IBM SPSS Decision Trees 26

참고

이 정보와 이 정보가 지원하는 제품을 사용하기 전에, 21 페이지의 『주의사항』에 있는 정보를 확인하십시오.

제품 정보

이 개정판은 새 개정판에서 별도로 명시하지 않는 한, IBM SPSS Statistics의 버전 26, 릴리스 0. 수정 0 및 모든 후속 릴리스와 수정에 적용됩니다.

목차

의사결정나무

다음 의사결정나무 기능은 SPSS® Statistics Professional Edition 또는 의사결정나무 옵션에 포함되어 있습니다.

의사결정나무 작성

의사결정나무 프로시저는 나무 기반의 분류 모형을 작성합니다. 독립(예측자) 변수값을 기반으로 종속(대상) 변수값을 예측하거나 케이스를 집단으로 분류합니다. 이 프로시저에서는 탐색 및 확인 분류 분석을 위한 검증 도구를 제공합니다.

다음에 대해 프로시저를 사용할 수 있습니다.

분석 방식. 특정 그룹의 구성원일 가능성이 큰 사람을 식별합니다.

층화. 케이스를 높은 위험 그룹, 중간 위험 그룹, 낮은 위험 그룹과 같은 여러 범주 중 하나에 할당합니다.

예측. 규칙을 생성하고 이러한 규칙을 사용하여 향후 이벤트(예: 누군가 대출금에 대한 채무를 이행하지 않을 우도 또는 차량이나 주택의 잠재적 재판매 가치)를 예측합니다.

데이터 축소 및 변수 선별. 큰 변수 집합에서 공식 모수 모델를 작성하는 데 사용할 유용한 예측자 부분변수를 선택합니다.

상호작용 식별. 특정 부집단에만 관련되어 있는 관계를 식별하여 공식 모수 모델에 지정합니다.

범주 합치기 및 연속형 변수 이산화. 정보 손실을 최소화하면서 집단 예측자 범주 및 연속형 변수를 코딩변경합니다.

예제. 은행에서는 신용 대출자가 적정 신용 위험도를 나타내는지 여부에 따라 범주화하려고 합니다. 과거 고객의 알려진 신용등급을 포함하여 다양한 요인을 기준으로 모형을 작성하여 미래 고객의 채무 이행 가능성을 예측할 수 있습니다.

나무 기반 분석은 다음과 같은 몇 가지 훌륭한 기능을 제공합니다.

- 위험도가 높거나 낮은 동질적 집단을 식별할 수 있습니다.
- 개별 케이스에 대한 예측을 작성하기 위한 규칙을 쉽게 구성할 수 있습니다.

데이터 고려사항

데이터. 종속변수 및 독립변수는 다음으로 사용될 수 있습니다.

- 명목*(Nominal)*. 변수의 값이 고유한 순위가 없는 범주를 나타내는 경우 해당 변수는 명목으로 취급될 수 있습니다. 예를 들어, 직원이 근무하는 회사의 부서가 있습니다. 종교, 우편번호 또는 종교 단체 등이 명목변수에 해당합니다.
- 순서*(Ordinal)*. 변수의 값이 고유한 순위가 있는 범주를 나타내는 경우 해당 변수는 순서로 취급될 수 있습니다. 예를 들어, 매우 불만족에서 매우 만족에 이르는 서비스 만족도 수준이 있습니다. 순서변수의 예로는 만족도나 신뢰도를 나타내는 태도 스코어 및 선호도 등급 스코어가 있습니다.
- 척도*(Scale)*. 해당 값이 의미 있는 메트릭으로 순서가 지정된 범주를 나타내므로 값 간 거리 비교가 적합한 경우 해당 변수는 척도(연속형)로 처리할 수 있습니다. 척도변수의 예로는 연령과 수입이 있습니다.

빈도 가중값 가중값을 적용하는 경우 분수 가중값은 가장 가까운 정수로 반올림되기 때문에 가중값이 0.5보다 작은 케이스는 가중값으로 0이 할당되어 분석에서 제외됩니다.

가정. 이 프로시저에서는 모든 분석 변수에 적절한 측정 수준이 할당되었다고 가정하며 일부 기능에서는 분석에 포함된 모든 종속변수 값에 값 레이블을 정의되었다고 가정합니다.

- **측정 수준.** 측정 수준은 나무 계산에 영향을 미치므로 모든 변수에 적절한 측정 수준이 할당되어야 합니다. 기본적으로 숫자변수는 척도로, 문자변수는 명목형으로 계산된다고 가정하여 참 측정 수준을 정확하게 반영하지 않을 수도 있습니다. 변수 목록의 각 변수 옆에 있는 아이콘은 변수 유형을 식별합니다.

소스 변수 목록에서 변수를 마우스 오른쪽 단추로 클릭하고 팝업 메뉴에서 측정 수준을 선택하여 변수에 대한 측정 수준을 임시로 변경할 수 있습니다.

- **값 레이블.** 이 프로시저의 대화 상자 인터페이스에서는 범주형(명목형, 순서형) 종속변수의 결측되지 않은 값 모두에 값 레이블이 정의되거나 있거나 모두에 값 레이블이 정의되지 않았다고 가정합니다. 범주형 종속변수의 결측되지 않은 값 중 최소 2개에 값 레이블이 없는 경우 일부 기능을 사용할 수 없습니다. 2개 이상의 결측되지 않은 값에 값 레이블이 정의되어 있는 경우 값 레이블이 없는 다른 값이 포함된 케이스를 분석에서 제외합니다.

의사결정나무를 구하는 방법

1. 메뉴에서 다음을 선택합니다.

   분석 > **분류** > **나무...**
2. 하나의 종속변수를 선택합니다.
3. 하나 이상의 독립변수를 선택합니다.
4. 성장방법을 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- 소스 목록의 변수 측정 수준을 변경합니다.
- 독립변수 목록의 첫 번째 변수를 첫 번째 분할변수로 모형에 사용합니다.

- 케이스가 나무 성장 과정에 미치는 영향력 정도를 정의하는 영향 변수를 선택합니다. 영향력 값이 낮은 케이스는 영향력이 적고 값이 높은 케이스는 영향력이 큽니다. 영향 변수 값은 양수여야 합니다.
- 나무를 검증합니다.
- 나무 성장 기준을 사용자 정의합니다.
- 터미널 노드 수, 예측값 및 예측 확률을 변수로 저장합니다.
- 모형을 XML(PMML) 형식으로 저장합니다.

측정 수준을 알 수 없는 필드

측정 수준 경보는 데이터 세트에서 하나 이상의 변수(필드)에 대해 측정 수준을 알 수 없을 때 표시됩니다. 측정 수준은 이 프로시저의 결과 계산에 영향을 미치기 때문에 모든 변수에 정의된 측정 수준이 있어야 합니다.

데이터 스캔
: 활성 데이터 세트의 데이터를 읽고 현재 알 수 없는 측정 수준이 있는 필드에 기본 측정 수준을 할당합니다. 데이터 세트가 큰 경우 다소 시간이 소요될 수 있습니다.

수동으로 할당
: 알 수 없는 측정 수준이 있는 필드를 모두 나열합니다. 해당 필드에 측정 수준을 할당할 수 있습니다. 데이터 편집기의 변수 목록 분할창에서도 측정 수준을 할당할 수 있습니다.

측정 수준이 이 프로시저에서 중요한 요소이므로 모든 필드에 대해 측정 수준이 정의될 때까지 이 프로시저를 실행할 수 없습니다.

