TITANIC\_SURVIVAL.MWX

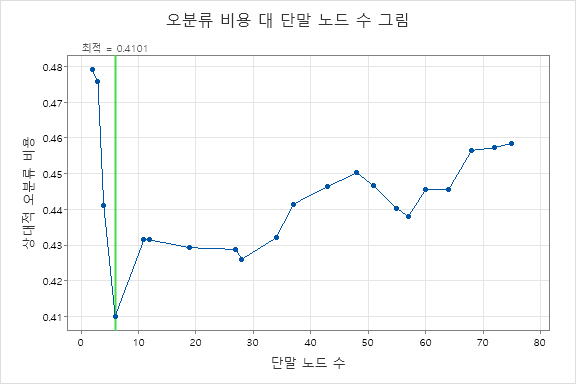
**6 노드 CART® 분류: Status 대 Age, SibSp, ParCh, Fare, Gender, Pclass, Embarked**

**방법**

|  |  |
| --- | --- |
| 사전 확률 | 모든 등급에 대해 같음 |
| 노드 분할 | 지니 |
| 최적 트리 | 최소 오분류 비용의 1 표준 오차 이내 |
| 모형 검증 | 10-접기 교차 검증 |
| 사용된 행 | 1309 |

**이항 반응 정보**

|  |  |  |  |
| --- | --- | --- | --- |
| **변수** | **등급** | **카운트** | **%** |
| Status | S (사건) | 500 | 38.20 |
|  | D | 809 | 61.80 |
|  | 모두 | 1309 | 100.00 |



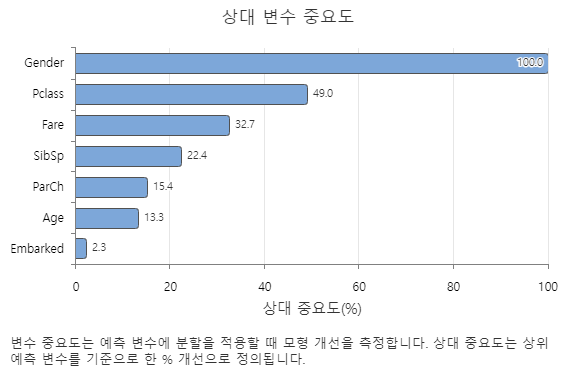
**최적 트리 다이어그램**

텍스트, 도표, 평면도, 라인이(가) 표시된 사진

자동 생성된 설명

**모형 요약**

|  |  |
| --- | --- |
| 전체 예측 변수 | 7 |
| 중요 예측 변수 | 7 |
| 단말 노드 수 | 6 |
| 최소 단말 노드 크기 | 20 |
| **통계량** | **교육** | | **검정** |
| 평균 음수 로그 우도 | 0.4392 | | 0.5065 |
| ROC 곡선 아래 면적 | 0.8264 | | 0.8129 |
| 95% CI | (0.3979, 1) | | (0.7866, 0.8392) |
| 향상도 | 2.4400 | | 2.3843 |
| 오분류 비용 | 0.4036 | | 0.4101 |

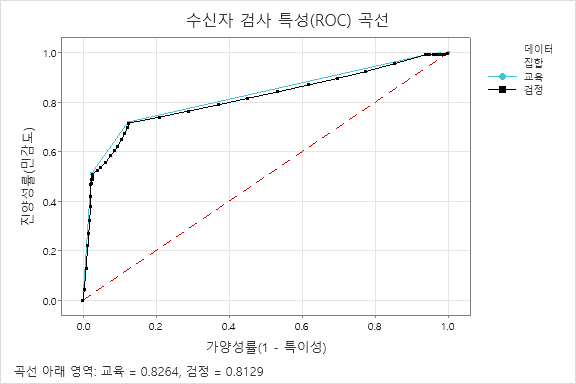


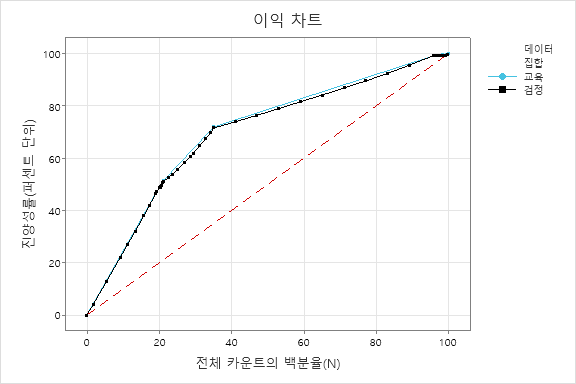
**오차 행렬**

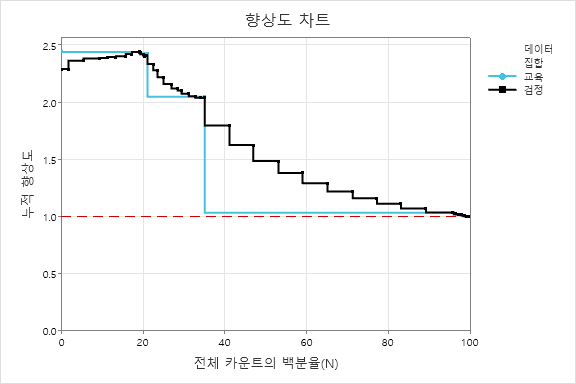
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **예측된 등급(교육)** | | | | | **예측된 등급(검정)** | | | |
| **실제 등급** | **카운트** | **S** | **D** | | **정답률(%)** | | **S** | | **D** | **정답률(%)** |
| S (사건) | 500 | 360 | 140 | | 72.0 | | 358 | | 142 | 71.6 |
| D | 809 | 100 | 709 | | 87.6 | | 102 | | 707 | 87.4 |
| 모두 | 1309 | 460 | 849 | | 81.7 | | 460 | | 849 | 81.4 |
| **통계량** | | | | **교육(%)** | | **검정(%)** | |
| 진양성률(민감도 또는 검정력) | | | | 72.0 | | 71.6 | |
| 가양성률(유형 I 오차) | | | | 12.4 | | 12.6 | |
| 가음성률(유형 II 오차) | | | | 28.0 | | 28.4 | |
| 진음성률(특이성) | | | | 87.6 | | 87.4 | |

**오분류**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **입력 오분류 비용** | | **예측된 등급** | | | |
| 실제 등급 | | S | | D | |
| S | |  | | 1.00 | |
| D | | 1.00 | |  | |
|  |  | | **교육** | | | | | **검정** | | |
| **실제 등급** | **카운트** | | **오분류됨** | | **오차율(%)** | | **비용** | **오분류됨** | **오차율(%)** | **비용** |
| S (사건) | 500 | | 140 | | 28.0 | | 0.2800 | 142 | 28.4 | 0.2840 |
| D | 809 | | 100 | | 12.4 | | 0.1236 | 102 | 12.6 | 0.1261 |
| 모두 | 1309 | | 240 | | 18.3 | | 0.2018 | 244 | 18.6 | 0.2050 |







**자료 분석**

최적의 노드(6개)로 모델을 적용한 결과 5개의 그룹으로 분할 되었으며 그 중 가장 생존율이 높은 그룹은 93.2%의 생존율을 보였으며 여성이면서, PClass (좌석 등급)가 1, 2 인 그룹이었다.

반대로 생존율이 가장 낮은 그룹은 생존율이 5%에 지나지 않았으며, 남성이면서 9.5세 이하이면서 형제, 자매 수가 2.5 이상인 그룹이었다.

생존에 가장 중요한 변수는 성별이었으며, 2번째 높은 변수인 PClass의 두배가 넘는 중요도를 보였다.