ACADENTIAL

Section 4. 손실 함수 심화 (Loss Function)

목차

- 섹션 0. 강의 소개
- 섹션 1. PyTorch 환경 설정
- 섹션 2. 딥러닝이란?
- 섹션 3. 손실 함수 (Loss Function)
- 섹션 4. 손실 함수에 대한 심화 이론 (Advanced Topics on Loss Function)
- 섹션 5. 경사 하강 (Gradient Descent)
- 섹션 6. 경사 하강에 대한 심화 이론 (Advanced Topics on Gradient Descent)

Objective

학습 목표

- One-hot-encoding에 대한 이해
- Entropy 개념에 대한 이해
- Cross Entropy Loss
- KL Divergence Loss

4-1. One-hot-encoding

One Hot Encoding

One Hot Encoding

- Categorical (범주형) 데이터를 처리하는데 사용되는 Encoding 방법
- 예를 들어서 동물의 종류 (3가지)에 대한 데이터:
  - “고양이”, “개”, “원숭이”
- 이것을 **One-hot-encoding**하면

“고양이”: [1, 0, 0]

“개”: [0, 1, 0]

“원숭이”: [0, 0, 1]

One Hot Encoding

One Hot Encoding

- Categorical (범주형) 데이터를 처리하는데 사용되는 Encoding 방법
- 이것을 **One-hot-encoding**하면

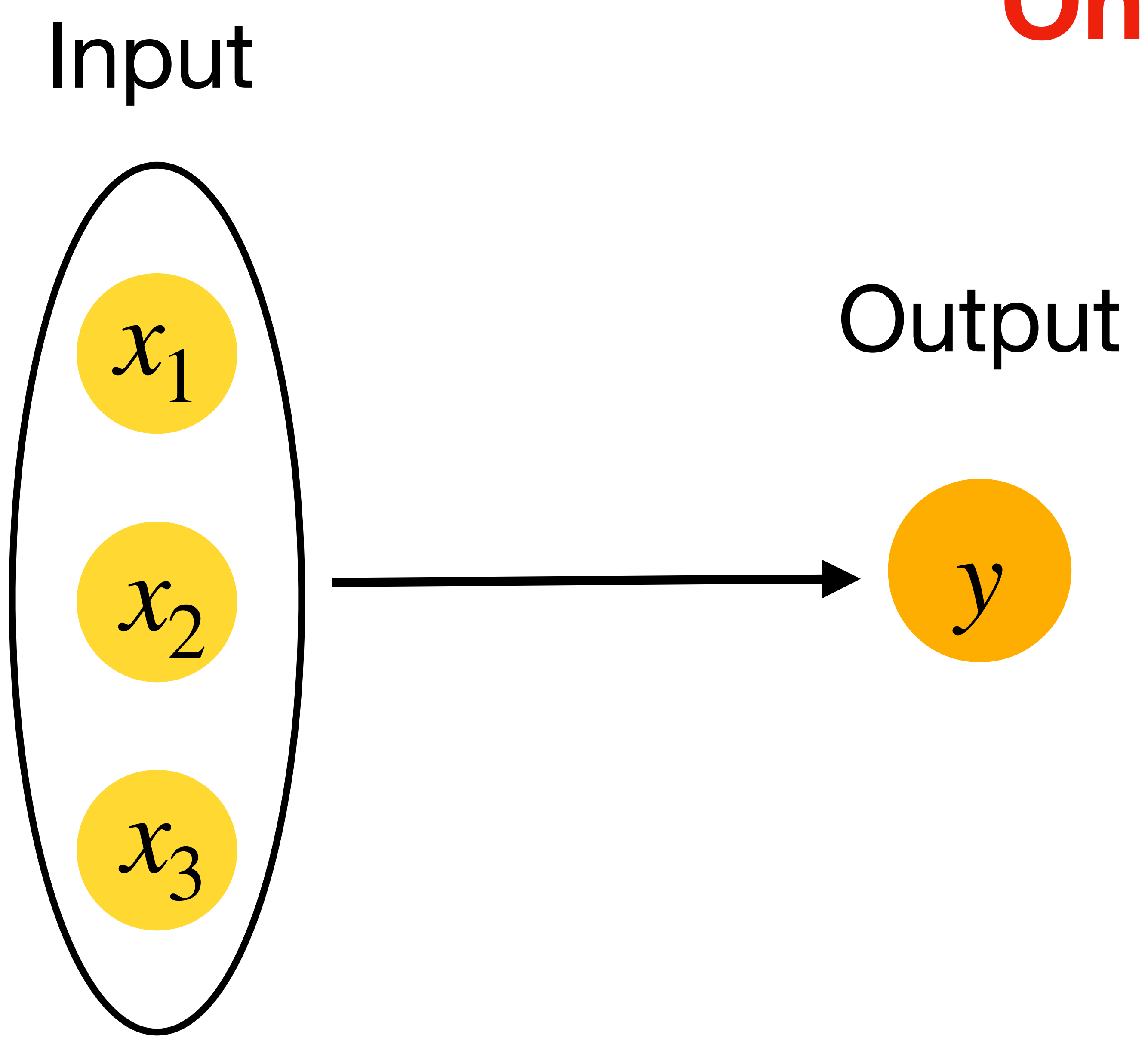
"고양이": [1, 0, 0]

"개": [0, 1, 0]

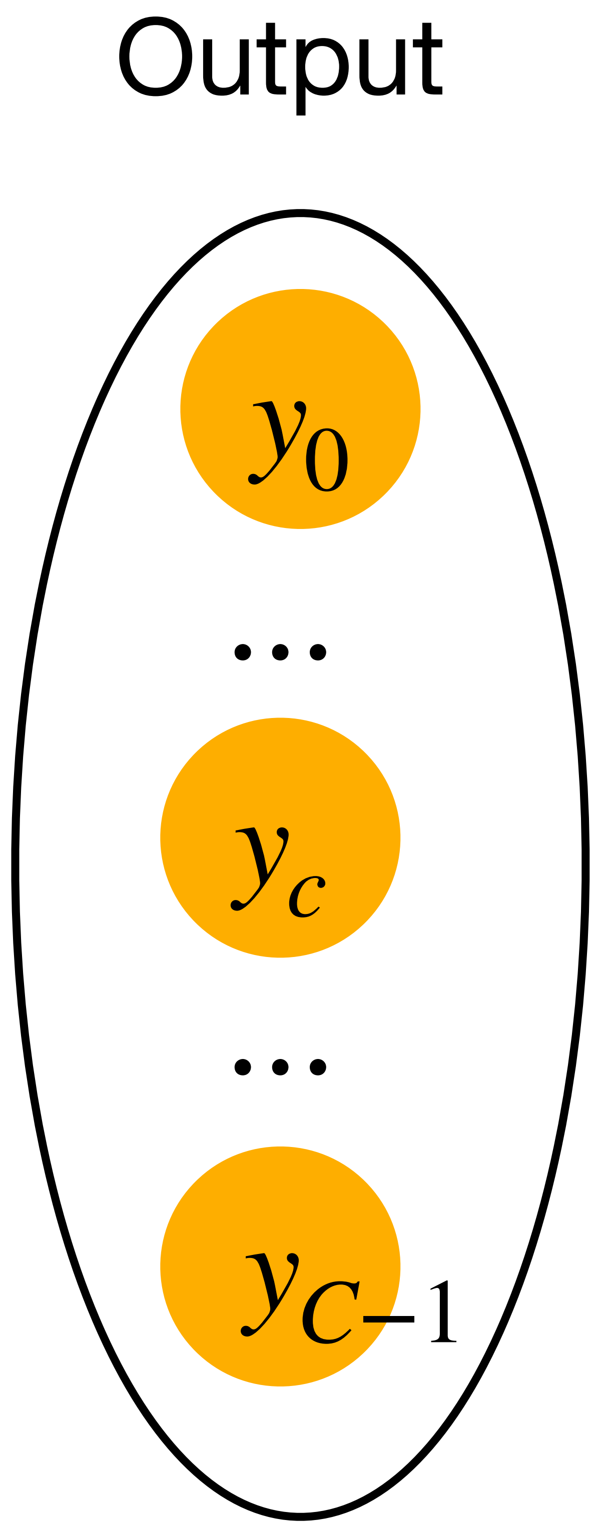
"원숭이": [0, 0, 1]