측정 수준 변경

1. 소스 목록에서 변수를 마우스 오른쪽 단추로 클릭합니다.
2. 팝업 메뉴에서 측정 수준을 선택합니다.

이렇게 하면 의사결정나무 프로시저에서 사용하도록 측정 수준이 일시적으로 변경됩니다.

성장방법

사용가능한 성장방법은 다음과 같습니다.

CHAID
: CHAID(Chi-squared Automatic Interaction Detection) 알고리즘입니다. 각 단계에서 CHAID는 종속변수와의 상호작용이 가장 강한 독립변수(예측자)를 선택합니다. 각 예측자의 범주는 종속변수와 크게 차이 나지 않는 한 합쳐집니다.

Exhaustive CHAID
: 각 예측자에 대해 가능한 모든 분할을 검사하는 CHAID 알고리즘을 수정한 것입니다.

CRT 분류 및 회귀분석 나무(Classification and Regression Trees)입니다. CRT는 종속변수와 가능한 동일한 세그먼트로 데이터를 분할합니다. 모든 케이스의 종속변수 값이 동일한 터미널 노드는 동일한 "순수" 노드입니다.

QUEST

신속하고 비편향적이며 효율적인 통계분석 나무입니다. 여러 범주가 있는 예측자 편에서 다른 방법의 편향성을 방지하는 신속한 방법입니다. 종속변수가 명목일 경우에만 QUEST를 지정할 수 있습니다.

각 방법의 장점 및 제한사항은 다음과 같습니다.

표 *1.* 성장방법의 기능.

기능	CHAID*	CRT	QUEST
카이제곱 기반**	X		
서로게이트 독립변수(예측자 변수)		X	X
나무 잘라내기		X	X
다원 노드 분할	X		
이분형 노드 분할		X	X
영향 변수	X	X	
사전확률		X	X
오분류 비용	X	X	X
신속 계산	X		X

*Exhaustive CHAID를 포함합니다.

**QUEST 또한 명목형 독립변수에 대해 카이제곱 측도를 사용합니다.

범주 선택

범주형(명목형, 순서형) 종속변수의 경우 다음을 수행할 수 있습니다.

- 분석에 포함되는 범주를 제어합니다.
- 관심 있는 대상 범주를 식별합니다.

범주 포함/제외

분석을 종속변수의 특정 범주로 제한할 수 있습니다.

- 제외 목록의 종속변수 값이 포함되어 있는 케이스는 분석에 포함되지 않습니다.
- 명목 종속변수의 경우 사용자 결측 범주도 분석에 포함시킬 수 있습니다. 기본적으로 사용자 결측 범주는 제외 목록에 표시됩니다.

목표 범주

선택된 범주가 분석에서 가장 관심있는 범주로 처리됩니다. 예를 들어, 채무 불이행 가능성이 가장 높은 개인을 식별하는 것이 주된 관심사인 경우 "불량" 신용 등급 범주를 대상 범주로 선택할 수 있습니다.

- 기본 대상 범주는 없습니다. 범주가 선택되지 않으면 일부 분류 규칙 옵션 및 이득 관련 결과를 사용할 수 없습니다.
- 다중 범주를 선택하는 경우 각 대상 범주에 대해 별도의 이득 표 및 도표가 생성됩니다.
- 한 개 이상의 범주를 대상 범주로 지정해도 나무 모형, 위험도 추정값 또는 오분류 결과에 영향을 미치지 않습니다.

범주

이 대화 상자를 사용하려면 종속변수의 값 레이블이 정의되어 있어야 합니다. 두 개 이상의 범주형 종속변수 값에 값 레이블이 정의되어 있지 않은 경우 이 대화 상자를 사용할 수 없습니다.

범주를 포함/제외하고 목표 범주를 선택하는 방법

1. 기본 의사결정나무 대화 상자에서 두 개 이상의 값 레이블이 정의된 범주형(명목형, 순서형) 종속변수를 선택합니다.
2. **범주**를 클릭합니다.

검증

검증을 사용하여 더 큰 모집단에 대해 나무 구조 일반화가 적합한지 평가할 수 있습니다. 교차 검증 및 분할 표본 검증의 두 가지 검증 방법을 사용할 수 있습니다.

교차 검증

교차 검증은 표본을 다수의 부표본 또는 **중첩**으로 나눕니다. 그런 다음 나무 모형을 생성하고 각 부표본에서 차례대로 데이터를 제외합니다. 첫 번째 나무는 첫 번째 표본 중첩의 케이스를 제외한 모든 케이스를 기준으로 하며, 두 번째 나무는 두 번째 표본 중첩의 케이스를 제외한 모든 케이스를 기준으로 하는 방식입니다. 나무 생성에서 제외된 부표본에 각 나무를 적용하여 해당 나무의 오분류 위험도를 추정합니다.

중요사항: 잘라내기가 선택되면 CRT 및 QUEST 방법에 교차 검증을 사용할 수 없습니다.

- 최대 25개의 표본 중첩을 지정할 수 있습니다. 값이 높을수록 각 나무 모형에서 제외되는 케이스 수가 적어집니다.
- 교차 검증은 하나의 최종 나무 모형을 생성합니다. 최종 나무의 교차 검증 위험도 추정값은 모든 나무의 위험도 평균으로 계산됩니다.

분할 표본 검증

분할 표본 검증으로 학습 표본을 사용하여 모형을 생성하고 검정용 표본으로 검정합니다.

- 전체 표본 결과에 대한 퍼센트로 표시되는 학습 표본 결과를 지정하거나 표본을 학습 및 검정 표본으로 분할하는 변수를 지정할 수 있습니다.
- 변수를 사용하여 학습 및 검정 표본을 정의하는 경우 변수의 값이 1인 케이스가 학습 표본에 할당되고 다른 모든 케이스가 검정 표본에 할당됩니다. 해당 변수는 종속변수, 가중변수, 영향 변수 또는 강제 독립변수가 될 수 없습니다.
- 학습 및 검정 표본 모두에 대해 또는 학습 표본에 대해서만 결과를 표시할 수 있습니다.
- 작은 데이터 파일(케이스 수가 적은 데이터 파일)에서는 분할 표본 검증을 주의하여 사용해야 합니다. 일부 범주에 케이스가 충분하지 않아 나무를 적절하게 성장시키지 못할 수 있기 때문에 학습 표본 결과가 작은 경우 잘못된 모형을 산출할 수 있습니다.

의사결정나무 검증

1. 기본 의사결정나무 대화 상자에서 **검증**을 클릭합니다.
2. **교차 검증** 또는 **분할 표본 검증**을 선택합니다.

참고: 두 검증 방법 모두가 표본 집단에 임의로 케이스를 할당합니다. 후속 분석에서 정확히 동일한 결과를 재연할 수 있으려면 처음으로 분석을 실행하기 전에 난수 시작값(변환 메뉴, 난수 생성기)을 설정한 후 후속 분석을 위해 난수 시작값을 해당 값으로 재설정해야 합니다.

나무 성장 기준

사용가능한 성장 기준은 성장방법, 종속변수의 측정 수준 또는 이 둘의 조합에 따라 달라질 수 있습니다.

확장 한계

확장 한계 대화 상자를 사용하면 나무의 수준 수를 제한하고 부모 및 자식 노드의 최소 케이스 수를 제어할 수 있습니다.

