

mgcvUI: mgcv(일반화 가법 모형) 패키지를 위한 대화형 인터페이스 (한국어판)

한국어 번역본. 인공지능의 도움을 받아 번역하고 편집부가 검토했습니다. 영문판이正本입니다.

William Bert Craytor

제 1 권 제 1 호 · 2026 년 여름

발행일: 2026 년 7 월 1 일

DOI: [10.67330/xtv1pa25](https://doi.org/10.67330/xtv1pa25)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

이 논문은 오픈 액세스로 공개되며 [크리에이티브 커먼즈 저작자표시 4.0 국제 라이선스 \(CC BY 4.0\)](#)의 조건에 따라 배포됩니다.

초록. mgcvUI 는 R mgcv 패키지를 위한 그래픽 사용자 인터페이스 (graphical user interface) 이다. mgcv 는 자동 평활도 선택 (smoothness selection) 을 갖춘 벌점 회귀 스플라인 (penalized regression spline) 으로 일반화 가법 모형 (generalized additive model, GAM) 을 적합한다. mgcvUI 는 범용 예측 모델링, 부동산 감정평가, 시장지역 분석 (market-area analysis) 의 세 가지 목적 모드를 제공하며, 데이터 가져오기, 평활항 (smooth term) 설정, 모형 적합, 진단 그림 및 효과 그림 확인, 보고서 내려받기에 이르는 과정을 사용자에게 안내한다. 본고는 mgcvUI 의 데이터 형식 요건, 모델링 작업 흐름, 출력 화면, 그리고 전체 기능 참조를 문서화한다.

키워드: 일반화 가법 모형, GAM, 벌점 스플라인, 평활화, mgcv, R, 그래픽 사용자 인터페이스, 부동산 감정평가

발행인 주 (2026 년 7 월): 인쇄 직전에 mgcvUI 는 earthUI(0.11.0 이후 버전) 에 통합되었으며, 본고에서 설명하는 GAM 작업 흐름은 이제 MARS 및 엘라스틱 넷 작업 흐름과 함께 earthUI 에서 제공된다. mgcvUI 는 최종 배포판 (0.3.0) 형태로 <https://github.com/wcraytor/mgcvUI>에 계속 공개되어 있으며, 향후 개발은 earthUI 에서 이어진다. 본고는 배포된 상태의 mgcvUI 를 다룬다.

목차

1	서론	4
1.1	역사와 배경	4
1.1.1	일반화 가법 모형	4
1.1.2	mgcv 패키지	4
1.1.3	MARS(earth) 및 엘라스틱 넷 (glmnet) 과의 비교	5
1.1.4	earthUI-mgcvUI 파이프라인	5
1.1.5	권장 작업 흐름	5
1.2	mgcvUI 란 무엇인가?	6
1.3	세 가지 목적 모드	6
1.4	부동산 전용 기능	6
1.5	시작하기	7
2	MLS 입력 데이터 요건	8
2.1	파일 형식과 구조	8
2.2	감정평가 모드에 필요한 열	8
2.3	특수 열 명명 규칙	9
2.4	데이터 품질과 완전성	9
2.5	대상 행의 위치	9

3	범용 모드	10
3.1	개요	10
3.2	사이드바 작업 흐름	10
3.3	earthUI 에서 가져오기 (선택 사항)	11
3.4	메인 패널 탭	11
3.5	설정의 유지	12
3.6	다크 모드	12
3.7	로케일과 지역 설정	12
4	감정평가 모드	14
4.1	대상 행의 처리	14
4.2	기준일과 매매 경과일	14
4.3	특수 열 지정	14
4.4	RCA 조정 개요	14
5	시장지역 분석 모드	15
5.1	감정평가 모드와의 차이	15
5.2	시장 모드를 사용하는 경우	15
6	변수 선택	16
6.1	목표변수	16
6.2	반응변수 변환	16
6.3	예측변수 표	16
6.4	데이터 유형 탐지와 재지정	17
6.5	특수 열 유형 참조	17
7	모수 선택	18
7.1	모수 프리셋	18
7.2	Family(분포족)	18
7.3	Method(방법)	18
7.4	Gamma(평활 승수)	18
7.5	교차검증	19
7.6	Select(수축)	19
7.7	기본 기저	19
7.8	기본 k(기저 차원)	19
7.9	텐서 유형	19
7.10	허용 상호작용	20
7.11	요인별 평활 상호작용	20
7.12	고급 모수	20
8	모형 적합	21
8.1	Fit 버튼	21
8.2	적합 중에 일어나는 일	21
8.3	백그라운드 적합과 진행 상황	21
9	결과 탭	22
9.1	Data(데이터)	22
9.2	Equation(식)	22
9.3	Summary(요약)	22
9.4	Variable Importance(변수 중요도)	22
9.5	기여도 그림	23
9.5.1	일변량 평활항 — $s(x)$	23

