Deep Learning Basic

Jaewon Kim, Dankook Univ.

Chapter 1-2 Probability and Statistics

Part 1

Cross-Entropy

Cross-Entropy (Cost Function)

$$cost = \frac{1}{m}\sum_{i=1}^{m}(H(x^{(i)}) - y^{(i)})^2$$

$$\underset{W,b}{\text{minimize}}\ \boxed{cost(W,b)}$$

Cross-Entropy (Classification Cost)

Entropy : 불확실성, 무질서도

: 정보 ↑ , 확률 ↓

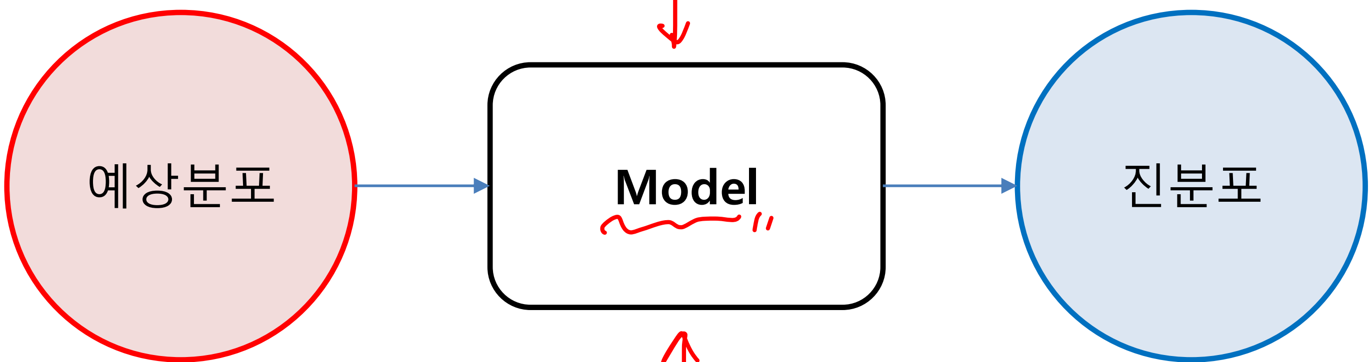
: 정보 ↓ , 확률 ↑

Cross-Entropy

실제 값과 예측 값의 차이를 줄이기 위한 Entropy

Cross-Entropy (Classification Cost)

- Apply to Machine Learning



Ex) *Newton' Second Law,* $\boxed{F = ma}$

Cross-Entropy (Classification Cost)

- Notation

$$\mathrm{CE}(p, q) = -\mathrm{E}_{x\sim p}[\log q(x)]$$
$$= -\sum_{x} p(x) \log q(x)$$

변수	분포
x	p
y	q

Difference with p and q ?

Cross-Entropy (Classification Cost)

KL-divergence Kullback-Leilbler divergence

KL-발산

$$D_{\mathrm{KL}}(p \parallel q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

$\log \frac{\text{진분포}}{\text{예상분포}}$

" *예상 분포 q(x)가 진 분포 p(x)를 얼마나 잘 따라 갔느냐?* "

If, p(x)=q(x) ?

$\log \frac{p}{p} = \log 1 = 0.$

Cross-Entropy (Classification Cost)

Expected value

KL- 발산,

기댓값.

$$\sum_x p(x) \log \frac{p(x)}{q(x)} = \mathrm{E}_{x \sim p} \left[\log \frac{p(x)}{q(x)} \right]$$

Cross-Entropy (Classification Cost)

Expected value

$$\underbrace{\sum_x p(x) \log \frac{p(x)}{q(x)}}_{} = \mathrm{E}_{x \sim p}\left[\log \frac{p(x)}{q(x)}\right]$$

$x \in \{a, b\}$

$$p(a) = \frac{1}{3} \qquad p(b) = \frac{2}{3}$$

$$q(a) = \frac{2}{3} \qquad q(b) = \frac{1}{3}$$

진 분포 $p(x)$

예상 분포 $q(x)$

$$D_{KL}(p \| q) = p(a) \log \frac{p(a)}{q(a)} + p(b) \log \frac{p(b)}{q(b)}$$

$$= \frac{1}{3} \log \frac{1}{2} + \frac{2}{3} \log 2$$

$$\left(= \underbrace{-\frac{1}{3} \log 2}_{p < q} + \underbrace{\frac{2}{3} \log 2}_{p > q} \right)$$

$p(x)$ < $q(x)$: 차이를 깎아준다,

$p(x)$ > $q(x)$: 부족하게 예상,,

Cross-Entropy (Classification Cost)