최대 나무 깊이

루트 노드 아래 성장의 최대 수준 수를 제어합니다. **자동** 설정은 CHAID 및 Exhaustive CHAID 방법의 경우 루트 노드 아래 세 수준으로, CRT 및 QUEST 방법의 경우 다섯 수준으로 나무를 제한합니다.

최소 케이스 수

노드의 최소 케이스 수를 제어합니다. 이 기준을 충족하지 않는 노드는 분할되지 않습니다.

- 최소값이 증가할수록 노드 수가 적은 나무를 생성합니다.
- 최소값이 감소할수록 노드 수가 많은 나무를 생성합니다.

케이스 수가 작은 데이터 파일에서 부모 노드 및 자식 노드의 케이스에 대해 각각 기본값인 100과 50을 유지하는 경우 루트 노드 아래 노드가 없는 나무를 생성할 수 있습니다. 이러한 경우 최소값을 낮추면 더 유용한 결과를 생성할 수도 있습니다.

확장 한계를 지정하는 방법

1. 기본 의사결정나무 대화 상자에서 **확장 한계**를 클릭합니다.

CHAID 기준

CHAID 및 Exhaustive CHAID 방법에서는 다음을 제어할 수 있습니다.

유의수준

노드 분할과 범주 합치기에 대한 유의확률을 제어할 수 있습니다. 두 기준 모두 기본 유의수준은 0.05입니다.

노드 분할

값이 0보다 크고 1보다 작아야 합니다. 값이 낮으면 노드 수가 적은 나무를 생성합니다.

범주 합치기

값이 0보다 크고 1보다 작거나 같아야 합니다. 범주 합치기를 방지하려면 값을 1로 지정합니다. 척도 독립변수의 경우, 최종 나무의 변수 범주 수가 지정된 구간 수(기본값은 10)임을 의미합니다. 자세한 정보는 8 페이지의 『CHAID 분석의 척도 구간』 주제를 참조하십시오.

카이제곱 통계량

순서형 종속변수의 경우 우도비 방법을 사용하여 노드 분할과 범주 합치기를 결정하는 카이제곱을 계산합니다. 명목형 종속변수의 경우 다음 방법을 선택할 수 있습니다.

피어슨

이 방법은 보다 빠른 계산을 제공하지만 작은 표본에서는 주의하여 사용해야 합니다. 기본 방법입니다.

우도비

이 방법은 Pearson보다 더 뛰어나지만 계산하는 데 시간이 오래 걸립니다. 작은 표본의 경우 이 방법을 사용하는 것이 좋습니다.

모형 추정

명목형 및 순서형 종속변수의 경우 다음을 지정할 수 있습니다.

최대반복수

기본값은 100입니다. 최대반복수에 도달하여 나무 성장이 중단되는 경우 최대값을 늘리거나 나무 성장을 제어하는 다른 기준 중 하나 이상을 변경하려고 할 수 있습니다.

셀 기대빈도의 최소 변화량

값은 0보다 크고 1보다 작아야 합니다. 기본값은 0.05입니다. 값이 낮을수록 노드 수가 적은 나무를 생성합니다.

Bonferroni 방법을 사용하여 유의성 값 조정

다중비교의 경우 합치기 및 분할 기준의 유의확률이 Bonferroni 방법을 사용하여 조정됩니다. 기본값입니다.

노드 내에서 합친 범주 재분할 허용

명시적으로 범주 합치기를 방지하지 않으면 프로시저에서 독립변수(예측자 변수) 범주를 서로

합쳐 모형을 설명하는 가장 단순한 나무를 생성합니다. 더 나은 솔루션을 제공하는 경우 이 옵션을 사용하여 프로시저에서 합쳐진 범주를 재분할할 수 있습니다.

CHAID 기준을 지정하는 방법

1. 기본 의사결정나무 대화 상자에서 성장방법으로 **CHAID** 또는 **Exhaustive CHAID**를 선택합니다.
2. **CHAID**를 클릭합니다.

CHAID 분석의 척도 구간: CHAID 분석의 경우 분석 전에 척도 독립변수(예측자 변수)가 항상 이산형 집단(예: 0–10, 11–20, 21–30 등)으로 일정하게 나뉩니다. 초기 분할 이후 프로시저에서 연속된 집단을 합칠 수 있지만 초기/최대 집단 수를 제어할 수 있습니다.

고정 숫자
모든 척도 독립변수가 초기에 동일한 집단 수로 나뉩니다. 기본값은 10입니다.

사용자 정의
각 척도 독립변수가 초기에 해당 변수에 대해 지정된 집단 수로 나뉩니다.

척도 독립변수 구간을 지정하는 방법

1. 기본 의사결정나무 대화 상자에서 하나 이상의 척도 독립변수를 선택합니다.
2. **성장방법**으로 **CHAID** 또는 **Exhaustive CHAID**를 선택합니다.
3. **구간**을 클릭합니다.

CRT 및 QUEST 분석에서는 모든 분할이 이분형이며 척도 및 순서형 독립변수가 동일한 방법으로 처리되기 때문에 척도 독립변수 구간 수를 지정할 수 없습니다.

CRT 기준

CRT 성장 법은 노드 내 동질성을 최대화합니다. 노드가 케이스의 동질적 부분집합을 나타내지 않는 정도는 **불순도**로 표시합니다. 예를 들어, 모든 케이스의 종속변수 값이 동일한 터미널 노드는 "순수"하기 때문에 분할이 더 이상 필요없는 동질적 노드입니다.

사용된 방법을 선택하여 불순도 및 노드 분할에 필요한 불순도의 최소 감소량을 측정할 수 있습니다.

불순도 측도
척도 종속변수의 경우 불순도의 최소 제곱편차(LSD) 측도를 사용합니다. 노드 내 분산으로 계산되며 빈도 가중값 또는 영향력 값에 대해 조정됩니다.범주형(명목형, 순서형) 종속변수의 경우 다음의 불순도 측도를 선택할 수 있습니다.

Gini 종속변수 값에 따라 자식 노드의 동질성이 최대가 되도록 분할합니다. Gini는 각 종속변수 범주의 멤버십 제곱 확률을 기준으로 합니다. 노드의 모든 케이스가 단일 범주에 속하면 최소값인 0에 도달합니다. 기본값 측도입니다.

투잉 종속변수 범주가 두 개의 서브클래스로 집단화됩니다. 두 집단이 가장 잘 분리되도록 분할합니다.

순서화 투잉

인접 범주만 집단화할 수 있다는 점을 제외하고는 투잉과 유사합니다. 순서형 종속변수에 대해서만 이 측도를 사용할 수 있습니다.

개선도의 최소 변화량

노드 분할에 필요한 불순도의 최소 감소량입니다. 기본값은 0.0001입니다. 값이 높을수록 노드 수가 적은 나무를 생성합니다.

CRT 기준을 지정하는 방법

1. **성장방법**으로 **CRT**를 선택합니다.
2. **CRT**를 클릭합니다.

QUEST 기준

QUEST 방법의 경우 노드 분할 유의수준을 지정할 수 있습니다. 유의수준이 지정된 값보다 작거나 같지 않은 경우 독립변수를 사용하여 노드를 분할할 수 없습니다. 값은 0보다 크고 1보다 작아야 합니다. 기본값은 0.05입니다. 값이 작을수록 더 많은 독립변수를 최종 모형에서 제외합니다.