- 범주형 데이터를 **Vector로 변환**하는 기법!

One Hot Encoding



One Hot Encoding



index $c = c'$ 이면,

$$y_{c=c'} = 1$$

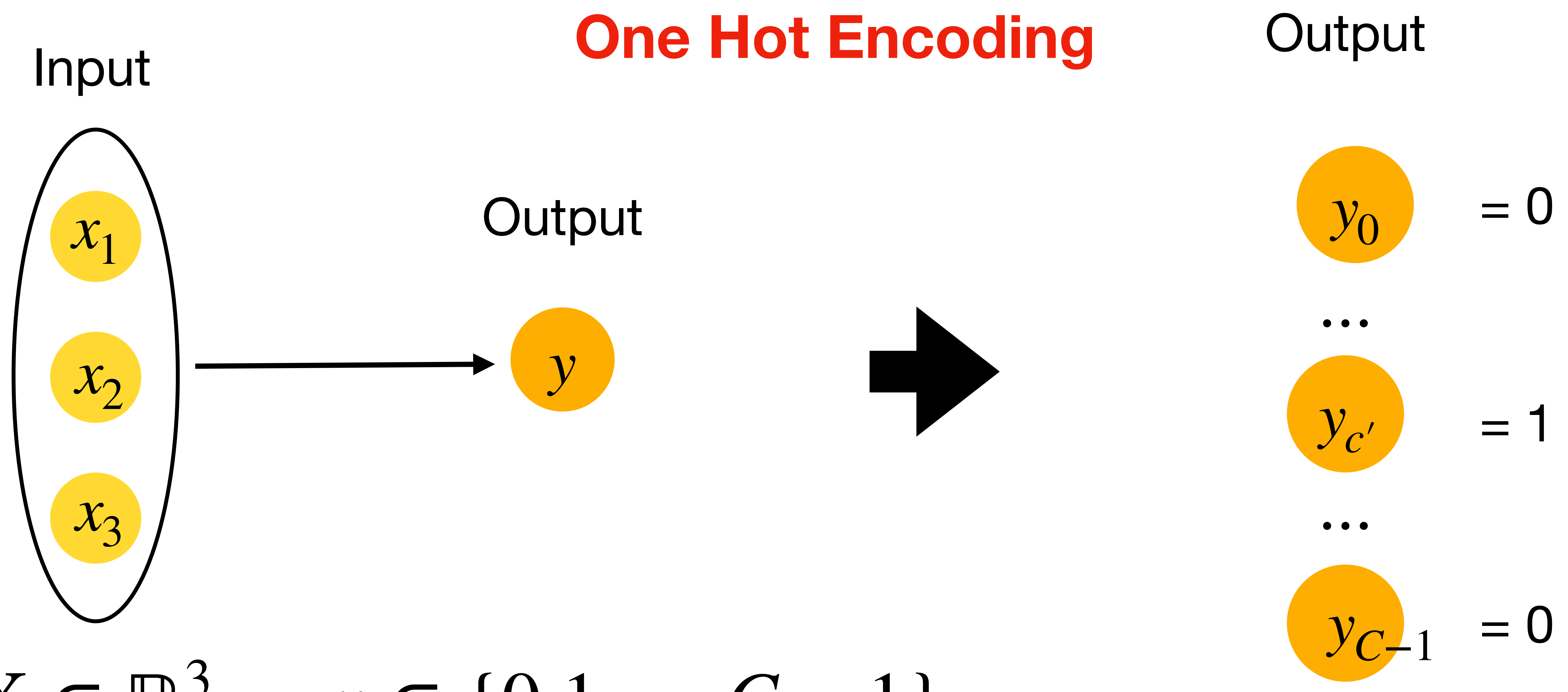
index $c \neq c'$ 이면

$$y_{c \neq c'} = 0$$

$\mathbf{X} \in \mathbb{R}^3$ $\quad y \in \{0, 1, \cdots, C-1\}$

Ground Truth Label = c'

One Hot Encoding



Ground Truth의 Index에 해당되는 element만 1의 가지고 나머지는 0의 값.

One-hot encoding!

$\mathbf{X} \in \mathbb{R}^3$ $\quad y \in \{0, 1, \cdots, C-1\}$

Ground Truth Label = c'

4-2. Entropy

Entropy

열역학 (Thermodynamics) 에서의 **Entropy**

물리 시스템의 **무질서한 정도**.

정보 이론 (Information Theory) 에서의 **Entropy**

확률 분포의 **불확실성**의 정도.

Entropy

정보 이론 (Information Theory) 에서의 **Entropy**

확률 분포의 **불확실성**의 정도.

예를 들어,

- 내일 해가 뜰 확률은? → $p_{sun} = 0.9999999\ldots$
- 사실상 1이다.
- 해가 뜨는 것에 대한 Entropy 낮음!

Entropy

Entropy

$$\sum_i -p_i \log p_i$$

- 예시: **Binary Label** (즉, $i \in \{0,1\}$)
- p = 데이터 샘플이 **Label을 가질 확률**에 대한 **모델의 예측값.**
- $1-p$ = **Label을 가지지 않을 확률**에 대한 예측값.

Entropy

Entropy

$$\sum_i - p_i \log p_i = - p \log p - (1-p) \log (1-p)$$

- 예시: **Binary Label** (즉, $i \in \{0,1\}$)
- p = 데이터 샘플이 **Label을 가질 확률**에 대한 **모델의 예측값.**
- $1 - p$ = **Label을 가지지 않을 확률**에 대한 예측값.