If you change the above equation,

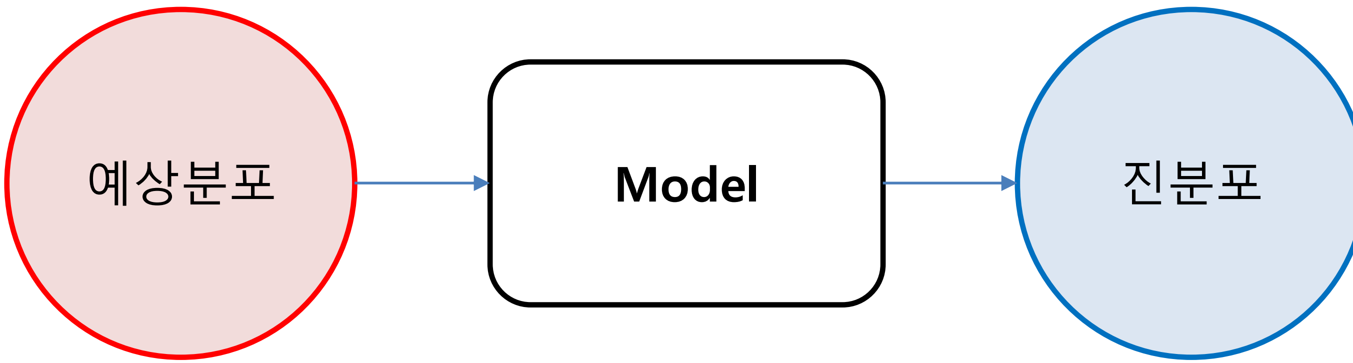
$(\because \log \frac{a}{b} = \log a - \log b)$

$$
\begin{aligned}
D_{\mathrm{KL}}(p \parallel q) &= \sum_x p(x) \log \frac{p(x)}{q(x)} \\
&= \sum_x p(x) \log p(x) - \sum_x p(x) \log q(x) \\
&= -\mathrm{H}(p) + \mathrm{CE}(p, q)
\end{aligned}
$$

Entropy

Cross Entropy

Cross-Entropy (Classification Cost)



If you apply to Machine Learing, You reduce KL-divergence !

Entropy ; Cross Entropy ,

Cross-Entropy (Classification Cost)

- At Classification

$$\text{MLP}(a) = \begin{pmatrix} P(Y_a = 1) \\ P(Y_a = 2) \\ \ldots \\ P(Y_a = C) \end{pmatrix}$$

Class : 1,2,....,C

Input : a

X_a : 진짜 클래스

Y_a : 예측 클래스

$$P(X_a = 1) = 1$$
$$P(X_a = c) = 0 \quad (c \neq 1)$$

$$\text{Entropy} \quad H(X_a) = -\sum_{c=1}^{C} P(X_a = c) \log P(X_a = c)$$

$$KL = \cancel{E} + CE$$

$$= -P(X_a = 1) \log P(X_a = 1)$$

$$= -1 \times \log 1 = 0$$

Cross-Entropy (Classification Cost)

$$D_{\mathrm{KL}}(p \parallel q) = \cancel{-\mathrm{H}(p)} + \mathrm{CE}(p, q) = \mathrm{CE}(p, q)$$

$-\log P$

로그 가능도

(log - Likelihood)

Cross-Entropy (Classification Cost) 기댓값.

- Why Cross-Entropy ?