QUEST 기준을 지정하는 방법

1. 기본 의사결정나무 대화 상자에서 명목형 종속변수를 선택합니다.
2. **성장방법**으로 **QUEST**를 선택합니다.
3. **QUEST**를 클릭합니다.

나무 잘라내기

CRT 및 QUEST 방법을 통해 나무 **잘라내기**를 수행하여 모형 과적합을 방지할 수 있습니다. 나무는 중지 기준이 충족될 때까지 성장한 후 지정된 최대 위험도 차이를 기준으로 하여 자동으로 가장 작은 부나무로 잘립니다. 위험도 값은 표준 오차로 표시됩니다. 기본값은 1이며 값은 음수가 아니어야 합니다. 위험도가 최소값인 부나무를 얻으려면 0을 지정합니다.

중요사항: 잘라내기가 선택되면 CRT 및 QUEST 방법에 교차 검증을 사용할 수 없습니다.

나무 잘라내기

1. 기본 의사결정나무 대화 상자에서 **성장방법**으로 **CRT** 또는 **QUEST**를 선택합니다.
2. **잘라내기**를 클릭합니다.

노드 잘라내기 및 노드 숨김

잘라낸 나무를 작성하는 경우 나무에서 잘라낸 모든 노드를 최종 나무에서 사용할 수 없습니다. 최종 나무에서 선택된 자식 노드를 대화형으로 숨기거나 표시할 수 있지만 나무 작성 과정에서 잘라낸 노드는 표시할 수 없습니다.

서로게이트

CRT 및 QUEST는 독립변수(예측자 변수)에 대해 **서로게이트**를 사용할 수 있습니다. 해당 변수의 값이 결측된 케이스의 경우 원래 변수와 밀접한 연관이 있는 다른 독립변수가 분류에 사용됩니다. 이러한 대체 예측자를 서로게이트라고 합니다. 모형에 사용할 최대 서로게이트 수를 지정할 수 있습니다.

- 기본적으로 최대 서로게이트 수는 독립변수 수보다 하나가 적습니다. 즉, 각 독립변수에 대해 다른 모든 독립변수를 서로게이트로 사용할 수 있습니다.
- 모형에서 서로게이트를 사용하지 않으려면 서로게이트 수를 0으로 지정합니다.

서로게이트를 지정하는 방법

1. 기본 의사결정나무 대화 상자에서 **성장방법**으로 **CRT** 또는 **QUEST**를 선택합니다.
2. **서로게이트**를 클릭합니다.

옵션

사용가능한 옵션은 성장방법, 종속변수의 측정 수준 및/또는 종속변수 값에 대해 지정된 변수값 레이블이 존재하는지 여부에 따라 달라질 수 있습니다.

오분류 비용

범주형(명목형, 순서형) 종속변수의 경우 오분류 비용을 사용하여 부정확한 분류와 연관된 상대적 페널티 정보를 포함시킬 수 있습니다. 예를 들어, 다음과 같습니다.

- 신용도가 높은 고객에게 대출을 거부하는 비용은 채무를 이행하지 않는 고객에게 대출을 연장하는 비용과 다릅니다.
- 심장병 위험이 높은 개인을 위험이 낮은 것으로 오분류하는 비용은 위험이 낮은 개인을 위험이 높은 것으로 오분류하는 비용보다 더 높습니다.
- 응답할 가능성이 없는 사람에게 대량 메일을 보내는 비용은 상당히 낮지만 응답할 가능성이 있는 사람에게 대량 메일을 보내지 않는 비용은 손실 수입으로 인해 상대적으로 더 높습니다.

참고: 두 개 이상의 범주형 종속변수 값에 값 레이블이 정의되어 있지 않은 경우 이 오분류 비용 대화 상자를 사용할 수 없습니다.

오분류 비용을 지정하는 방법

1. 기본 의사결정나무 대화 상자에서 두 개 이상의 값 레이블이 정의된 범주형(명목형, 순서형) 종속변수를 선택합니다.
2. **오분류 비용**을 클릭합니다.
3. **사용자 정의**를 클릭합니다.
4. 예측 범주 눈금에 하나 이상의 오분류 비용을 입력합니다. 값은 음수가 아니어야 합니다. 정확한 분류는 대각선으로 표시되며 항상 0입니다.

행렬 채우기

대부분의 인스턴스에서는 비용이 대칭이 되기를 원할 수 있습니다. 즉, A를 B로 오분류하는 비용이 B를 A로 오분류하는 비용이 동일한 것입니다. 다음 제어를 사용하면 대칭 비용 행렬을 보다 쉽게 작성할 수 있습니다.

아래쪽 삼각형 중복

행렬의 아래쪽 삼각형(대각선 아래) 값을 해당 위쪽 삼각형 셀에 복사합니다.

위쪽 삼각형 중복

행렬의 위쪽 삼각형(대각선 위) 값을 해당 아래쪽 삼각형 셀에 복사합니다.

평균 셀 값 사용

행렬의 각 절반에서 각각의 셀에 대해 두 값(위쪽 삼각형 및 아래쪽 삼각형)의 평균을 계산하여 그 평균으로 두 값을 대체합니다. 예를 들어, A를 B로 오분류하는 비용이 1이고 B를 A로 오분류하는 비용이 3인 경우 이 제어는 두 값을 평균인 (1+3)/2 = 2로 대체합니다.

이익

범주형 종속변수의 경우 수입 및 비용 값을 종속변수 수준에 할당할 수 있습니다.

- 수입에서 비용을 빼서 이익을 계산합니다.
- 이윤 값은 이익 표의 평균 이윤 및 ROI(투자 수익률) 값에 영향을 미칩니다. 기본 판별 모형구조에는 영향을 미치지 않습니다.
- 수입 및 비용 값은 숫자여야 하고 눈금에 표시된 모든 종속변수 범주에 대해 지정되어야 합니다.

참고: 이 대화 상자를 사용하려면 종속변수의 값 레이블이 정의되어 있어야 합니다. 두 개 이상의 범주형 종속변수 값에 값 레이블이 정의되어 있지 않은 경우 이 대화 상자를 사용할 수 없습니다.

이익을 지정하는 방법

1. 기본 의사결정나무 대화 상자에서 두 개 이상의 값 레이블이 정의된 범주형(명목형, 순서형) 종속변수를 선택합니다.
2. **이익**을 클릭합니다.
3. **사용자 정의**를 클릭합니다.
4. 눈금에 나열된 모든 종속변수 범주에 대한 수입 및 비용 값을 입력합니다.

사전확률

범주형 종속변수가 포함된 CRT 및 QUEST 나무의 경우 소속집단의 사전확률을 지정할 수 있습니다. **사전확률**은 독립변수(예측자 변수) 값에 대한 정보를 얻기 전에 각 종속변수 범주에 대한 전체적인 상대 빈도를 추정한 값입니다. 사전확률을 사용하면 전체 모집단을 대표하지 않는 표본에서 데이터에 의한 나무 성장을 수정할 수 있습니다.

학습 표본에서 가져오기(실제 사전확률)

데이터 파일의 종속변수 값 분포가 모집단 분포를 대표하는 경우 이 설정을 사용합니다. 분할 표본 검증을 사용하는 경우 학습 표본의 케이스 분포를 사용합니다.

참고: 분할 표본 검증의 학습 표본에 케이스가 임의로 할당되므로 학습 표본에 있는 케이스의 실제 분포를 사전에 알지 못합니다. 자세한 정보는 5 페이지의 『검증』 주제를 참조하십시오.