Entropy

$$\sum_i -p_i \log p_i$$

$$= -p\log p - (1-p)\log(1-p)$$

$$0 \le p < 0.5$$

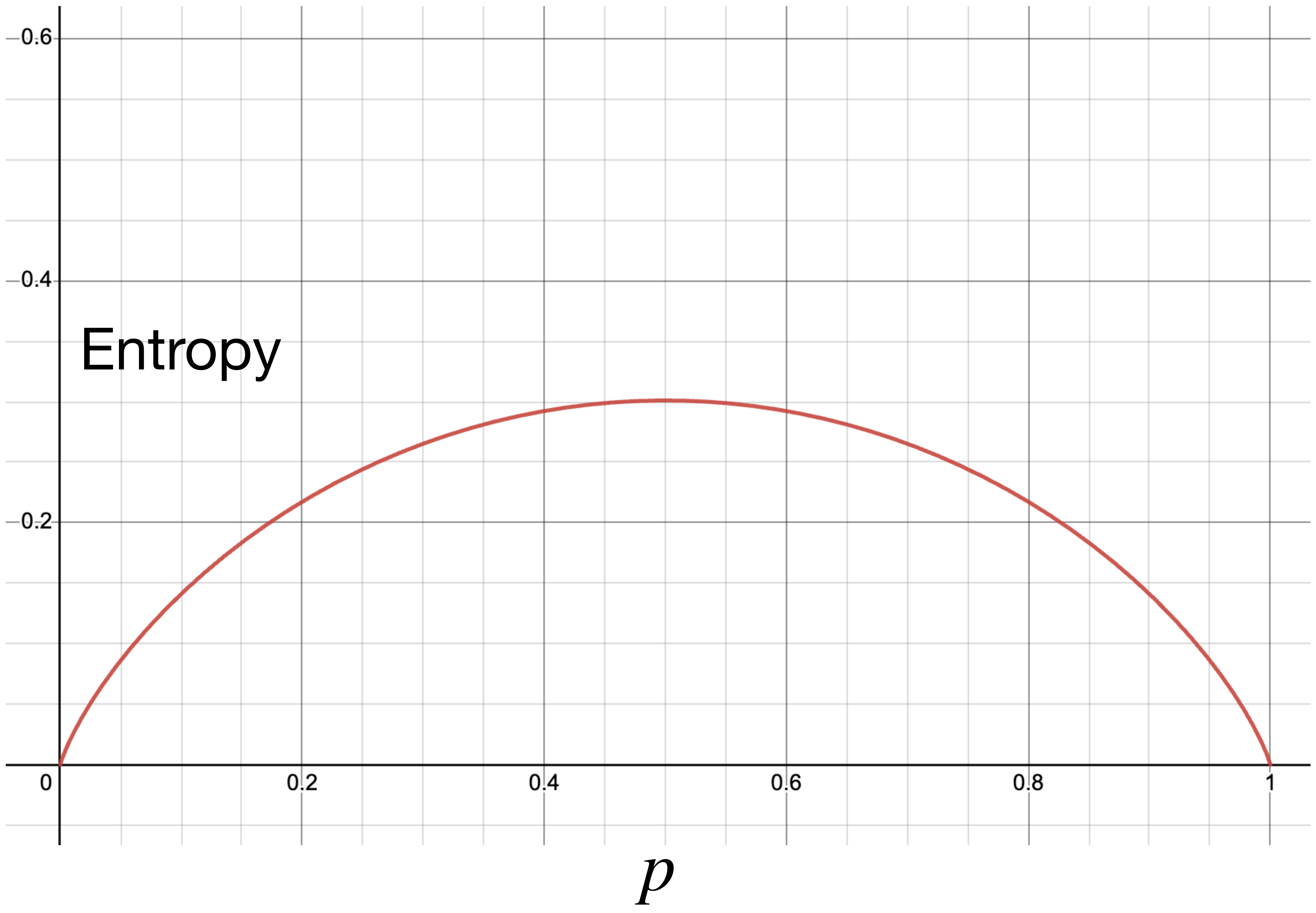
- p가 커지면서 Entropy도 증가

$$0.5 \le p \le 1$$

- p가 커지면서 Entropy는 감소

$$p = 0.5$$

- Entropy가 최대치



Entropy

$$\sum_i - p_i \log p_i$$

$$= - p \log p - (1 - p)\log(1 - p)$$

$p = 0$

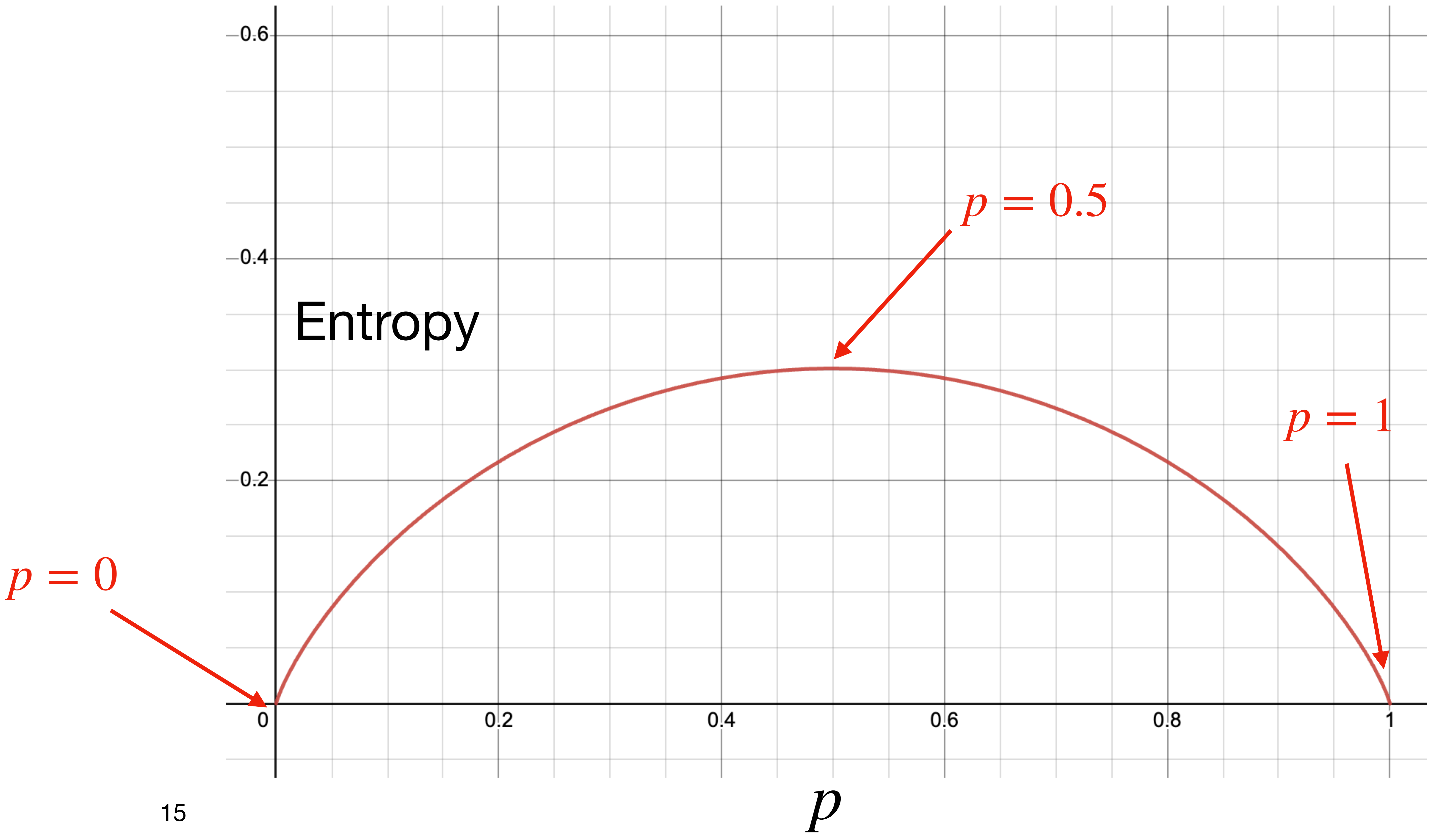
- Label을 가지지 않는 것에 확신
- 불확실함 최소 → Entropy 최소

