 2\times\frac{1}{6} + 3\times\frac{1}{6} + 4\times\frac{1}{6} + 5\times\frac{1}{6} + 6\times\frac{1}{6} = 3.5$$

Basic Probability

Notation for Supervised Learning

y: data x: data θ: Learnable parameter

$$\max_{\theta} P(Y|X;\theta)$$

; parameterized by

$*$ optimal

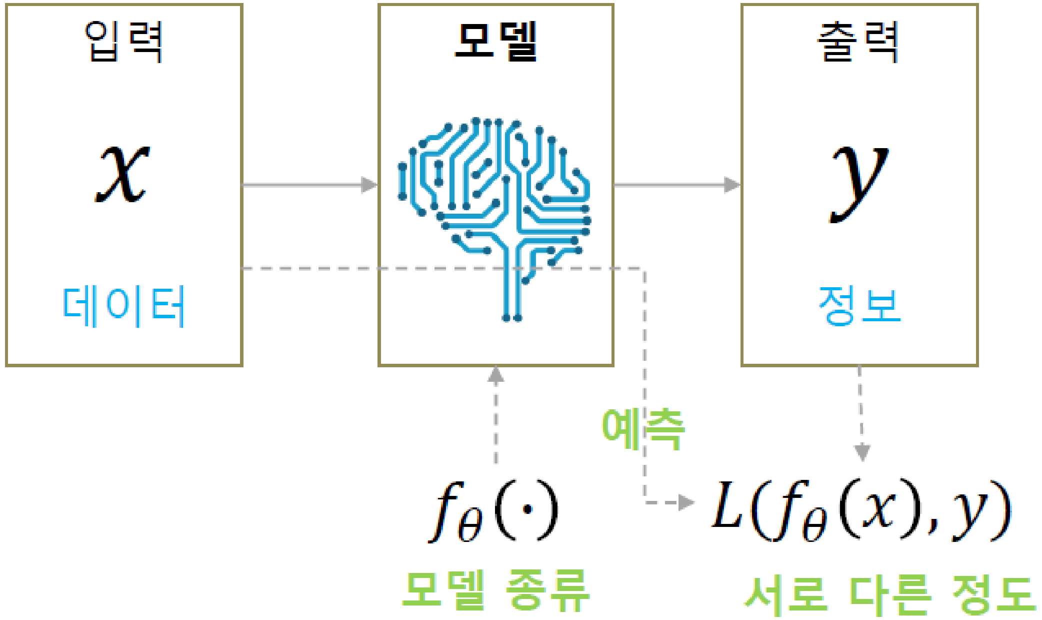
$$\theta^* = arg\max_{\theta} P(Y|X;\theta)$$

get θ when P is maximum

SMART DESIGN LAB KAIST

Maximum Likelihood

❑Classic Machine Learning

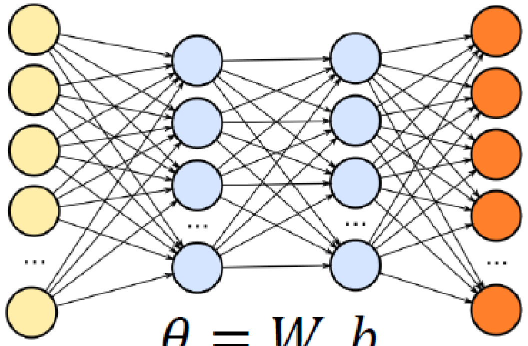


$$\theta^* = arg\min_\theta L(f_\theta(x), y)$$ 주어진 데이터를 제일 잘 설명하는 모델 찾기

$$y_{new} = f_\theta\ (x_{new})$$ 고정 입력, 고정 출력

❑Deep Neural Networks

$f_\theta(\cdot)$ Deep Neural Network



$\theta = W, b$

파라미터는 웨이트와 바이어스

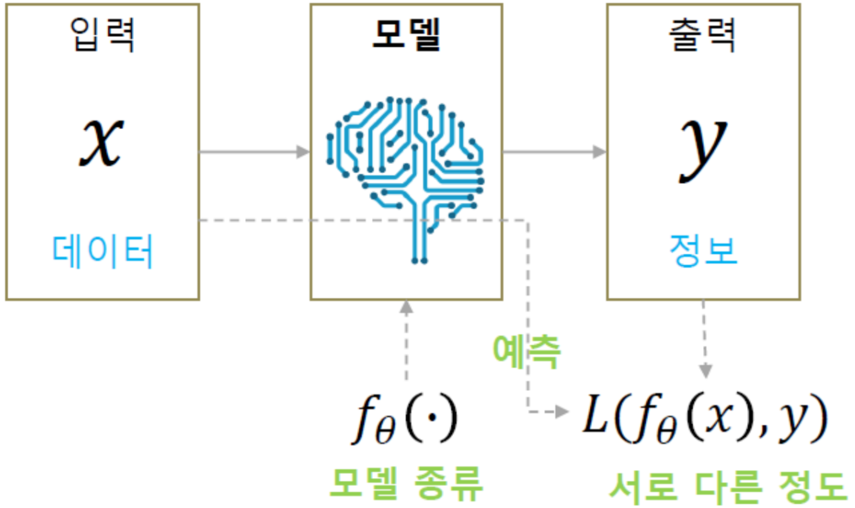
$$L(f_\theta(x), y) = \sum_i L(f_\theta(x_i), y_i)$$

