

3.3 성능 예측

3.3.2 오차행렬

		예측값	
		0	1
실제값	0	TN	FP
	1	FN	TP

정밀도(precision) = $\frac{TP}{TP+FP}$: 양성이라고 판단한 것들 중에 진짜 양성 비율

재현율(recall) = $\frac{TP}{TP+FN}$: 실제 양성인 것들 중에 진짜 양성이라고 판단한 비율

3.3.3 정밀도와 재현율

		예측값	
		0	1
실제값	0	75	5
	1	15	5

$$\text{정확도} = \frac{TN+TP}{TN+FP+FN+TP} = \frac{75+5}{75+5+15+5} = 0.8$$

$$\text{정밀도(precision)} = \frac{TP}{TP+FP} = \frac{5}{5+5} = 0.5$$

$$\text{재현율(recall)} = \frac{TP}{TP+FN} = \frac{5}{5+15} = 0.25$$

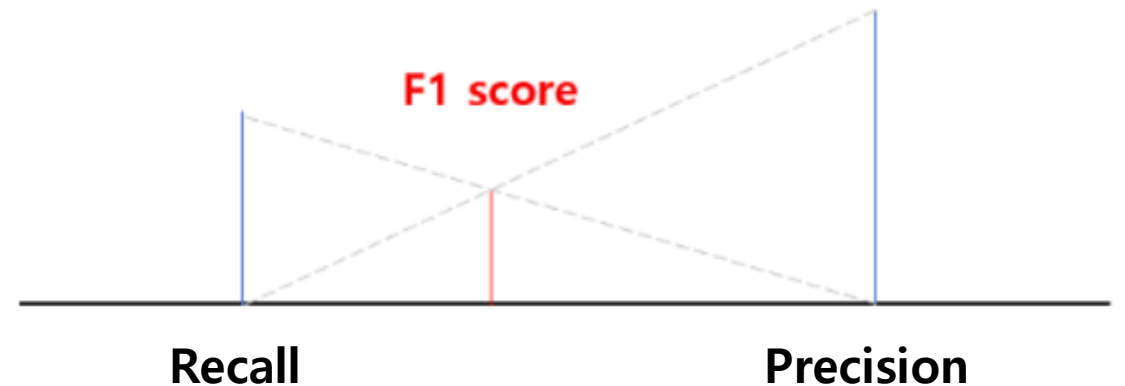
3.3.3 정밀도와 재현율

		예측값	
		0	1
실제값	0	75	5
	1	15	5

$$F_1 = \frac{2}{\frac{1}{\text{정밀도}} + \frac{1}{\text{재현율}}} = \frac{2}{\frac{1}{0.5} + \frac{1}{0.25}} = 0.33$$

○ 조화평균

$$\text{조화평균} = \frac{2}{\frac{1}{a} + \frac{1}{b}} = \frac{a+b}{2ab}$$

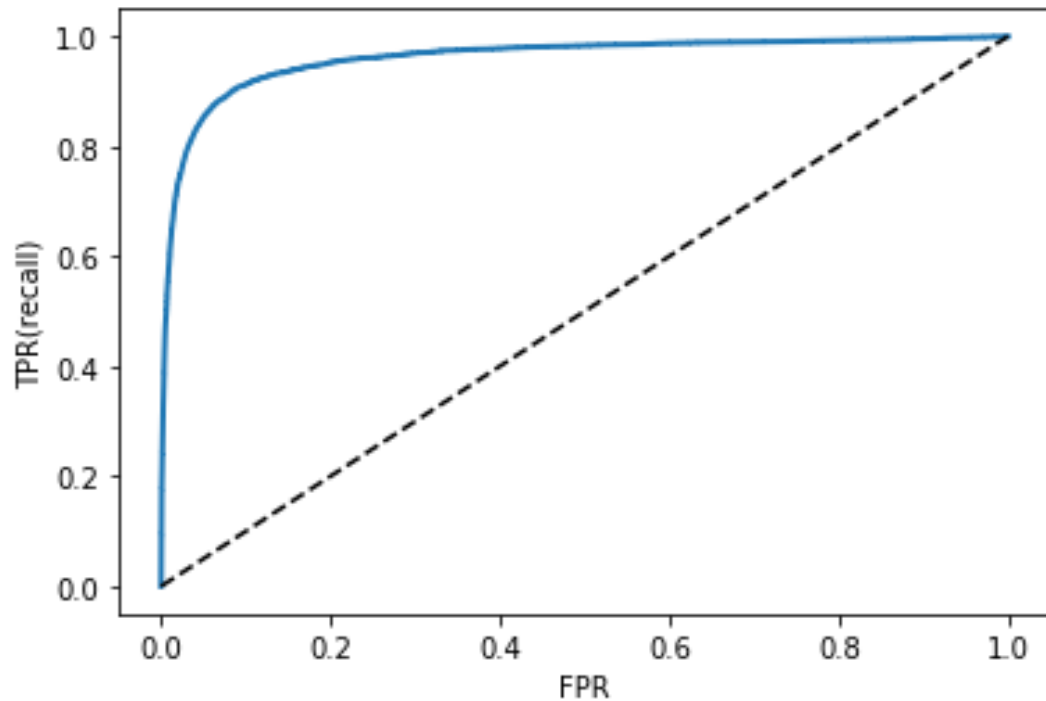


3.3.4 정밀도/재현율 트레이드오프

트레이드오프 현상 : 정밀도 $\uparrow \rightarrow$ 재현율 \downarrow
 정밀도 $\downarrow \rightarrow$ 재현율 \uparrow

정밀도가 재현율보다 중요한 경우	재현율이 정밀도보다 중요한 경우
<ul style="list-style-type: none">- 양성을 음성으로 분류(FN)하면 손실이 적을 경우- 음성을 양성으로 분류(FP)하면 손실이 클 경우- PR곡선 사용	<ul style="list-style-type: none">- 양성을 음성으로 분류(FN)하면 손실이 클 경우- 음성을 양성으로 분류(FP)하면 손실이 적을 경우- ROC곡선 사용

3.3.5 ROC곡선



$$\text{FNR(특이도)} = \frac{TN}{TN+FP}$$

실제 음성인 것들 중에 진짜 양성이라고
판단한 비율

* 카피 통계량

$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

Pr(a) : observed accuracy

Pr(e) : expected accuracy

		예측값		합계
		0	1	
실제값	0	75	5	80
	1	15	5	20
합계		90	10	

예측) 양성 확률 = 0.1/ 음성 확률 = 0.9

실제) 양성 확률 = 0.2/ 음성 확률 = 0.8

실제와 예측 모두 양성일 확률 = $0.1 \times 0.2 = 0.02$

실제와 예측 모두 음성일 확률 = $0.9 \times 0.8 = 0.72$

확률적으로 우연히 일치할 확률 = $0.02 + 0.72 = 0.74$

* 카피 통계량

$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

Pr(a) : observed accuracy

Pr(e) : expected accuracy

$$K = \frac{0.8 - 0.74}{1 - 0.74} = 0.23$$