$p = 1$

- Label을 가지는 것에 확신
- 불확실함 최소 → Entropy 최소

$p = 0.5$

- Label을 가질 확률을 완전히 Random
- 불확실함 최대 → Entropy 최대



1. One-hot-encoding
2. Entropy

4-3. Cross Entropy Loss

Cross Entropy Loss

- **L1 Loss (MAE Loss)**

$$\sum_{c=1}^{C} | Y_{i,c} - \hat{Y}_{i,c} |$$

- **Cross Entropy Loss**

$$\sum_{c=1}^{C} - Y_{i,c} \log \hat{Y}_{i,c}$$

- **Kullback-Leibler Divergence Loss (KL Divergence)**

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log \left(\frac{Y_{i,c}}{\hat{Y}_{i,c}} \right)$$

Cross Entropy Loss

- **Cross Entropy Loss (CE Loss)**

$$\sum_{c=1}^{C} \boxed{- Y_{i,c} \log \hat{Y}_{i,c}}$$

Hard Label & Ground Truth class label가 c' 로 가정

$c \neq c'$ 의 경우

$$-Y_{i,c} \log \hat{Y}_{i,c}$$

$c = c'$ 의 경우

$$-Y_{i,c} \log \hat{Y}_{i,c}$$

ACADENTIAL

Cross Entropy Loss

- **Cross Entropy Loss (CE Loss)**

$$\sum_{c=1}^{C} - Y_{i,c} \log \hat{Y}_{i,c}$$

Ground Truth class label가 c' 로 가정

$c \neq c'$ 의 경우

$$-\overset{=0}{Y_{i,c}} \log \hat{Y}_{i,c} = 0$$

$c = c'$ 의 경우

$$-Y_{i,c} \log \hat{Y}_{i,c}$$

Cross Entropy Loss

- **Cross Entropy Loss (CE Loss)**

$$\sum_{c=1}^{C} -Y_{i,c} \log \hat{Y}_{i,c}$$

Ground Truth class label가 c' 로 가정

$c \neq c'$ 의 경우

$$-\underbrace{Y_{i,c}}_{=0} \log \hat{Y}_{i,c} = 0$$

$c = c'$ 의 경우

$$-\underbrace{Y_{i,c}}_{=1} \underbrace{\log \hat{Y}_{i,c}}_{\leq 0} \geq 0$$

Cross Entropy Loss

- **Cross Entropy Loss (CE Loss)**

$$\underbrace{\sum_{c=1}^{C} - Y_{i,c} \log \hat{Y}_{i,c}}$$

$$\downarrow$$

$$-Y_{i,c=c'} \log(\hat{Y}_{i,c=c'})$$

$$\downarrow$$

$$-\log(\hat{Y}_{i,c=c'})$$

$c \neq c'$ 의 경우

$$-\underbrace{\cancel{Y_{i,c}}}_{=0} \log(\hat{Y}_{i,c}) = 0$$

$c = c'$ 의 경우

$$-\underbrace{\cancel{Y_{i,c}}}_{=1} \underbrace{\log(\hat{Y}_{i,c})}_{\leq 0} \geq 0$$

Cross Entropy Loss

- **Cross Entropy Loss (CE Loss)**

$$\sum_{c=1}^{C} - Y_{i,c} \log \hat{Y}_{i,c}$$

$$\downarrow$$

$$-Y_{i,c=c'} \log(\hat{Y}_{i,c=c'})$$

$$\downarrow$$

$$\boxed{-\log(\hat{Y}_{i,c=c'})}$$

Ground Truth Label c' 일때,

$\hat{Y}_{i,c=c'}$ 가 높을수록 CE Loss가 낮아짐.

$\hat{Y}_{i,c=c'}$ 가 낮을수록 CE Loss가 높아짐.

Cross Entropy Loss

- **Cross Entropy Loss (CE Loss)**

$$\underbrace{\sum_{c=1}^{C} - Y_{i,c} \log \hat{Y}_{i,c}}$$

$$\downarrow$$

$$-Y_{i,c=c'} \log(\hat{Y}_{i,c=c'})$$

$$\downarrow$$

$$\boxed{-\log(\hat{Y}_{i,c=c'})}$$

Ground Truth Label c' 일때,

$\hat{Y}_{i,c=c'}$ 가 높을수록 **(잘 맞춘 것)** CE Loss가 **낮아짐.**

$\hat{Y}_{i,c=c'}$ 가 낮을수록 **(못 맞춘 것)** CE Loss가 **높아짐.**

4-4. Kullback-Leibler Divergence Loss

Kullback-Leibler Divergence Loss

- **L1 Loss (MAE Loss)**

$$\sum_{c=1}^{C} |Y_{i,c} - \hat{Y}_{i,c}|$$

- **Cross Entropy Loss**

$$\sum_{c=1}^{C} -Y_{i,c} \log \hat{Y}_{i,c}$$

- **Kullback-Leibler Divergence Loss (KL Divergence)**

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log \left(\frac{Y_{i,c}}{\hat{Y}_{i,c}} \right)$$

Kullback-Leibler Divergence Loss

- **Kullback-Leibler Divergence Loss (KL Divergence)**

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log \left(\frac{Y_{i,c}}{\hat{Y}_{i,c}} \right) = \sum_{c=1}^{C} Y_{i,c} \log Y_{i,c} - Y_{i,c} \log \hat{Y}_{i,c}$$

ACADENTIAL

Kullback-Leibler Divergence Loss

- **Kullback-Leibler Divergence Loss (KL Divergence)**

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = \sum_{c=1}^{C} \boxed{Y_{i,c} \log Y_{i,c}} - Y_{i,c} \log \hat{Y}_{i,c}$$