Backpropagation을 통해 DNN을 학습시키기 위한 조건

1. Total loss of DNN over training samples is the sum of loss for each training sample
2. Loss for each training example is a function of final output of DNN

Maximum Likelihood

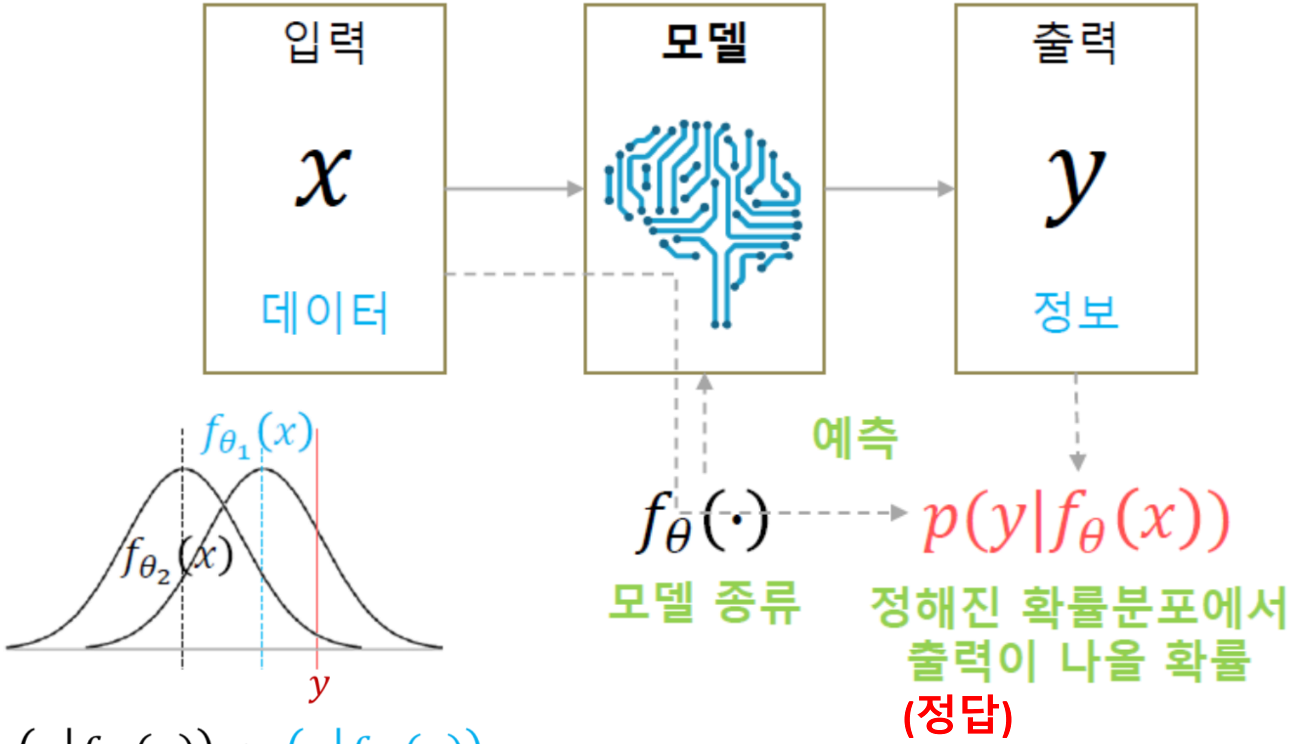
❑Classic Machine Learning



$\theta^* = arg\min_\theta L(f_\theta(x), y)$ 주어진 데이터를 제일 잘 설명하는 모델 찾기

$y_{new} = f_\theta\ (x_{new})$ 고정 입력, 고정 출력

❑Maximum Likelihood 관점



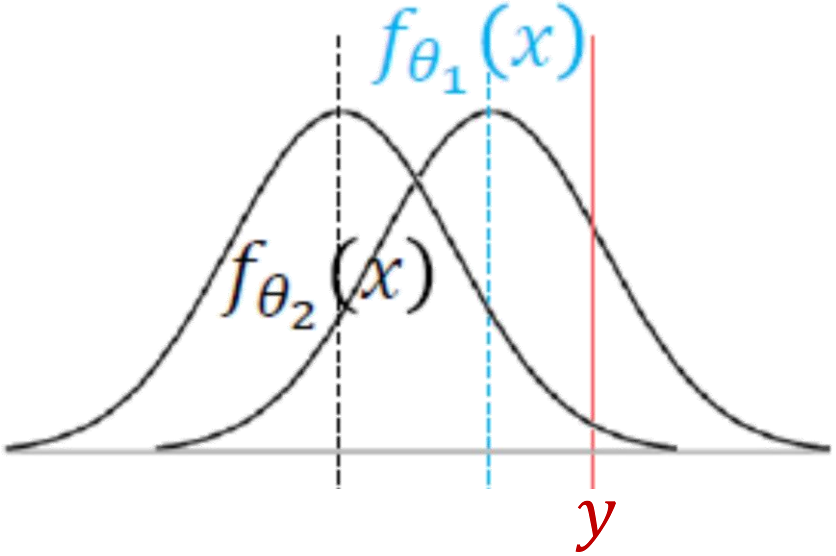
$p(y|f_{\theta_2}(x)) < p(y|f_{\theta_1}(x))$

$\theta^* = arg\min_\theta[-\log(p(y|f_\theta(x)))]$ 주어진 데이터를 제일 잘 설명하는 모델 찾기

Negative log-likelihood

$y_{new} \sim p(y|f_\theta(x_{new}))$ 고정 입력, 다른 출력

❑ Maximum Likelihood 관점



$$\max_{\theta} p(y|f_{\theta}(x))$$

정해진 확률분포에서 출력이 나올 확률을 최대화

$$p(y|f_{\theta_2}(x)) < p(y|f_{\theta_1}(x))$$

Maximum Likelihood

❑ Negative log-likelihood

independent and identically distributed

i.i.d Condition on

- Assumption 1: Independent

All of our data is independent of each other

$$p(y|f_\theta(x)) = \prod_i p_{D_i}(y|f_\theta(x_i))$$

- Assumption 2: Identically Distributed

Our data is identically distributed

$$p(y|f_\theta(x)) = \prod_i p(y|f_\theta(x_i))$$

Negative log-likelihood ***곱이 합이 될수 있음!***

$$-log\left(p(y|f_\theta(x))\right) = -\sum_i log(p(y_i|f_\theta(x_i)))$$

Backpropagation을 통해 DNN을 학습시키기 위한 조건 만족시킴

1. Total loss of DNN over training samples is the sum of loss for each training sample

2. Loss for each training example is a function of final output of DNN

Maximum Likelihood