9.5.2	요인별 평활항 — $s(x, \text{by} = \text{factor})$	23
9.5.3	상호작용항 — $ti(x_1, x_2)$ 및 $te(x_1, x_2)$	23
9.5.4	모수항 — 선형항과 요인항	23
9.6	Correlation(상관)	23
9.7	Diagnostics(진단)	23
9.8	RCA Adjustments(RCA 조정)	23
9.9	Anova(분산분석)	24
9.10	Mgecv Output(mgecv 출력)	24
9.11	Sign Check(부호 점검)	24
9.12	Concurvity(공곡선성)	24
9.13	Basis Check(기저 점검)	24
10	데이터 내려받기	25
10.1	출력 열	25
10.2	CQA 점수	25
11	RCA 계산과 내려받기	26
11.1	RCA 대화상자 열기	26
11.2	CQA 점수 보간	26
11.3	출력 열	26
12	거래사례비교 그리드	27
12.1	비교사례 선택 대화상자	27
12.2	통합문서 구성	27
12.3	시트 보호와 잔차 칸	27
13	보고서 내려받기	28
13.1	1 단계: Quarto 보고서 생성	28
13.2	2 단계: Quarto 보고서 변환	28
13.3	보고서 내용	28
14	모형 함수 내보내기	29
14.1	R 함수	29
14.2	Python 함수	29
14.3	C++ 헤더	29
14.4	SQLite 데이터베이스	29
15	earthUI 및 glmnetUI 와의 비교	30
15.1	주요 차이	30
15.2	공통 기능	30
15.3	권장 작업 흐름	30
16	시연용 데이터셋: Appraisal_1.csv	31
16.1	설명	31
16.2	열	31
16.3	빠른 시작 예시	32
17	시스템 요건과 문제 해결	33
17.1	지원 플랫폼	33
17.2	R 버전	33
17.3	필수 R 패키지	33
17.4	선택적 의존성	33
17.5	플랫폼별 유의 사항	34

17.5.1 macOS	34
17.5.2 Windows	34
17.5.3 Linux	34
17.6 점진적 기능 축소	34
17.7 문제 해결	35
17.7.1 “PDF option is not available”	35
17.7.2 “Settings will not persist”	35
17.7.3 “Diagnostics plot unavailable”	35
17.7.4 “Report error: number of items to replace...”	35
17.7.5 포트 7880 이 이미 사용 중	35
17.7.6 Linux 에서 패키지 설치 실패	35

1 서론

1.1 역사와 배경

1.1.1 일반화 가법 모형

일반화 가법 모형 (generalized additive model, GAM) 은 Trevor Hastie 와 Robert Tibshirani 가 1990 년 저서에서 정식화한 모형이다. GAM 은 각 선형항 $\beta_j x_j$ 를 평활 함수 (smooth function) $f_j(x_j)$ 로 대체함으로써 선형 모형을 확장하며, 모형은 다음과 같은 형태가 된다:

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_p(x_p) + \varepsilon$$

각 f_j 는 벌점 회귀 스플라인 (penalized regression spline) 을 이용해 데이터로부터 추정된다. 이러한 평활 함수 덕분에 모형은 사용자가 함수 형태를 미리 지정하지 않아도 비선형 관계를 포착할 수 있다. 벌점은 각 평활항의 굴곡 정도 (wiggliness) 를 제어하여, 실제 패턴을 따라갈 만한 유연성은 남겨 두면서 과적합을 방지한다.

1.1.2 mgcv 패키지

R 패키지 mgcv(Mixed GAM Computation Vehicle) 는 베스 대학교의 Simon Wood 가 개발했다 (Wood 2017). 2001 년에 처음 공개된 이후 현재는 R 에서 GAM 을 구현하는 표준으로 자리 잡았으며, 모든 R 설치본에 함께 배포된다. 그 기반 방법론은 안정적인 다중 평활 모수 추정 (Wood 2004), 박판 회귀 스플라인 (thin-plate regression spline) (Wood 2003), 빠르고 안정적인 REML 추정 (Wood 2011), 일반 평활 모형의 선택 (Wood, Pya, and Säfken 2016) 등을 다룬 일련의 논문 에 기술되어 있다. 주요 기능은 다음과 같다:

- 자동 평활도 선택 (smoothness selection): 각 항의 평활 모수는 모형 적합 과정의 일부로 추정되며, REML(제한 최대가능도), GCV(일반화 교차검증), ML(최대가능도) 중 하나를 사용한다.
- 다양한 기저 유형: 박판 회귀 스플라인 (tp), 삼차 회귀 스플라인 (cr), P-스플라인 (ps), B-스플라인 (bs) 이 있으며 각각 성질이 다르다.
- 텐서곱 평활: `te()` 와 `ti()` 는 차원의 저주 없이 변수 간 평활 상호작용을 구현한다.
- 변수 선택: `select = TRUE` 옵션은 평활항을 0 으로 수축시킬 수 있는 추가 벌점을 더하여 사실상 변수 선택을 수행한다.
- 공곡선성 진단 (concurvity diagnostics): 평활항에서 다중공선성에 대응하는 현상을 측정한다.

