


- Logistic Regression Goal.

⇒ Binary Classification

- Linear Regression의 확장.

① 연속형 값을 출력.

② 각 설명변수의 영향력을 알 수 있음.

③ 각 설명변수의 비율의 변화에 따라
종속변수의 class를 설명하는.

- Logistic Regression 흐름.

$$- \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_n x_n$$

우리의 \hat{y} 은 {0, 1}의 범위형 값을 갖음.

but, LR의 출력 범위는 $(-\infty, \infty)$

- 목표수정 1.

\hat{y} 예측값을 구하는 대신 확률값을 목표
수정.

$$\hat{y} \in [0, 1].$$

- 목표수정 2.

Odds > 1일 때

$$\hat{y} = \frac{P}{1-P} \in (0, \infty) \rightarrow \text{Asymmetric}$$

- 목표수정 3.

$$\log \text{ odds} \rightarrow \text{symmetric}$$

$$\hat{y} = \log\left(\frac{P}{1-P}\right) \in (-\infty, \infty)$$

$$- \hat{y} = \log\left(\frac{P}{1-P}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_n x_n$$

$$\Rightarrow P = \frac{\exp(\hat{\beta}_0 + \dots + \hat{\beta}_n x_n)}{1 + \exp(\hat{\beta}_0 + \dots + \hat{\beta}_n x_n)} = \sigma(x|\beta)$$

- 디자이 품질 평가.

- 디자이 품질 평가.

⇒ 출판은 출판인가 예제. likelihood를 이용하여
디자이 평점을 찾기.

⇒ Maximum Likelihood Estimation

$$L(X, y | \beta) = \prod_{i=1}^n P(X_i, y_i | \beta)$$

$$= \prod_{i=1}^n \delta(X_i | \beta)^{y_i} \cdot (1 - \delta(X_i | \beta))^{1-y_i}$$

⇒ Log MLE.

$$\log L(X, y | \beta) = \sum_{i=1}^n y_i \log \delta(X_i | \beta)$$

$$+ (1 - y_i) (1 - \delta(X_i | \beta))$$

⇒ 목표! Maximum !!

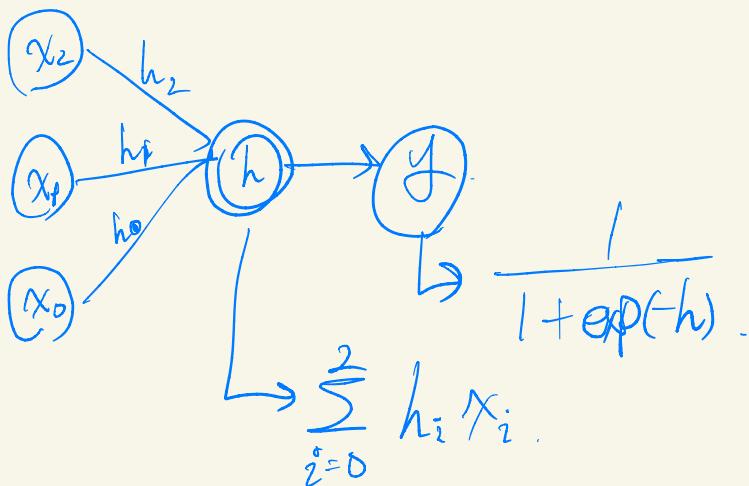


Negative Log MLE. 목표

$$\text{Min} \left[-\frac{1}{n} \sum_{i=1}^n y_i \log \delta(x_i|\beta) + (1-y_i) (1-\delta(x_i|\beta)) \right]$$

↓
해를 찾기 위해!

• Gradient Descent. Algorithm



- Chain Rule을 이용하여 각 설명변수의 coefficient 값이 어떻게 업데이트되는지 방향과 정도를 결정한다.

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \sigma} \cdot \frac{\partial \sigma}{\partial w}$$

$$\frac{\partial \sigma}{\partial w} = \sigma(x)(1-\sigma(x))$$

$$\frac{\partial L}{\partial w_j} = \frac{1}{n} \sum_{i=1}^n [\sigma(x_i|\beta) - y_i] x_{i,j}$$

$$w_{\text{new}} = w_{\text{old}} - \alpha \cdot \frac{\partial L}{\partial w}$$

• 결과 분석.

Linear Regression \rightarrow coefficients.

$x_1 : 1 \rightarrow 2$. 증가.

$\hat{y} : 0 \rightarrow 0 + \hat{\beta}_1$. 증가.

Logistic Regression

Not intuitive

$\hat{\beta}_i$: Positive $\rightarrow e^{\hat{\beta}_i} > 1$,

\Rightarrow odds Ratio $\uparrow \Rightarrow$ $P \uparrow$

$\hat{\beta}_i$: Negative $\rightarrow P \downarrow$.

* Coefficients positive

\Rightarrow Positively correlation, success class.

Classification Performance Evaluation

* Confusion Matrix.

		Prediction	
		P	N
Actual	P	tp	fn
	N	fp	tn

$$* \text{Recall} = \frac{tp}{tp + fn}$$

] 3类评估.

$$* \text{Precision} = \frac{tp}{tp + fp}$$

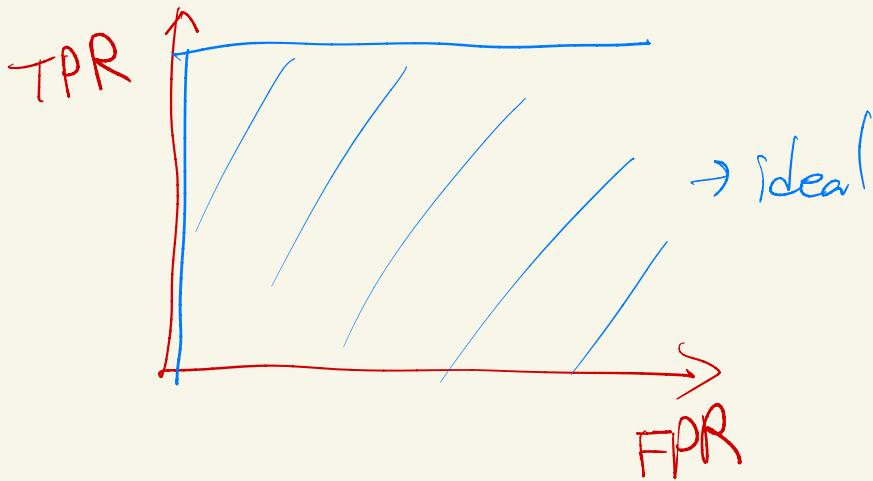
$$* \text{F1-score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

$$* \text{BCR} = \sqrt{\frac{tp}{tp + fn} \times \frac{tn}{tn + fp}}$$

$$* TPR = \frac{tp}{tp + fn} \quad (\text{Recall})$$

$$* FPR = \frac{fp}{tn + fp}$$

* ROC - Curve.



X축: FPR, Y축: TPR

⇒ 가장 좋은 모델은 시가지 좋은 빨간 선을 성능 지표면(TPR)

증가하다가, 100% 순간부터 인종은 블루

성능 지표가 증가하니. (FPR)