❑ 딥러닝 모델의 Loss function

Maximum Likelihood $\max p(y|f_\theta(x))$

Negative log-likelihood $\min -log(p(y|f_\theta(x)))$

Sum of loss for each training sample $\min -\sum_i log(p(y_i|f_\theta(x_i)))$

Loss계산필요

Likelihood의 분포를 가정

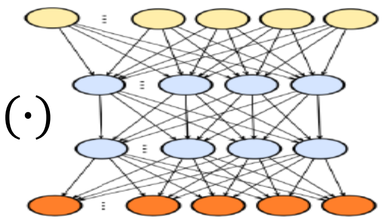
Case1) Gaussian distribution 이면
Mean Squared Error(MSE) Loss

Case2) Categorical distribution 이면
Cross-Entropy Loss

Output node

$f_\theta(x_i)$

DNN $f_\theta(\cdot)$



Input node

x_i

Value	O/X	Category
이게 얼마가 될거같니?	기냐? 아니냐?	종류중에 요건 뭐냐?
output을 그냥 받는다.	output에 sigmoid를 먹인다.	output에 softmax를 먹인다.

https://ratsgo.github.io/deep%20learning/2017/09/24/loss/

Maximum Likelihood

Univariate case

$$-log(p(y_i|f_\theta(x_i)))$$

Gaussian distribution

$$f_\theta(x_i) = \mu_i, \sigma_i = 1$$

$$p(y_i|\mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} exp(-\frac{(y_i - \mu_i)^2}{2\sigma_i^2})$$

$$log(p(y_i|\mu_i, \sigma_i)) = log \frac{1}{\sqrt{2\pi}\sigma_i} - \frac{(y_i - \mu_i)^2}{2\sigma_i^2}$$

$$-log(p(y_i|\mu_i)) = -log \frac{1}{\sqrt{2\pi}} + \frac{(y_i - \mu_i)^2}{2}$$

$$-log(p(y_i|\mu_i)) \propto \frac{(y_i - \mu_i)^2}{2} = \frac{(y_i - f_\theta(x_i))^2}{2}$$

Mean Squared Error (MSE)

Bernoulli distribution (categorical)

$$f_\theta(x_i) = p_i$$

$$p(y_i|p_i) = p_i^{y_i}(1 - p_i)^{1-y_i}$$

$$log(p(y_i|p_i)) = y_i log p_i + (1 - y_i) log(1 - p_i)$$

$$-log(p(y_i|p_i)) = -[y_i log p_i + (1 - y_i) log(1 - p_i)]$$

Cross-entropy

SMART DESIGN LAB KAIST

What Questions Do You Have?

nwkang@kaist.ac.kr

www.smartdesignlab.org