

分類において良し悪しを評価するには？

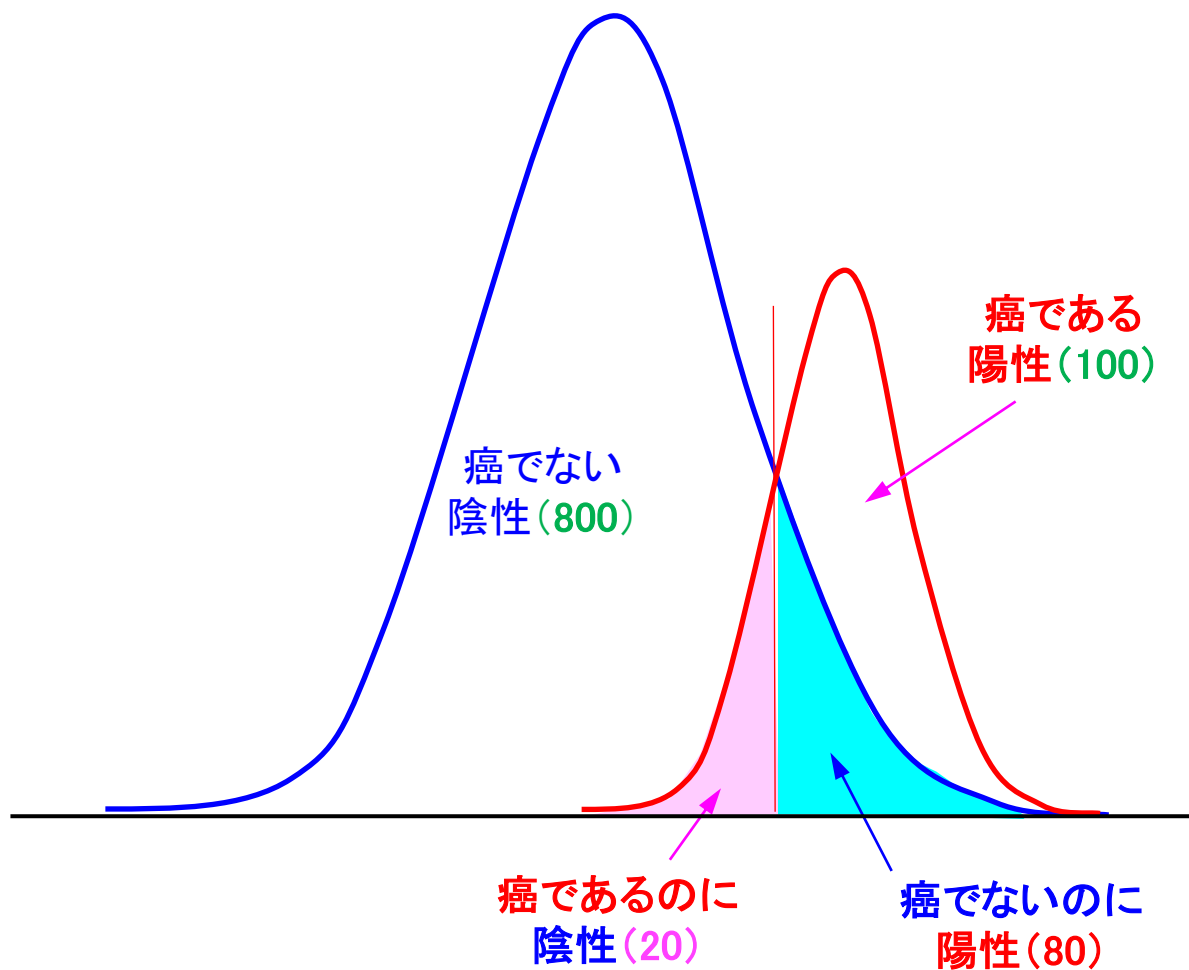
		予測データ	
		異常	正常
実際のデータ	異常	a	b
	正常	c	d

例		予測データ	
		異常	正常
実際のデータ	異常	100	20
	正常	80	800



評価指標	算出式	意味	例
正解数(Accuracy)	$\frac{a + d}{a + b + c + d}$	全てのサンプルで実測と予測値が一致した割合	$\frac{100 + 800}{100 + 20 + 80 + 800} = \frac{900}{1000} = 90\%$
適合率(Precision)	$\frac{a}{a + c}$	適合率と再現率はトレードオフの関係	$\frac{100}{100 + 80} = \frac{100}{180} = 56\%$
再現率(Recall)	$\frac{a}{a + b}$		$\frac{100}{100 + 20} = \frac{100}{120} = 83\%$
F値(F-measure)	$\frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}}$	適合率と再現率の調和平均	$\frac{2}{\frac{1}{56} + \frac{1}{83}} = \frac{900}{1000} = 67\%$

例		診断	
		陽性	陰性
実際	癌	100	20
	正常	80	800



評価指標	例
正解数(Accuracy)	$\frac{100 + 800}{100 + 20 + 80 + 800} = \frac{900}{1000} = 90\%$
適合率(Precision)	$\frac{100}{100 + 80} = \frac{100}{180} = 56\%$
再現率(Recall)	$\frac{100}{100 + 20} = \frac{100}{120} = 83\%$
F値(F-measure)	$\frac{2}{\frac{1}{56} + \frac{1}{83}} = \frac{900}{1000} = 67\%$