Negative Entropy of $Y_{i,c}$

Entropy

$$\sum_{i} - p_i \log p_i$$

Kullback-Leibler Divergence Loss

- **Kullback-Leibler Divergence Loss (KL Divergence)**

Cross Entropy Term

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = \sum_{c=1}^{C} \boxed{Y_{i,c} \log Y_{i,c}} \boxed{- Y_{i,c} \log \hat{Y}_{i,c}}$$

Negative Entropy of $Y_{i,c}$

Cross Entropy Loss

$$\sum_{c=1}^{C} - Y_{i,c} \log \hat{Y}_{i,c}$$

ACADENTIAL

Kullback-Leibler Divergence Loss

- **Kullback-Leibler Divergence Loss (KL Divergence)**

Cross Entropy Term

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = \sum_{c=1}^{C} \boxed{Y_{i,c} \log Y_{i,c}} \boxed{- Y_{i,c} \log \hat{Y}_{i,c}}$$

Negative Entropy of $Y_{i,c}$

Negative Entropy

$$\sum_{c}^{C} Y_{i,c} \log Y_{i,c}$$

Cross Entropy Loss

$$\sum_{c=1}^{C} - Y_{i,c} \log \hat{Y}_{i,c}$$

Kullback-Leibler Divergence Loss

- **Kullback-Leibler Divergence Loss (KL Divergence)**

Cross Entropy Term

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = \sum_{c=1}^{C} \boxed{Y_{i,c} \log Y_{i,c}} \boxed{- Y_{i,c} \log \hat{Y}_{i,c}}$$

Negative Entropy of $Y_{i,c}$

Negative Entropy

$$\sum_{c}^{C} Y_{i,c} \log Y_{i,c}$$

Cross Entropy Loss

$$\sum_{c=1}^{C} - Y_{i,c} \log \hat{Y}_{i,c}$$

$\hat{Y}_{i,c} \rightarrow Y_{i,c}$ 이면 Entropy가 된다!

KL Divergence의 해석

- **Kullback-Leibler Divergence Loss (KL Divergence)**

Negative Entropy of $Y_{i,c}$

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = \sum_{c=1}^{C} \boxed{Y_{i,c} \log Y_{i,c}} - Y_{i,c} \log \hat{Y}_{i,c}$$

- Negative Entropy가 높다 → 해당 샘플의 Label에 대해서 더 확신.
- 해당 샘플에 대해서는 더 잘 맞춰야함.
- 따라서, 더 높은 Loss을 주는 셈.

ACADENTIAL

KL Divergence의 해석

- **Kullback-Leibler Divergence Loss (KL Divergence)**

Negative Entropy of $Y_{i,c}$

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = \sum_{c=1}^{C} \boxed{Y_{i,c} \log Y_{i,c}} - Y_{i,c} \log \hat{Y}_{i,c}$$

그런데 만약 Label이 Soft label이 아닌 Hard label일 경우?

ACADENTIAL

KL Divergence의 해석

Hard Label $Y_{i,c} \in \{0,1\}$

$$\textbf{Entropy} = \sum_{c} -Y_{i,c} \log Y_{i,c} = 0$$

$$\textbf{KL Div. Loss} = \sum_{c=1}^{C} \cancelto{=0}{Y_{i,c} \log Y_{i,c}} - Y_{i,c} \log \hat{Y}_{i,c}$$