모든 범주에서 동일

종속변수 범주가 모집단에 동일하게 나타나는 경우 이 설정을 사용합니다. 예를 들어, 네 개의 범주가 있는 경우 각각의 범주에 케이스가 약 25%씩 있습니다.

사용자 정의

눈금에 나열된 각 종속변수 범주에 대해 음수가 아닌 값을 입력합니다. 값은 비율, 퍼센트, 빈도 수 또는 범주 전체의 값 분포를 나타내는 다른 값이 될 수 있습니다.

오분류 비용을 사용하여 사전확률 수정

사용자 정의 오분류 비용을 정의하는 경우 이러한 비용을 기준으로 사전확률을 조정할 수 있습니다. 자세한 정보는 10 페이지의 『오분류 비용』 주제를 참조하십시오.

참고: 이 대화 상자를 사용하려면 종속변수의 값 레이블이 정의되어 있어야 합니다. 두 개 이상의 범주형 종속변수 값에 값 레이블이 정의되어 있지 않은 경우 이 대화 상자를 사용할 수 없습니다.

사전확률을 지정하는 방법

1. 기본 의사결정나무 대화 상자에서 두 개 이상의 값 레이블이 정의된 범주형(명목형, 순서형) 종속변수를 선택합니다.
2. **성장방법**으로 **CRT** 또는 **QUEST**를 선택합니다.
3. **사전확률**을 클릭합니다.

점수

순서 종속변수가 있는 CHAID 및 Exhaustive CHAID의 경우 각 종속변수 범주에 사용자 정의 점수를 할당할 수 있습니다. 스코어는 종속변수 범주 간 순서 및 거리를 정의합니다. 점수를 사용하여 순서 값 사이의 상대적 거리를 증가 또는 감소시키거나 값의 순서를 변경할 수 있습니다.

각 범주의 순서 순위 사용

최저 종속변수 범주에 점수 1을 할당하고 다음으로 높은 범주에 점수 2를 할당하는 방법입니다. 기본값입니다.

사용자 정의

눈금에 나열된 각 종속변수 범주에 대한 숫자 점수 값을 입력합니다.

참고: 이 대화 상자를 사용하려면 종속변수의 값 레이블이 정의되어 있어야 합니다. 두 개 이상의 범주형 종속변수 값에 값 레이블이 정의되어 있지 않은 경우 이 대화 상자를 사용할 수 없습니다.

예제

표 2. 사용자 정의 점수 값

값 레이블	원래 값	점수
비숙련 노동자	1	1
숙련 노동자	2	4
사무직	3	4.5
전문직	4	7
관리직	5	6

- 점수로 인해 비숙련 노동자와 숙련 노동자 사이의 상대적 거리는 증가하고 숙련 노동자와 사무직 사이의 상대적 거리는 감소합니다.
- 점수는 관리직과 전문직의 순서를 뒤바꿉니다.

점수를 지정하는 방법

1. 기본 의사결정나무 대화 상자에서 두 개 이상의 값 레이블이 정의된 순서형 종속변수를 선택합니다.
2. **성장방법**으로 **CHAID** 또는 **Exhaustive CHAID**를 선택합니다.
3. **점수**를 클릭합니다.

결측값

결측값 대화 상자에서는 명목, 사용자 결측, 독립변수(예측자 변수) 값 처리를 제어합니다.

- 순서 및 척도 사용자 결측 독립변수 값의 처리는 성장방법에 따라 다릅니다.
- 명목형 종속변수의 처리는 범주 대화 상자에서 지정됩니다. 자세한 정보는 4 페이지의 『범주 선택』 주제를 참조하십시오.
- 순서 및 척도 종속변수의 경우 시스템 결측값 또는 사용자 결측 종속변수 값이 있는 케이스는 항상 제외됩니다.

명목형 독립변수의 사용자 결측값

결측값으로 처리

사용자 결측값을 시스템 결측값과 같이 처리합니다. 시스템 결측값 처리는 성장방법에 따라 다릅니다.

유효한 값으로 처리

명목 독립변수의 사용자 결측값을 나무 성장 및 분류의 일반 변수로 처리합니다.

방법 종속 규칙

전체가 아닌 일부 독립변수 값이 시스템 또는 사용자 결측값인 경우 다음을 적용합니다.

- CHAID 및 Echaustive CHAID의 경우 시스템 및 사용자 결측 독립변수 값이 하나의 결합된 범주로 분석에 포함됩니다. 척도 및 순서 독립변수의 경우 알고리즘이 먼저 유효한 값을 사용하여 범주를 생성한 후 결측 범주를 가장 유사한(유효한) 범주와 합칠 것인지 별도 범주로 유지할 것인지 결정합니다.
- CRT 및 QUEST의 경우 결측 독립변수 값이 있는 케이스가 나무 성장 과정에서 제외하지만 방법에 서로게이트가 포함되어 있는 경우 서로게이트를 사용하여 분류됩니다. 명목 사용자 결측값을 결측으로 처리하는 경우에도 이러한 방법으로 처리합니다. 자세한 정보는 10 페이지의 『서로게이트』 주제를 참조하십시오.

명목형, 독립 사용자 결측 처리를 지정하는 방법

1. 기본 의사결정나무 대화 상자에서 하나 이상의 명목형 독립변수를 선택합니다.
2. **결측값**을 클릭합니다.

모형정보 저장

모형정보를 작업 데이터 파일에 변수로 저장하며 전체 모형을 외부 파일에 XML(PMML) 형식으로 저장할 수 있습니다.

저장 변수

터미널 노드 번호

각 케이스가 할당되는 터미널 노드입니다. 값은 나무 노드 수입니다.

예측값

모형에서 예측한 종속변수의 계층(집단) 또는 값입니다.

예측 확률

모형의 예측과 연관된 확률입니다. 각 종속변수 범주에 대해 하나의 변수가 저장됩니다. 척도 종속변수에 대해서는 사용할 수 없습니다.

표본 할당(학습/검정)

분할 표본 검증의 경우 이 변수는 케이스가 학습 표본에서 사용되었는지, 검정 표본에서 사용되었는지를 나타냅니다. 학습 표본의 경우 값이 1이고 검정 표본의 경우 0입니다. 분할 표본 검증을 선택하지 않은 경우 사용할 수 없습니다. 자세한 정보는 5 페이지의 『검증』 주제를 참조하십시오.

나무 모형을 XML로 내보내기

전체 나무 모형을 XML(PMML) 형식으로 저장할 수 있습니다. 스코어링 목적으로 이 모형 파일을 사용하여 모형 정보를 다른 데이터 파일에 적용할 수 있습니다.

학습 표본

모형을 지정된 파일에 기록합니다. 분할 표본 검증 나무의 경우 이는 학습 표본용 모형입니다.

검정 표본

검정 표본 모형을 지정된 파일에 기록합니다. 분할 표본 검증을 선택하지 않은 경우 사용할 수 없습니다.

결과

사용가능한 결과 옵션은 성장방법, 종속변수의 측정 수준 및 기타 설정에 따라 달라집니다.

나무 표시

나무의 초기 외형을 제어하거나 나무가 완전히 표시되지 않도록 설정할 수 있습니다.

나무 기본적으로 나무 다이어그램이 결과 탭에 표시되는 결과에 포함됩니다. 나무 다이어그램을 결과에서 제외하려면 이 옵션을 선택 취소합니다.