Jensen's Equation

$$f(\mathrm{E}_{x\sim p}[x]) \leq \mathrm{E}_{x\sim p}[f(x)]$$

$$\downarrow$$

$$f\left(\frac{a+b}{2}\right) \leq \frac{f(a)+f(b)}{2}$$

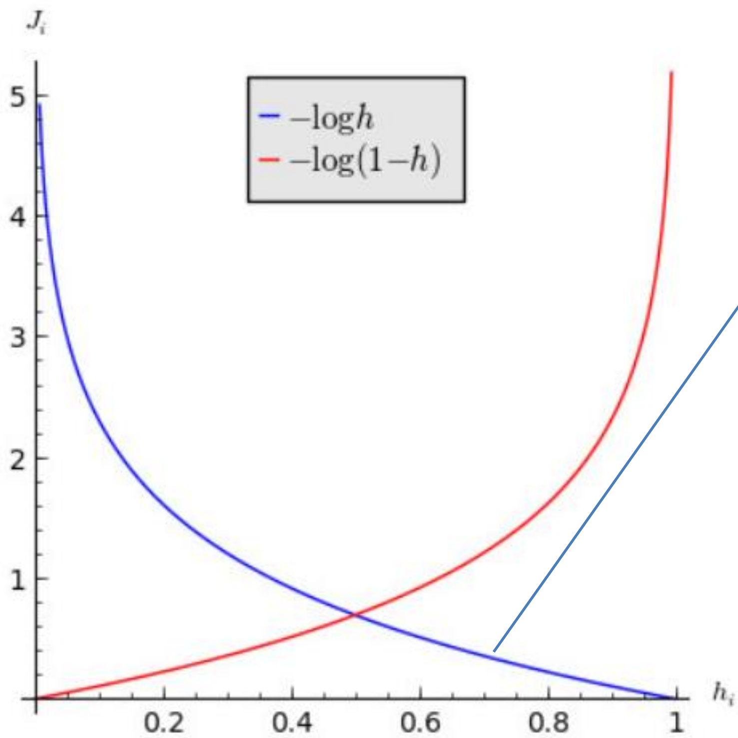
Cross-Entropy (Classification Cost)

- Why Cross-Entropy ?

Jensen's Equation

$$f(\mathrm{E}_{x\sim p}[x]) \leq \mathrm{E}_{x\sim p}[f(x)]$$

볼록 함수



$$f(x) = -\log(x)$$

Cross-Entropy (Classification Cost)

$$\mathrm{CE}(X_a, Y_a) = -\sum_{c=1}^{C} P(X_a=c) \log P(Y_a=c)$$

$$= -\sum_{c=1}^{C} P(c) \log q(c)$$

$$= \sum_{c=1}^{C} p(c) \, -\log q(c)$$

$$= \sum_{c=1}^{C} p(c) f(q(c)) = E_{c\sim p}[f(q(c))]$$

$$\boxed{f(E_{c\sim p}[q(c)]) \le E_{c\sim p}[f(q(c))]}$$

$$\boxed{f(E_{c\sim p}[q(c)]) = f\left(\sum_{c=1}^{C} p(c) q(c)\right)}$$

$$f\left(\sum_{c=1}^{C} p(c)q(c)\right) = f\left(P(X_a = Y_a)\right)$$

X_a P

Y_a q

$$f(x) = -\log x$$

주사위,,

① ②

$$\frac{1}{2} \times \frac{1}{2}$$

Cross-Entropy (Classification Cost)

$$f\left(\sum_{c=1}^{C} p(c)q(c)\right) = f\left(P(X_a = Y_a)\right)$$

Cross-Entropy (Classification Cost)

- Cross-Entorpy 부터 정리 $f(\mathrm{E}_{x\sim p}[x]) \le \mathrm{E}_{x\sim p}[f(x)]$

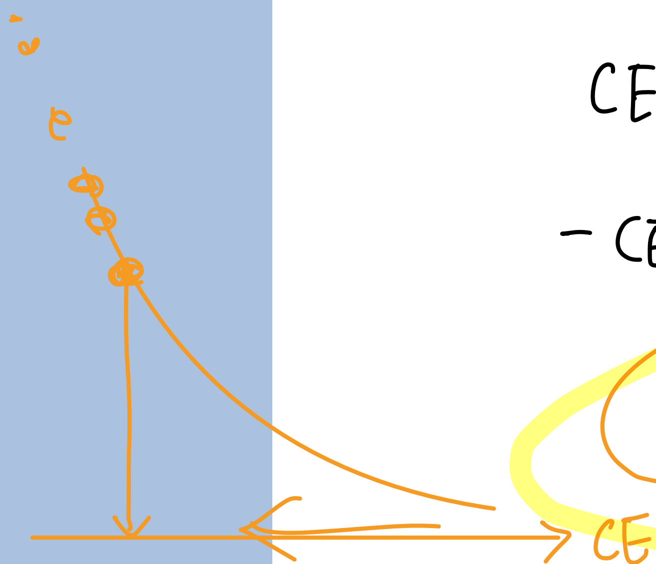
$$CE(X_a, Y_a) \ge f(P(X_a = Y_a)) \quad (\because f(x) = -\log x)$$

$$\downarrow$$

$$CE(X_a, Y_a) \ge -\log(P(X_a = Y_a))$$

$$-CE(X_a, Y_a) \le \log(P(X_a = Y_a))$$

$$e^{-CE(X_a, Y_a)} \le P(X_a = Y_a)$$



CE $\downarrow$ e $\uparrow$

$$\boxed{e^{-\mathrm{CE}(X_a, Y_a)} \le P(X_a = Y_a)}$$

Thank you..!!!

Reference : https://theeluwin.postype.com/post/

Mail : wodnjs284@naver.com