$$= \sum_{c} -Y_{i,c} \log \hat{Y}_{i,c} = \textbf{CE Loss}$$

ACADENTIAL

Soft Label vs. Hard Label

Output

y_0 = 0

…

$y_{c'}$ = 1

…

y_{C-1} = 0

Hard Label

↓

KL Divergence Loss

=

Cross Entropy Loss

Output

y_0 = 0

…

$y_{c'}$ = 0.8

$y_{c''}$ = 0.2

…

y_{C-1} = 0

Soft Label

KL Divergence Loss

=

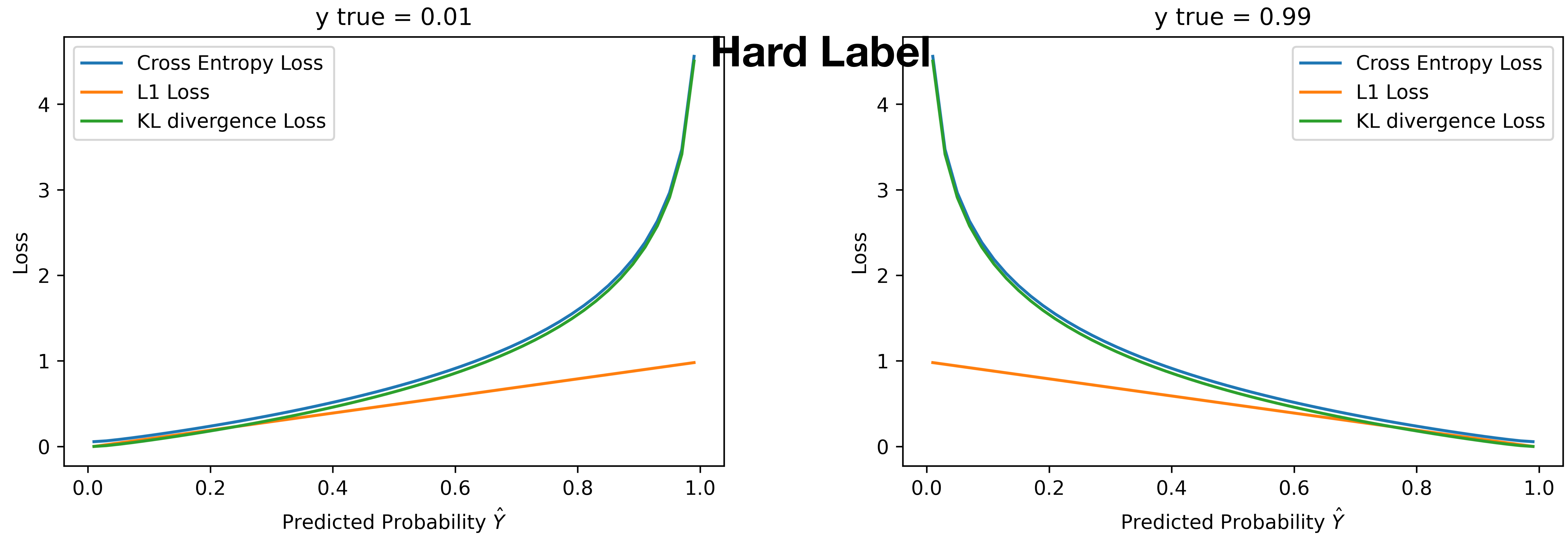
Negative Entropy

+

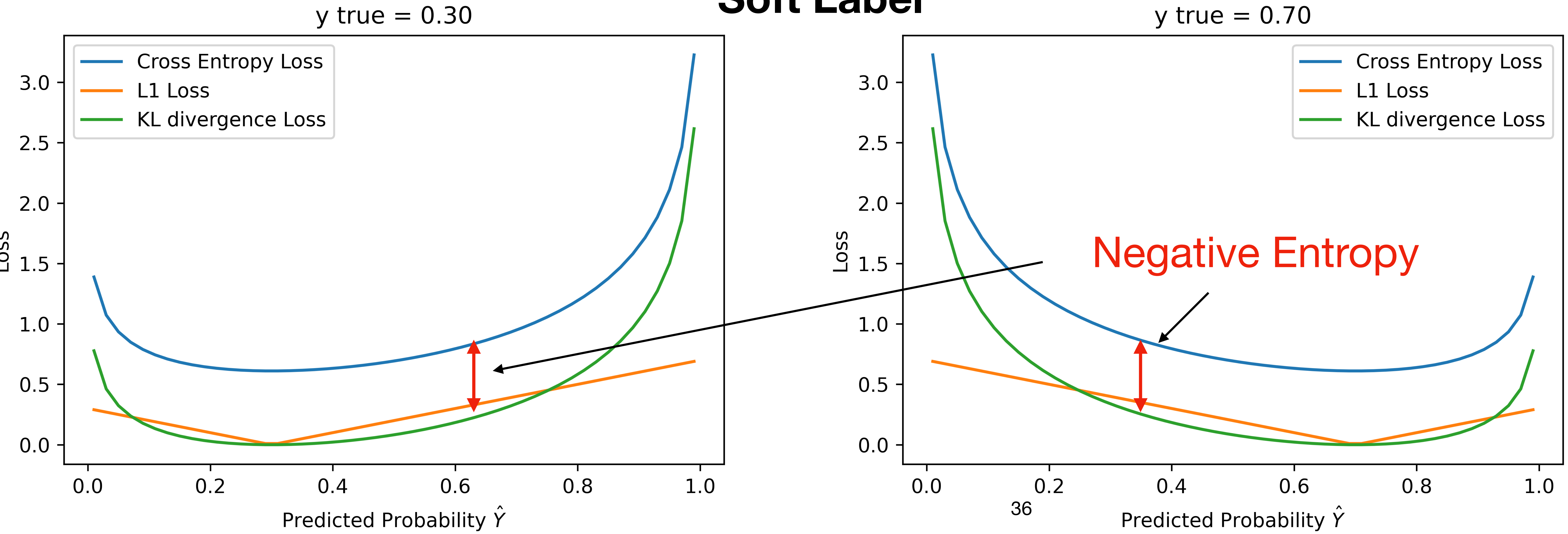
Cross Entropy Loss

ACADENTIA

Hard Label



Soft Label



Negative Entropy

$$= \sum_{c=1}^{C} Y_{i,c} \log Y_{i,c}$$

4-5. KL Divergence 2번째 해석

ACADENTIAL

KL Divergence의 2번째 해석

- **Kullback-Leibler Divergence Loss (KL Divergence)**

Cross Entropy Term

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = \sum_{c=1}^{C} Y_{i,c} \log Y_{i,c} - Y_{i,c} \log \hat{Y}_{i,c}$$

Negative Entropy of $Y_{i,c}$

KL Divergence의 2번째 해석

- **Kullback-Leibler Divergence Loss (KL Divergence)**

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = \sum_{c=1}^{C} Y_{i,c} \log Y_{i,c} - Y_{i,c} \log \hat{Y}_{i,c}$$

$$= \sum_{c=1}^{C} Y_{i,c} \left(\log Y_{i,c} - \log \hat{Y}_{i,c}\right)$$

KL Divergence의 2번째 해석

- **Kullback-Leibler Divergence Loss (KL Divergence)**

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = \sum_{c=1}^{C} Y_{i,c} \log Y_{i,c} - Y_{i,c} \log \hat{Y}_{i,c}$$

$$= \sum_{c=1}^{C} Y_{i,c} \left(\log Y_{i,c} - \log \hat{Y}_{i,c}\right)$$

$$\sum_{i} p_i \cdot f(p_i) = \mathrm{E}_p[f]$$

ACADENTIAL

KL Divergence의 2번째 해석

- **Kullback-Leibler Divergence Loss (KL Divergence)**

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = Y_{i,c}\log(Y_{i,c}) - Y_{i,c}\log\hat{Y}_{i,c}$$

$$= \sum_{c=1}^{C} Y_{i,c}\left(\log Y_{i,c} - \log\hat{Y}_{i,c}\right)$$

$$= \boxed{\mathrm{E}_{Y_{i,c}}\left[\log Y_{i,c} - \log\hat{Y}_{i,c}\right]}$$

확률 분포 $Y_{i,c}$에 대한 $\log Y_{i,c} - \log\hat{Y}_{i,c}$ 의 기댓값입니다.

(Expectation of log difference between $Y_{i,c}$ and $\hat{Y}_{i,c}$ with respect to $Y_{i,c}$ distribution)

KL Divergence의 2번째 해석

ACADENTIAL

- **Kullback-Leibler Divergence Loss**

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log \left(\frac{Y_{i,c}}{\hat{Y}_{i,c}} \right) = \mathrm{E}_{Y_{i,c}} \left[\log Y_{i,c} - \log \hat{Y}_{i,c} \right]$$

- $\log(Y_{i,c})$와 $\log(\hat{Y}_{i,c})$ **간의 차이에 대한** $Y_{i,c}$ **분포의 기댓값.**
- **즉, "ground truth label** $Y_{i,c}$ **분포의 입장"에서 본** $\log(Y_{i,c})$**와** $\log(\hat{Y}_{i,c})$ **간의 차이.**

KL Divergence의 2번째 해석

- **Kullback-Leibler Divergence Loss (KL Divergence)**

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = \mathrm{E}_{Y_{i,c}}\left[\log Y_{i,c} - \log \hat{Y}_{i,c}\right]$$

$$\boxed{= \mathrm{E}_{Y_{i,c}}\left[\log Y_{i,c}\right] - \mathrm{E}_{Y_{i,c}}\left[\log \hat{Y}_{i,c}\right]}$$

- KL Divergence은 확률 분포 $Y_{i,c}$에 대해서 $\log Y_{i,c}$의 기댓값과 $\log \hat{Y}_{i,c}$가 같아지도록 하는 것!

KL Divergence의 2번째 해석

- **Kullback-Leibler Divergence Loss (KL Divergence)**

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = \mathrm{E}_{Y_{i,c}}\left[\log Y_{i,c} - \log \hat{Y}_{i,c}\right]$$

$$\boxed{= \mathrm{E}_{Y_{i,c}}\left[\log Y_{i,c}\right] - \mathrm{E}_{Y_{i,c}}\left[\log \hat{Y}_{i,c}\right]}$$

참고 사항:

- KL Divergence가 0이 되는 것은 $Y = \hat{Y}$에 대한 필요 조건이지 충분 조건은 아니다.
- 기댓값이 일치하지만 분산 (Variance) 혹은 Higher order statistics이 같아지도록 강제 하지 않음.

