

7.6

1

この分類器は、糖尿病である(label=0)事例のうち、糖尿病であると正しく予測することが求められ、間違ってもいいが取りこぼしが許されない状況であるという点で、再現性を高くすることが重要である。正解度は、分類したい結果の重要度が同じである場合に分類器の性能評価に利用するが、この場合、糖尿病であるかないかというラベルの重要性が異なるため、性能評価に用いるのは適切ではない。そのため、正解度という指標のみではこの二値分類器は糖尿病の予測に役に立つと判断することができない。

2

精度：

$$39/(15+39)=0.722\dots$$

再現率：

$$39/(28+39)=0.582\dots$$

3

0.3

4

人口 1.2億 のうち 0.001% \Rightarrow 1200 人 感染			
・ 感度 70%	$\frac{\text{予測正陽性}}{\text{真に陽性}} = 0.7$		
実際に感染検出される人数は			
	$1200 \times 0.7 = 840$	人	
	$1200 - 840 = 360$	人が感染しているのに検出されない	
・ 特異度	真に陰性 \rightarrow 陰性	99.9%	
	真に陰性 \rightarrow 陽性	0.1%	
感染していない人数			
	$1.2\text{億} \times 0.99999$	\Rightarrow 119,998,800	人
このうち、陰性であるのに陽性だと誤って検査される人数は			
	$119,998,800 \times 0.001$	$=$ 119,998.8	人

以上のことから、次のような問題点が生じると考えられる；

- 実際には感染しているにもかかわらず、検査の結果が誤って陰性であった 360 人が日常生活を送ることになり、さらなる感染拡大が懸念される
- 実際には感染していないにもかかわらず、検査の結果が誤って陽性であった約 10 万人もの人が入院することになり、病床の確保が必要になり、また強制隔離による人権侵害などの問題が生じる。