표시 이 옵션은 결과 탭에서 나무 다이어그램의 초기 외형을 제어합니다. 생성된 나무를 편집하여 이러한 모든 속성을 수정할 수도 있습니다.

방향 나무는 루트 노드가 맨 위에 있는 상태로 위에서 아래로, 왼쪽에서 오른쪽 또는 오른쪽에서 왼쪽으로 표시할 수 있습니다.

노드 내용

노드는 표, 도표 또는 모두를 표시할 수 있습니다. 범주형 종속변수의 경우 표는 빈도 개수와 퍼센트를 표시하고 도표는 막대도표입니다. 척도 종속변수의 경우 표는 평균, 표준 편차, 케이스 수 및 예측값을 표시하고 도표는 히스토그램입니다.

척도 기본적으로 큰 나무는 나무를 페이지에 맞추도록 자동으로 크기가 줄어듭니다. 사용자 정의 척도 퍼센트를 200%까지 지정할 수 있습니다.

독립변수 통계량

CHAID 및 Exhaustive CHAID의 경우 통계량은 유의확률 및 자유도와 척도 종속변수의 *F* 값 또는 범주형 종속변수의 카이제곱 값을 포함합니다. CRT의 경우 향상 값이 표시됩니다. QUEST의 경우 척도 및 순서형 독립변수에 대해 *F*, 유의확률 및 자유도가 표시됩니다. 명목형 독립변수에 대해서는 카이제곱, 유의확률 및 자유도가 표시됩니다.

노드 정의

노드 정의는 각 노드 분할에 사용된 독립변수의 값을 표시합니다.

표 형식의 나무

나무의 각 노드에 대한 요약 정보에는 부모 노드 수, 독립변수 통계량, 노드의 독립변수 값, 척도 종속변수의 평균과 표준편차 또는 범주형 종속변수의 개수 및 퍼센트 등이 포함됩니다.

초기 나무 표시를 제어하는 방법

1. 기본 의사결정나무 대화 상자에서 **나무**를 클릭합니다.

통계

사용가능한 통계량 표는 종속변수의 측정 수준, 성장방법 및 기타 설정에 따라 달라집니다.

모형

요약 요약에는 사용된 방법, 모형에 포함된 변수, 지정되었으나 모형에 포함되지 않은 변수 등이 포함됩니다.

위험도

위험도 추정값 및 해당 표준 오차입니다. 나무의 예측 정확도 측도입니다.

- 범주형 종속변수의 경우 위험도 추정값은 사전확률 및 오분류 비용에 대해 조정한 후 부정확하게 분류된 케이스의 비율입니다.
- 척도 종속변수의 경우 위험도 추정값은 노드 내부 분산입니다.

분류표

범주형(명목형, 순서형) 종속변수의 경우 이 표는 각 종속변수 범주에 대해 정확하게 분류된 케이스 수 및 부정확하게 분류된 케이스 수를 나타냅니다. 척도 종속변수에 대해서는 사용할 수 없습니다.

비용, 사전확률, 점수 및 이익 값

범주형 종속변수의 경우 이 표는 분석에 사용된 비용, 사전확률, 점수 및 이익 값을 나타냅니다. 척도 종속변수에 대해서는 사용할 수 없습니다.

독립변수

모형에 대한 중요도

CRT 성장방법의 경우 해당 모형에 대한 중요도에 따라 독립변수(예측자 변수)의 순위를 지정합니다. QUEST 또는 CHAID 방법에 대해서는 사용할 수 없습니다.

서로게이트에 의한 분리

CRT 및 QUEST 성장 방법은 모형에 서로게이트가 포함된 경우 나무의 각 분할에 대한 서로게이트를 나열합니다. CHAID 방법에 대해서는 사용할 수 없습니다. 자세한 정보는 10 페이지의 『서로게이트』 주제를 참조하십시오.

노드 성능

요약 척도 종속변수의 경우 표에는 노드 수, 케이스 수 및 종속변수의 평균값이 포함됩니다. 이익이 정의되어 있는 범주형 종속변수의 경우 표에는 노드 수, 케이스 수, 평균 이익 및 ROI(투자 수익율) 값이 포함됩니다. 이익이 정의되어 있지 않은 범주형 종속변수에 대해서는 사용할 수 없습니다. 자세한 정보는 11 페이지의 『이익』 주제를 참조하십시오.

목표 범주별

대상 범주가 정의되어 있는 범주형 종속변수의 경우 표에는 노드 및 백분위수 집단별 퍼센트 이득, 반응 퍼센트 및 지수 퍼센트(리프트)가 포함됩니다. 각 대상 범주마다 별

도의 표가 생성됩니다. 대상 범주가 정의되어 있지 않은 척도 종속변수 또는 범주형 종속변수에 대해서는 사용할 수 없습니다. 자세한 정보는 4 페이지의 『범주 선택』 주제를 참조하십시오.

행 노드 성능 표에는 터미널 노드, 백분위수 또는 모두에 따라 결과가 표시될 수 있습니다. 모두를 선택하는 경우 각 대상 범주마다 두 개의 표가 생성됩니다. 백분위수 표는 정렬 순서를 기준으로 각 백분위수의 누적값을 표시합니다.

정렬 순서

값은 종속변수의 측정 수준에 따라 다르며 이득 요약과 이득 표에 대해 다릅니다.

백분위수 증가

백분위수 표의 경우 백분위수 증가(1, 2, 5, 10, 20, 25)를 선택할 수 있습니다.

누적 통계량 표시

터미널 노드 표의 경우 누적 결과가 포함된 각 표에 추가 열이 표시됩니다.

통계량 결과를 선택하는 방법

1. 기본 의사결정나무 대화 상자에서 **통계량**을 클릭합니다.

도표

사용가능한 도표는 종속변수의 측정 수준, 성장방법 및 기타 설정에 따라 달라집니다.

모형에 대한 독립변수 중요도

독립변수(예측자 변수)별 모형 중요도에 대한 막대도표입니다. CRT 성장방법에만 사용할 수 있습니다.

노드 성능

이득 이득은 각 노드의 대상 범주에 있는 전체 케이스의 퍼센트이며 (노드 대상 n / 전체 대상 n) x 100으로 계산됩니다. 이득 도표는 누적 백분위수 이득의 선도표이며 (누적 백분위수 대상 n / 전체 대상 n) x 100으로 계산됩니다. 각 대상 범주마다 별도의 선도표가 생성됩니다. 대상 범주가 정의되어 있는 범주형 종속변수에 대해서만 사용할 수 있습니다. 자세한 정보는 4 페이지의 『범주 선택』 주제를 참조하십시오.

누적값을 보고하는 백분위수에 대한 이득 표의 이득 퍼센트 열에 표시되는 값과 동일한 값을 이득 도표에서 도표화합니다.

지수 지수는 전체 표본의 전체 대상 범주 반응 퍼센트와 비교한 대상 범주의 노드 반응 퍼센트 비율입니다. 지수 도표는 누적 백분위수 지수 값의 선도표입니다. 범주형 종속변수에 대해서만 사용할 수 있습니다. 누적 백분위수 지수는 (누적 백분위수 반응 퍼센트 / 전체 반응 퍼센트) x 100으로 계산됩니다. 각 대상 범주마다 별도의 도표가 생성되고 대상 범주가 정의되어야 합니다.

지수 도표는 백분위수에 대한 이득 표의 지수 열에 표시되는 값과 동일한 값을 도표화합니다.

