

< Trading Off Precision and Recall >

Logistic Regression : $0 < f_{\vec{w},b}(\vec{x}) < 1$

$$\begin{cases} \text{Precision} = \frac{TP}{TP + FP} \\ \text{Recall} = \frac{TP}{TP + FN} \end{cases}$$

output threshold $\begin{cases} \text{Predict 1 if } f_{\vec{w},b}(\vec{x}) \geq 0.5 \\ \text{Predict 0 if } f_{\vec{w},b}(\vec{x}) < 0.5 \end{cases}$

① Suppose we want to predict $y=1$ (rare disease) only if very confident.

(매우 확실한때만 $y=1$ 이라고 결정하고 싶다면)

\Rightarrow set higher threshold $\begin{cases} \text{Predict 1 if } f_{\vec{w},b}(\vec{x}) \geq 0.7 \\ \text{Predict 0 if } f_{\vec{w},b}(\vec{x}) < 0.7 \end{cases} \Rightarrow \text{higher precision lower recall}$

* higher threshold = 실제로 rare disease가 아닌데 (actual class=0) rare disease로 분류하는 비율이 줄어든다는 것 (FP↓)
반대로 실제로 rare disease가 맞는데 (actual class=1) rare disease가 아니라고 분류하는 비율이 늘어난다 (FN↑)

\therefore higher precision lower Recall

② Suppose we want to avoid missing too many case of rare disease (avoid False Negative)

(많은 case의 rare disease를 놓치는 것을 피하고 싶다면 = rare disease 인데 아니라고 판단하는 경우를 줄이고 싶다면)

\Rightarrow set lower threshold $\begin{cases} \text{Predict 1 if } f_{\vec{w},b}(\vec{x}) \geq 0.3 \\ \text{Predict 0 if } f_{\vec{w},b}(\vec{x}) < 0.3 \end{cases} \Rightarrow \text{lower precision higher recall}$

* lower threshold = 실제로 rare disease가 아닌데 (actual class=0) rare disease로 분류하는 비율이 늘어난다는 것 (FP↑)
반대로 실제로 rare disease가 맞는데 (actual class=1) rare disease가 아니라고 판단하는 비율이 줄어든다 (FN↓)

\therefore lower Precision higher Recall

\Rightarrow 일반적으로 모델의 output 이 threshold 보다 클 때 1로 예측함

$f_{\vec{w},b}(\vec{x}) \geq \text{threshold} \Rightarrow \text{Predict 1}$



\Rightarrow 최적의 threshold를 찾는 것이 중요