

01. - (3)

①

$PZ(\text{Odds})$

: 성공한 확률이 실패한 확률의 몇 배인지.

$\Rightarrow 0$

②

로지스틱 회귀분석의 종속변수는 범주형일 때 사용 가능.

$\Rightarrow 0$

③

$$\frac{P}{1-P} = e^{\alpha + \beta_1 X_1 + \dots + \beta_k X_k} \text{ 이고}$$

$$\log\left(\frac{P}{1-P}\right) = \alpha + \beta_1 X_1 + \dots + \beta_k X_k \text{ 이므로}$$

회귀계수가 5일 때

$$\log\left(\frac{P}{1-P}\right) = 5 \iff e^5 = \frac{P}{1-P} = (\text{odds})$$

$\therefore e^5$ 이다.

$\Rightarrow X$

④

$$\frac{1}{1} = 1$$

02. - ④

④

종속변수가 범주형인 경우

→ 카이제곱 통계량, 지니 지수, 엔트로피 지수

.. 연속형인 경우

→ ANOVA F-통계량, 불순감소량 사용

03.

'중기 계획'에 대한 설명

04. - ①

지니 지수

$$= \sum_{i \neq j} p_i \quad , \quad p_i = 1 - \sum_{i=1}^n p_i^2 \text{ 이므로}$$

$$1 - \left\{ \left(\frac{3}{5} \right)^2 + \left(\frac{2}{5} \right)^2 \right\}$$

$$= 1 - \left(\frac{9}{25} + \frac{4}{25} \right) = 1 - \frac{13}{25} = \frac{12}{25}$$

\Rightarrow ①

05. - ③

① 배깅 (Bootstrap Aggregating)

: 여러 개의 부스트랩 집계

\Rightarrow O

② 부스트랩

: 원본 데이터와 같은 크기의 표본을 랜덤 복원 추출한
샘플 데이터.

\Rightarrow O

③ 랜덤 포레스트

: 큰 분산 갖는 의사결정나무의 단점 보완

\rightarrow 분산 감소

모든 분류기들이 높은 비상관성.

이상값에 민감 X

\Rightarrow X

④ 앙상블 분석

: 모형의 예측력 높이고자

여러 개의 분류 모형에 의한

결과 종합하여 최종 결과.

\Rightarrow O

06. - ③

부스팅

: 이전 분류기에 비해 잘못 분류된 data에
더 큰 가중치

∴ 약한 모델들 결합

⇒ ③

07. - ①

① 은닉층, 은닉노드의 수가 너무 많으면
과적합 발생 가능.

⇒ X

08. - ④

ROC 커브

· x축 : FPR (1-특이도)

· y축 : TPR (민감도)

· AUROC 가 1에 가까울수록 good

⇒ ④

09. - 0.7

$$F1 \text{ Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad 0.2$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{15}{15+10} = \frac{15}{25} = \frac{3}{5}$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{15}{15+5} = \frac{15}{20} = \frac{3}{4} \quad 0.25$$

$$\therefore \frac{2 \times \frac{3}{5} \times \frac{3}{4}}{\frac{3}{5} + \frac{3}{4}} = \frac{\frac{18}{20}}{\frac{12+15}{20}} = \frac{18}{27} = \frac{2}{3}$$

$$= 0.666...$$

$$\approx 0.7$$