응답 지정된 대상 범주에 있는 노드 내 케이스 퍼센트입니다. 반응 도표는 누적 백부위수 반

응의 선도표이며 (누적 백분위수 대상 n / 누적 백분위수 총계 n) x 100으로 계산됩니다. 대상 범주가 정의되어 있는 범주형 종속변수에 대해서만 사용할 수 있습니다.

반응 도표는 백분위수에 대한 이득 표의 반응 열에 표시되는 값과 동일한 값을 도표화합니다.

평균 종속변수에 대한 누적 백분위수 평균값의 선도표입니다. 척도 종속변수에 대해서만 사용할 수 있습니다.

평균 이익

누적 평균 이익의 선도표입니다. 이익이 정의되어 있는 범주형 종속변수에 대해서만 사용할 수 있습니다. 자세한 정보는 11 페이지의 『이익』 주제를 참조하십시오.

평균 이익 도표는 백분위수에 대한 이득 요약 표의 이익 열에 표시되는 값과 동일한 값을 도표화합니다.

투자 수익률(ROI)

누적 투자 수익률(ROI)의 선도표입니다. ROI는 비용에 대한 이익의 비율로 계산됩니다. 이익이 정의되어 있는 범주형 종속변수에 대해서만 사용할 수 있습니다.

ROI 도표는 백분위수에 대한 이득 요약 표의 *ROI* 열에 표시되는 값과 동일한 값을 도표화합니다.

백분위수 증가

모든 백분위수 도표에 대해 이 설정은 도표에 표시되는 백분위수 증가(1, 2, 5, 10, 20, 25)를 제어합니다.

도표 결과를 선택하는 방법

1. 기본 의사결정나무 대화 상자에서 **도표**를 클릭합니다.

선택 및 스코어링 규칙

규칙 대화 상자에서는 명령문, SQL 또는 단순(일반 영어) 텍스트 양식의 선택 또는 분류/예측 규칙을 생성하는 기능을 제공합니다. 이러한 규칙을 결과 탭에 표시하고/표시하거나 외부 파일에 저장할 수 있습니다.

분류 규칙 생성

선택 및 점수화 규칙 설정을 사용하려면 선택합니다.

명령문

결과 탭에 표시된 결과 및 외부 파일에 저장된 선택 규칙에서 선택 규칙의 양식을 제어합니다.

SPSS Statistics

명령문 언어. 케이스 부분집합 선택에 사용될 수 있는 필터 조건을 정의하는 명령문 집합 또는 케이스 스코어링에 사용될 수 있는 COMPUTE 명령문으로 규칙이 표현됩니다.

SQL 표준 SQL 규칙을 생성하여 데이터베이스에서 레코드를 선택 또는 추출하거나 이러한 레코드에 값을 할당합니다. 생성된 SQL 규칙에는 표 이름이나 기타 데이터 소스 정보가 포함되지 않습니다.

단순 텍스트

일반 영어 의사 코드. 각 노드에 대한 모형의 분류 또는 예측을 설명하는 논리 "if...then" 명령문 집합으로 규칙이 표현됩니다. 이 양식의 규칙은 정의된 변수 및 값 레이블 또는 변수 이름 및 데이터 값을 사용할 수 있습니다.

유형 IBM® SPSS Statistics 및 SQL 규칙의 경우 생성된 규칙의 유형(선택 또는 스코어링 규칙)을 제어합니다.

케이스에 값 할당

규칙을 사용하여 노드 멤버십 기준을 충족하는 케이스에 모형 예측을 지정할 수 있습니다. 노드 멤버십 기준을 충족하는 각 노드에 대해 별도의 규칙이 생성됩니다.

케이스 선택

규칙을 사용하여 노드 멤버십 기준을 충족하는 케이스를 선택할 수 있습니다. IBM SPSS Statistics 및 SQL 규칙의 경우 단일 규칙을 생성하여 선택 기준을 충족하는 모든 케이스를 선택합니다.

SPSS Statistics 및 SQL 규칙에 서로게이트 포함

CRT 및 QUEST의 경우 규칙에 모형의 서로게이트 예측자를 포함시킬 수 있습니다. 서로게이트를 포함하는 규칙은 매우 복잡할 수 있습니다. 일반적으로 나무에 대한 개념적 정보를 유도하려는 경우 서로게이트를 제외합니다. 케이스에 불완전한 독립변수(예측자 변수) 데이터가 있고 나무와 유사한 규칙을 원하는 경우 서로게이트를 포함시킵니다. 자세한 정보는 10 페이지의 『서로게이트』 주제를 참조하십시오.

노드 생성된 규칙의 범위를 제어합니다. 범위에 포함된 각 노드에 대해 별도의 규칙이 생성됩니다.

모든 터미널 노드

각 터미널 노드에 대한 규칙을 생성합니다.

가장 좋은 터미널 노드

지수 값을 기준으로 상위 *n*개의 터미널 노드에 대한 규칙을 생성합니다. 개수가 나무의 터미널 노드 수를 초과하는 경우 모든 터미널 노드에 대한 규칙이 생성됩니다

지정된 케이스 퍼센트까지의 가장 좋은 터미널 노드.

지수 값을 기준으로 상위 *n* 케이스 퍼센트의 터미널 노드에 대한 규칙을 생성합니다

지수 값이 분리점 값보다 크거나 같은 터미널 노드.

지수 값이 지정된 값보다 크거나 같은 모든 터미널 노드에 대한 규칙을 생성합니다. 지수 값이 100보다 크다는 것은 해당 노드의 대상 범주에 있는 케이스 퍼센트가 루트 노드의 퍼센트를 초과한다는 것을 의미합니다

모든 노드

모든 노드에 대한 규칙을 생성합니다.

참고:

- 지수 값을 기준으로 하는 노드 선택은 대상 범주가 정의되어 있는 범주형 종속변수에 대해서만 사용할 수 있습니다. 다중 대상 범주를 지정한 경우 각 대상 범주에 대해 별도의 규칙 집합이 생성됩니다.
- 케이스 선택을 위한 IBM SPSS Statistics 및 SQL 규칙의 경우(값 할당을 위한 규칙이 아님) **모든 노드** 및 **모든 터미널 노드**가 분석에 사용된 모든 케이스를 선택하는 규칙을 효과적으로 생성합니다.

파일에 규칙 내보내기
규칙을 외부 텍스트 파일에 저장합니다.

또한 최종 나무 모형에서 선택된 노드를 기준으로 선택 또는 스코어링 규칙을 대화형으로 생성하고 저장할 수 있습니다.

참고: 명령문 양식의 규칙을 다른 데이터 파일에 적용하는 경우, 최종 모형에 포함된 독립변수와 이름이 같고 동일한 메트릭으로 측정되며 동일한 사용자 정의 결측값이 있는(존재하는 경우) 변수가 해당 데이터 파일에 포함되어 있어야 합니다.

선택 또는 점수화 규칙을 지정하는 방법

1. 기본 의사결정나무 대화 상자에서 **규칙**을 클릭합니다.

주의사항

이 정보는 미국에서 제공되는 제품 및 서비스용으로 작성된 것입니다. 본 자료는 다른 언어로도 제공될 수 있습니다. 그러나 자료에 접근하기 위해서는 해당 언어로 된 제품 또는 제품 버전의 사본이 필요할 수 있습니다.

