

모델평가지표

가. 다음의 혼동행렬을 보고 문제 1~2번에 답하시오

		실제(Actual)	
		긍정(Positive)	부정(Negative)
예측	긍정	TP	(1)
	부정	(2)	TN

문제 1. (1)에 들어갈 알맞은 기호는?

① FN

② FP

③ TP

④ TN

문제 2. (2)에 들어갈 알맞은 기호는?

① FN

② FP

③ TP

④ TN

나. 다음의 혼동행렬을 보고 문제 3 ~ 4번에 답하시오

		예측(Predict)	
		긍정(Positive)	부정(Negative)
실제	긍정	TP	FN
	부정	FP	TN

문제 3. 민감도(=재현율)를 구하는 수식을 작성하시오.

$$\text{민감도(Sensitivity)} = \frac{TP}{(\quad) + (\quad)}$$

문제 4. 특이도(Specificity)를 구하는 수식을 작성하시오.

$$\text{특이도(Specificity)} = \frac{TN}{(\quad) + (\quad)}$$

다. 다음의 혼동행렬을 보고 문제 5~6번에 답하시오

		예측(Predict)	
		긍정(Positive)	부정(Negative)
실제	긍정	10	2
	부정	4	7

문제 5. 정확도, 정밀도, 재현율을 구하시오.(답은 분수로 표현)

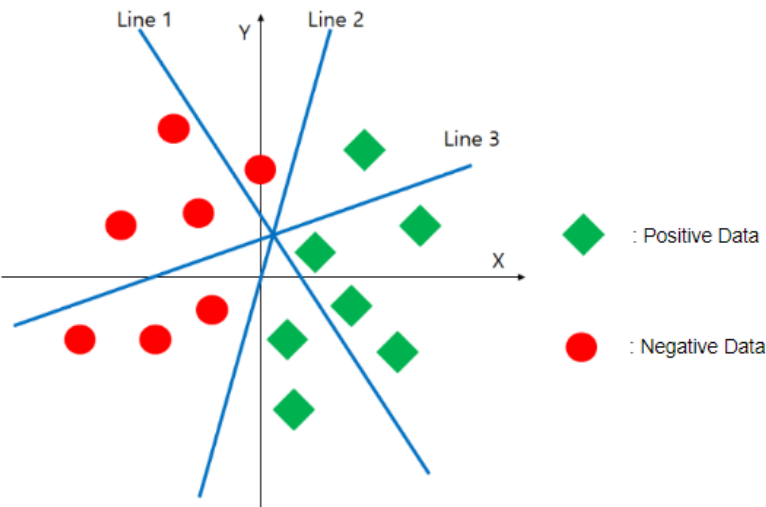
문제 6. 민감도, 특이도를 구하시오.(답은 분수로 표현)

라. 다음의 혼동행렬을 보고 문제 7번에 답하시오.

		실제	
		긍정(Positive)	부정(Negative)
예측	긍정	12	5
	부정	6	10

문제 7. 정확도, 정밀도, 재현율, 민감도, 특이도를 구하시오(답은 분수로 표현)

마. 다음의 그림은 SVM 알고리즘을 이용한 데이터 분류를 3번 수행한 결과이다. line1, line2, line3이 각각 SVM 모델이라고 할때 문제 8~10번에 답하시오. 단, 각 모델이 분류한 긍정 및 부정의 영역은 분류결과 내 해당 데이터가 많은 영역으로 결정된다고 가정한다.(예, line1 모델의 경우 선 아래 영역이 부정, 선 위의 영역이 긍정으로 결정)



문제 8. 세 가지 모델 중 정확도가 가장 높은 모델은?

- ① line1
- ② line2
- ③ line3

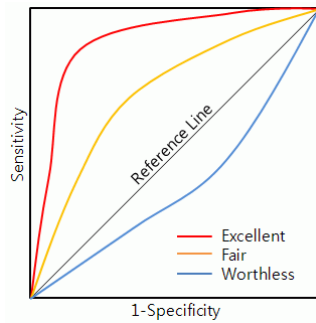
문제 9. 세 가지 모델 중 정밀도가 가장 낮은 모델은?

- ① line1
- ② line2
- ③ line3

문제 10. 세 가지 모델 중 재현율(=민감도)이 가장 낮은 모델은?

- ① line1
- ② line2
- ③ line3

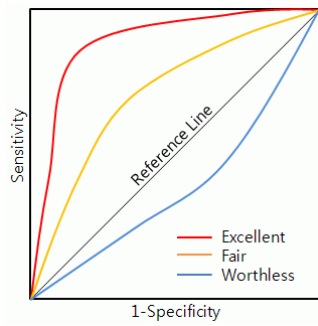
문제 11. 분류모형 성과 분석 중 ROC Curve는 x축에 1-특이도, y축에 민감도를 나타낸다. 아래와 같은 오분류표가 있을때 ROC 그래프의 y축을 표현한 것으로 가장 적절한 것은?



		예측(Predict)	
		긍정(Positive)	부정(Negative)
실제	긍정	TP	FN
	부정	FP	TN

- ① $FP / (FP + TN)$
- ② $TN / (FP + TN)$
- ③ $TP / (TP + FP)$
- ④ $TP / (TP + FN)$

문제 12. 분류모형 성과 분석 중 ROC Curve는 x축에 1-특이도, y축에 민감도를 나타낸다. 아래와 같은 오분류표가 있을때 ROC 그래프의 x축을 표현한 것으로 가장 적절한 것은?



		예측(Predict)	
		긍정(Positive)	부정(Negative)
실제	긍정	TP	FN
	부정	FP	TN

- ① $FP / (FP + TN)$
- ② $TN / (FP + TN)$
- ③ $TP / (TP + FP)$
- ④ $TP / (TP + FN)$

문제 13. 다음 중 아래 오분류표를 이용하여 구한 F1 값은 얼마인가?

		예측(Predict)		합계
		긍정(Positive)	부정(Negative)	
실제	긍정	40	60	100
	부정	60	40	100
합계		100	100	200

- ① 0.15
- ② 0.3
- ③ 0.4
- ④ 0.55

문제 14. 다음 중 ROC Curve의 x축에 1-특이도와 같은 의미를 가지는 것은?

- ① TPR
- ② Recall
- ③ FPR
- ④ Precision

문제 15. ROC 커브의 밑부분 면적을 의미하는 용어로 이 값이 클수록 모형의 성능이 높아짐을 의미하는 것은?

답안

1. 2
2. 1
3. TP, FN
4. TN, FP
5. 정확도 = $17/23$, 정밀도 = $5/7$, 재현율 = $5/6$
6. 민감도 = $5/6$, 특이도 = $7/11$
7. 정확도 = $2/11$, 정밀도 = $12/17$, 재현율 = 민감도 = $2/3$, 특이도 = $2/3$
8. 2
9. 3
10. 1
11. 4
12. 1
13. 3
14. 3