ACADENTIAL

4-6.Cross Entropy와 KL Divergence 에 대한 경사

Cross Entropy와 KL Divergence

- **Kullback-Leibler Divergence Loss (KL Divergence)**

Negative Entropy of $Y_{i,c}$

$$L_{KL}(\hat{Y}_i, Y_i) = \sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = \sum_{c=1}^{C} \boxed{Y_{i,c} \log Y_{i,c}} \boxed{- Y_{i,c} \log \hat{Y}_{i,c}}$$

Cross Entropy Term

Cross Entropy와 KL Divergence

- **Kullback-Leibler Divergence Loss (KL Divergence)**

Negative Entropy of $Y_{i,c}$

$$L_{KL}(\hat{Y}_i, Y_i) = \sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = \sum_{c=1}^{C} \boxed{Y_{i,c} \log Y_{i,c}} \boxed{- Y_{i,c} \log \hat{Y}_{i,c}}$$

Cross Entropy Term

- **Gradient Descent 핵심**

경사하강은 경사의 음의 방향으로 모델의 weight을 update 해주는 것.

$$W_{t+1} = W_t - \lambda \cdot \frac{dL}{dW}$$

Cross Entropy와 KL Divergence

- **Kullback-Leibler Divergence Loss (KL Divergence)**

$$L_{KL}(\hat{Y}_i, Y_i) = \sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = \sum_{c=1}^{C} \boxed{Y_{i,c} \log Y_{i,c} - Y_{i,c} \log \hat{Y}_{i,c}}$$

- **Gradient Descent 핵심**

경사하강은 경사의 음의 방향으로 모델의 weight을 update 해주는 것.

$$W_{t+1} = W_t - \lambda \cdot \frac{\boxed{dL}}{dW}$$

Cross Entropy와 KL Divergence

- **KL Divergence Loss에 대한 경사**

$$W_{t+1} = W_t - \lambda \cdot \frac{dL_{KL}}{dW}$$

$$= W_t - \lambda \cdot \sum_{c=1}^{C} \frac{d}{dW}\left(Y_{i,c}\log Y_{i,c} - Y_{i,c}\log \hat{Y}_{i,c}\right)$$

$Y_{i,c} \perp\!\!\!\perp W$, $\hat{Y}_{i,c} \not\!\perp\!\!\!\perp W$

Cross Entropy와 KL Divergence

- **KL Divergence Loss에 대한 경사**

$$W_{t+1} = W_t - \lambda \cdot \frac{dL_{KL}}{dW}$$

$$= W_t - \lambda \cdot \sum_{c=1}^{C} \frac{d}{dW} \left(\cancelto{=0}{Y_{i,c} \log Y_{i,c}} - Y_{i,c} \log \hat{Y}_{i,c} \right)$$

$Y_{i,c} \perp\!\!\!\perp W$, $\hat{Y}_{i,c} \not\!\perp\!\!\!\perp W$

Cross Entropy와 KL Divergence

ACADENTIAL

- **KL Divergence Loss에 대한 경사**

$$W_{t+1} = W_t - \lambda \cdot \frac{dL_{KL}}{dW}$$

$$= W_t - \lambda \cdot \sum_{c=1}^{C} \frac{d}{dW} \left(\underbrace{\cancel{Y_{i,c} \log Y_{i,c}}}_{=0} \underbrace{\boxed{- Y_{i,c} \log \hat{Y}_{i,c}}}_{\text{Cross Entropy Term}} \right)$$

$Y_{i,c} \perp\!\!\!\perp W$, $\hat{Y}_{i,c} \not\perp\!\!\!\perp W$

$$= W_t - \lambda \cdot \frac{dL_{CE}}{dW}$$

Cross Entropy와 KL Divergence

경사 하강 기반의 최적화:

KL Divergence Loss = Cross Entropy Loss

4. Section 4 요약

Section Summary

One Hot Encoding

- Categorical (범주형) 데이터를 처리하는데 사용되는 Encoding 방법
- 이것을 **One-hot-encoding**하면

“고양이”: [**1**, 0, 0]

“개”: [0, **1**, 0]

“원숭이”: [0, 0, **1**]

- 범주형 데이터를 **Vector로 변환**하는 기법!

ACADENTIAL

Section Summary

Entropy

열역학 (Thermodynamics) 에서의 **Entropy**

물리 시스템의 **무질서한 정도**.

정보 이론 (Information Theory) 에서의 **Entropy**

확률 분포의 **불확실성**의 정도.

Cross Entropy Loss

Cross Entropy Loss (CE Loss)

$$\sum_{c=1}^{C} - Y_{i,c} \log \hat{Y}_{i,c}$$

$$\downarrow$$

$$- Y_{i,c=c'} \log(\hat{Y}_{i,c=c'})$$

$$\downarrow$$

$$\boxed{-\log(\hat{Y}_{i,c=c'})}$$

Ground Truth Label c' 일때,

$\hat{Y}_{i,c=c'}$ 가 높을수록 **(잘 맞춘 것)** CE Loss가 **낮아짐.**

$\hat{Y}_{i,c=c'}$ 가 낮을수록 **(못 맞춘 것)** CE Loss가 **높아짐.**

Section Summary

KL Divergence

- **Kullback-Leibler Divergence Loss (KL Divergence)**

Cross Entropy Term

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) = \sum_{c=1}^{C} \boxed{Y_{i,c} \log Y_{i,c}} \boxed{- Y_{i,c} \log \hat{Y}_{i,c}}$$

Negative Entropy of $Y_{i,c}$

Negative Entropy

$$\sum_{c}^{C} Y_{i,c} \log Y_{i,c}$$

Cross Entropy Loss

$$\sum_{c=1}^{C} - Y_{i,c} \log \hat{Y}_{i,c}$$

Section Summary

KL Divergence

- **Kullback-Leibler Divergence Loss (KL Divergence)**

확률 분포 $Y_{i,c}$에 대한 $\log Y_{i,c} - \log \hat{Y}_{i,c}$ 의 기댓값.