IBM은 다른 국가에서 이 책에 기술된 제품, 서비스 또는 기능을 제공하지 않을 수도 있습니다. 현재 사용할 수 있는 제품 및 서비스에 대한 정보는 한국 IBM 담당자에게 문의하십시오. 이 책에서 IBM 제품, 프로그램 또는 서비스를 언급했다고 해서 해당 IBM 제품, 프로그램 또는 서비스만을 사용할 수 있다는 것을 의미하지는 않습니다. IBM의 지적 재산권을 침해하지 않는 한, 기능상으로 동등한 제품, 프로그램 또는 서비스를 대신 사용할 수도 있습니다. 그러나 비IBM 제품, 프로그램 또는 서비스의 운영에 대한 평가 및 검증은 사용자의 책임입니다.

IBM은 이 책에서 다루고 있는 특정 내용에 대해 특허를 보유하고 있거나 현재 특허 출원 중일 수 있습니다. 이 책을 제공한다고 해서 특허에 대한 라이센스까지 부여하는 것은 아닙니다. 라이센스에 대한 의문사항은 다음으로 문의하십시오.

07326
서울특별시 영등포구
국제금융로 *10, 3IFC*
한국 아이.비.엠 주식회사
대표전화서비스*: 02-3781-7114*

2바이트(DBCS) 정보에 관한 라이센스 문의는 한국 IBM에 문의하거나 다음 주소로 서면 문의하시기 바랍니다.

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

IBM은 타인의 권리 비침해, 상품성 및 특정 목적에의 적합성에 대한 묵시적 보증을 포함하여(단, 이에 한하지 않음) 묵시적이든 명시적이든 어떠한 종류의 보증 없이 이 책을 "현상태대로" 제공합니다. 일부 국가에서는 특정 거래에서 명시적 또는 묵시적 보증의 면책사항을 허용하지 않으므로, 이 사항이 적용되지 않을 수도 있습니다.

이 정보에는 기술적으로 부정확한 내용이나 인쇄상의 오류가 있을 수 있습니다. 이 정보는 주기적으로 변경되며, 변경된 사항은 최신판에 통합됩니다. IBM은 이 책에서 설명한 제품 및/또는 프로그램을 사전 통지 없이 언제든지 개선 및/또는 변경할 수 있습니다.

이 정보에서 언급되는 비IBM 웹 사이트는 단지 편의상 제공된 것으로, 어떤 방식으로든 이들 웹 사이트를 옹호하고자 하는 것은 아닙니다. 해당 웹 사이트의 자료는 본 IBM 제품 자료의 일부가 아니므로 해당 웹 사이트 사용으로 인한 위험은 사용자 본인이 감수해야 합니다.

IBM은 귀하의 권리를 침해하지 않는 범위 내에서 적절하다고 생각하는 방식으로 귀하가 제공한 정보를 사용하거나 배포할 수 있습니다.

(i) 독립적으로 작성된 프로그램과 기타 프로그램(본 프로그램 포함) 간의 정보 교환 및 (ii) 교환된 정보의 상호 이용을 목적으로 본 프로그램에 관한 정보를 얻고자 하는 라이센스 사용자는 다음 주소로 문의하십시오.

07326
서울특별시 영등포구
국제금융로 *10, 3IFC*
한국 아이.비.엠 주식회사
대표전화서비스*: 02-3781-7114*

이러한 정보는 해당 조건(예를 들면, 사용료 지불 등)하에서 사용될 수 있습니다.

이 정보에 기술된 라이센스가 부여된 프로그램 및 프로그램에 대해 사용 가능한 모든 라이센스가 부여된 자료는 IBM이 IBM 기본 계약, IBM 프로그램 라이센스 계약(IPLA) 또는 이와 동등한 계약에 따라 제공한 것입니다.

인용된 성능 데이터와 고객 예제는 예시 용도로만 제공됩니다. 실제 성능 결과는 특정 구성과 운영 조건에 따라 다를 수 있습니다.

비IBM 제품에 관한 정보는 해당 제품의 공급업체, 공개 자료 또는 기타 범용 소스로부터 얻은 것입니다. IBM에서는 이러한 제품들을 테스트하지 않았으므로, 비IBM 제품과 관련된 성능의 정확성, 호환성 또는 기타 청구에 대해서는 확신할 수 없습니다. 비IBM 제품의 성능에 대한 의문사항은 해당 제품의 공급업체에 문의하십시오.

IBM이 제시하는 방향 또는 의도에 관한 모든 언급은 특별한 통지 없이 변경될 수 있습니다.

이 정보에는 일상의 비즈니스 운영에서 사용되는 자료 및 보고서에 대한 예제가 들어 있습니다. 이들 예제에는 개념을 가능한 완벽하게 설명하기 위하여 개인, 회사, 상표 및 제품의 이름이 사용될 수 있습니다. 이들 이름은 모두 가공의 것이며 실제 인물 또는 기업의 이름과 유사하더라도 이는 전적으로 우연입니다.

저작권 라이센스:

이 정보에는 여러 운영 플랫폼에서의 프로그래밍 기법을 보여주는 원어로 된 샘플 응용프로그램이 들어 있습니다. 귀하는 이러한 샘플 프로그램의 작성 기준이 된 운영 플랫폼의 애플리케이션 프로그래밍 인터페이스(API)에 부합하는 애플리케이션을 개발, 사용, 판매 또는 배포할 목적으로 IBM에 추가 비용을 지불하지 않고 이들 샘플 프로그램을 어떠한 형태로든 복사, 수정 및 배포할 수 있습니다. 이러한

샘플 프로그램은 모든 조건하에서 완전히 테스트된 것은 아닙니다. 따라서 IBM은 이러한 프로그램의 신뢰성, 서비스 가능성 또는 기능을 보증하거나 진술하지 않습니다. 본 샘플 프로그램은 일체의 보증 없이 "현상태대로" 제공됩니다. IBM은 귀하의 샘플 프로그램 사용과 관련되는 손해에 대해 책임을 지지 않습니다.

이러한 샘플 프로그램 또는 파생 제품의 각 사본이나 그 일부에는 반드시 다음과 같은 저작권 표시가 포함되어야 합니다.

상표

IBM, IBM 로고 및 ibm.com은 전세계 여러 국가에 등록된 International Business Machines Corp.의 상표 또는 등록상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 타사의 상표입니다. 현재 IBM 상표 목록은 "저작권 및 상표 정보" 웹 페이지(www.ibm.com/legal/copytrade.shtml)에 있습니다.

Adobe, Adobe 로고, PostScript 및 PostScript 로고는 미국 및/또는 기타 국가에서 사용되는 Adobe Systems Incorporated의 등록상표 또는 상표입니다.

Intel, Intel 로고, Intel Inside, Intel Inside 로고, Intel Centrino, Intel Centrino 로고, Celeron, Intel Xeon, Intel SpeedStep, Itanium 및 Pentium은 미국 또는 기타 국가에서 사용되는 Intel Corporation 또는 그 계열사의 상표 또는 등록상표입니다.

Linux는 미국 또는 기타 국가에서 사용되는 Linus Torvalds의 등록상표입니다.

Microsoft, Windows, Windows NT 및 Windows 로고는 미국 또는 기타 국가에서 사용되는 Microsoft Corporation의 상표입니다.

UNIX는 미국 및 기타 국가에서 사용되는 The Open Group의 등록상표입니다.

Java 및 모든 Java 기반 상표와 로고는 Oracle 및/또는 그 계열사의 상표 또는 등록상표입니다.

색인

[가]

[나]

[마]

[바]

[사]

[아]

[자]

[차]

[타]

C

G

Q

S