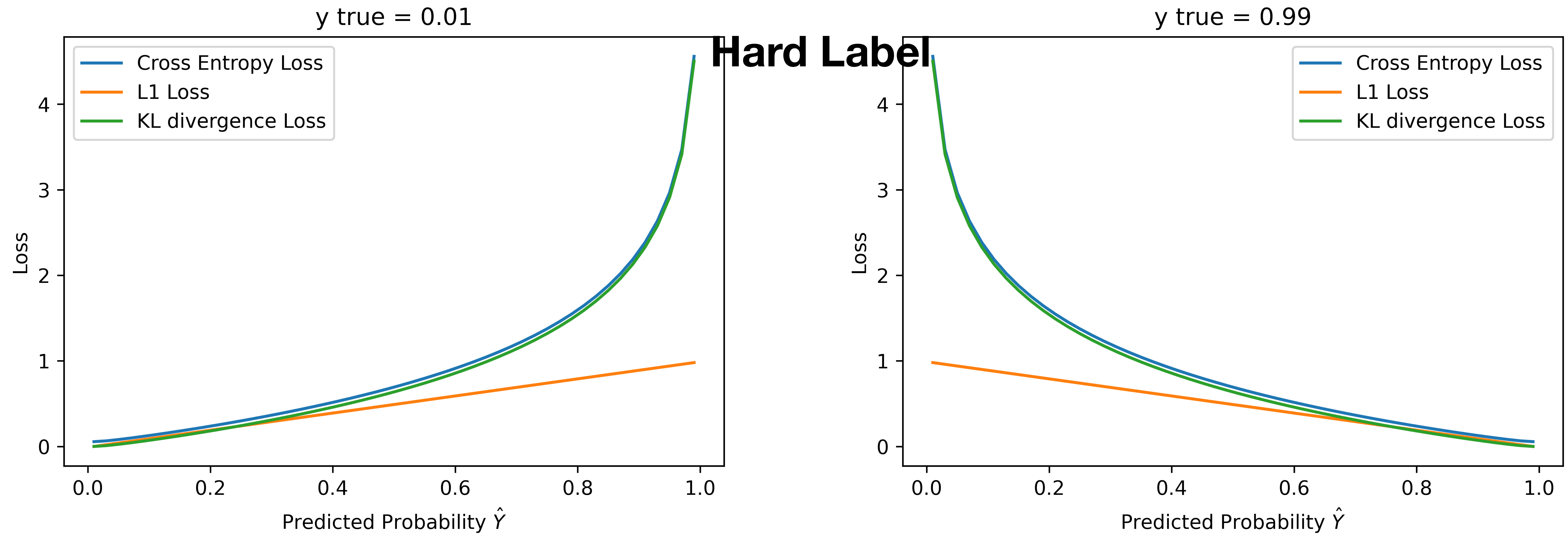
(Expectation of log difference between $Y_{i,c}$ and $\hat{Y}_{i,c}$ with respect to $Y_{i,c}$ distribution)

$$\sum_{c=1}^{C} Y_{i,c} \cdot \log\left(\frac{Y_{i,c}}{\hat{Y}_{i,c}}\right) \boxed{= \mathrm{E}_{Y_{i,c}}\left[\log Y_{i,c} - \log \hat{Y}_{i,c}\right]}$$

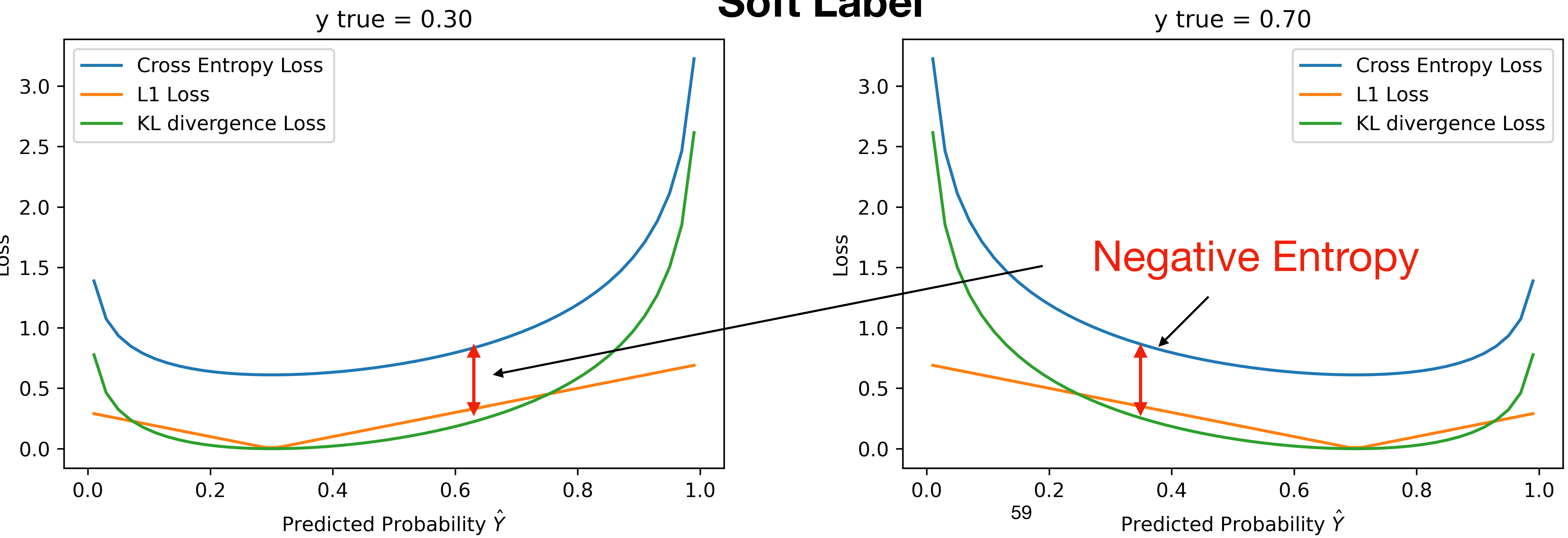
확률 분포 $Y_{i,c}$에 대해서 $\log Y_{i,c}$의 기댓값과 $\log \hat{Y}_{i,c}$의 기댓값의 차이

$$\boxed{= \mathrm{E}_{Y_{i,c}}\left[\log Y_{i,c}\right] - \mathrm{E}_{Y_{i,c}}\left[\log \hat{Y}_{i,c}\right]}$$

Hard Label



Soft Label



Negative Entropy

$$= \sum_{c=1}^{C} Y_{i,c} \log Y_{i,c}$$

Section Summary

- **KL Divergence Loss에 대한 경사**

$$W_{t+1} = W_t - \lambda \cdot \frac{dL_{KL}}{dW}$$

$$= W_t - \lambda \cdot \sum_{c=1}^{C} \frac{d}{dW} \left(\underbrace{\cancel{Y_{i,c} \log Y_{i,c}}}_{=0} \underbrace{\boxed{- Y_{i,c} \log \hat{Y}_{i,c}}}_{\text{Cross Entropy Term}} \right)$$

$Y_{i,c} \perp\!\!\!\perp W$, $\hat{Y}_{i,c} \not\perp\!\!\!\perp W$

$$= W_t - \lambda \cdot \frac{dL_{CE}}{dW}$$

Next Up!

Next Up

What is Deep Learning?

섹션 3. 손실 함수 (Loss Function)

- Neural Network가 학습되는 과정 = weight값이 최적화되는 과정

Gradient Descent (경사 하강)을 통한 **Loss function (손실 함수)** 값을 최소화하도록

weight 값을 최적화하여 **점진적으로** 모델의 예측 정확도를 높인다.

섹션 5. 경사 하강법 (Gradient Descent)

Next Up

What is Deep Learning?

섹션 3. 손실 함수 (Loss Function)

- Neural Network가 학습되는 과정 = weight값이 최적화되는 과정

Gradient Descent (경사 하강)을 통한 **Loss function (손실 함수)** 값을 최소화하도록

weight 값을 최적화하여 **점진적으로** 모델의 예측 정확도를 높인다.

섹션 5. 경사 하강법 (Gradient Descent)

Loss Function을 최소화하는 방법인 Gradient Descent에 대해서 배워보겠습니다!