1.1.3 MARS(earth) 및 엘라스틱 넷 (glmnet) 과의 비교

earth 응용 프로그램과 두 개의 후처리 모듈 (glmnet 과 mgcv) 은 서로 다른 모델링 엔진을 사용한다. 다음 표는 주요 차이를 요약한 것이다:

특성	mgcv(GAM)	earth(MARS)	glmnet(엘라스틱 넷)
모형 유형	평활 비선형 (별점 스플라인)	구간별 선형 (적응 스플라인)	선형 (정칙화 적응)
비선형성	평활 함수 (박판, 삼차 등) 를 통해 자동 처리	데이터 기반 매듭점을 갖는 힌지 함수로 자동 처리	상호작용이나 기저 확장을 통해서만 가능
변수 선택	이중 별점 수축 (select=TRUE) 을 통한 선택적 기능	GCV 가지치기를 동반한 전진/후진 단계적 선택으로 내장	L_1 별점 (라쏘) 을 통해 내장
상호작용	텐서곱 평활 (te(), ti())	힌지 기반 곱 (해석 가능, 최대 3 차)	쌍별 교차곱 (해석이 어려움)
해석 가능성	평활한 부분효과 곡선으로 매우 직관적	g-함수가 변수별 구간별 선형 부분효과를 보여 줌	계수가 선형 가중치이므로 개별 설명이 쉬움
적합한 용도	평활한 비선형 관계, 유연한 모델링	비선형성과 상호작용의 자동 탐지	고차원 데이터, 변수 선택, 소표본
RCA 조정	부분효과를 통한 평활 조정	g-함수를 통한 구간별 선형 조정	변수별 선형 조정
과적합 제어	REML/GCV 평활 모수 추정	GCV 기반 가지치기	교차검증된 람다

MARS(Multivariate Adaptive Regression Splines, 다변량 적응 회귀 스플라인) 는 1991 년 Jerome Friedman 이 제안했다. 이 방법은 힌지 함수와 그 매듭점 위치를 데이터로부터 적응적으로 선택하여 구간별 선형 모형을 구축한다. R 구현은 Stephen Milborrow 의 **earth** 패키지이다.

엘라스틱 넷 (elastic net) 회귀는 Friedman, Hastie, Tibshirani (2010) 의 **glmnet** 패키지로 구현되어 있으며, 라쏘 (L_1) 와 릿지 (L_2) 별점을 결합하여 변수 선택과 계수 수축을 동시에 수행한다.

1.1.4 earthUI-mgcvUI 파이프라인

earthUI 의 자동 매듭점 탐색과 mgcvUI 의 평활 추정을 결합하면 강력한 작업 흐름이 만들어진다. earthUI 가 .rds 결과 파일을 내보내고 mgcvUI 가 이를 가져오면 다음과 같이 진행된다:

1. **earth** 가 매듭점을 찾아낸다 — MARS 는 각 변수에 대해 데이터에 기반한 힌지 위치 (변화점) 를 찾는다.
2. **mgcvUI** 가 이를 기준으로 사용한다 — earth 의 매듭점은 GAM 에서 삼차 회귀 스플라인 (cr 기저) 의 바탕이 되며, 데이터의 실제 구조를 반영하는 매듭점 위치로 평활향을 유리하게 출발시킨다.
3. **mgcv** 가 적합을 정교화한다 — 별점 스플라인 추정은 earth 의 구간별 선형 적합을 매끄럽게 다듬어, earth 가 발견한 전반적 형태를 유지하면서 평활한 부분효과 곡선을 산출한다.

이 파이프라인은 탐색적 모델링 (earth) 과 확증적 모델링 (GAM) 을 이어 주며, 두 접근법의 장점을 결합한다.

1.1.5 권장 작업 흐름

해석 가능하고 방어 가능한 모형이 필요한 부동산 감정평가 및 이와 유사한 용도에서는 다음을 권장한다:

1. **earthUI** 로 시작하여 데이터에서 중요한 변수, 비선형 관계, 상호작용을 발견한다. earth의 자동 기저 선택과 g-함수는 훌륭한 초기 통찰을 제공한다.
2. 부분효과를 더 매끄럽게 만들고 싶다면 **mgcvUI** 로 정교화한다. earthUI의 매듭점을 mgcvUI로 가져오면 구간별 선형 모형에서 평활 모형으로 자연스럽게 이행할 수 있다.
3. 공격적인 변수 선택 (라쏘)이 필요할 때, 예측변수의 개수가 관측치 수에 근접하거나 이를 초과할 때, 또는 방어 가능성을 위해 정칙화된 순수 선형 모형을 원할 때에는 **glmnetUI** 를 사용한다.

1.2 mgcvUI란 무엇인가?

mgcvUI는 R mgcv 패키지를 위한 그래픽 사용자 인터페이스이다. 로컬 Shiny 응용 프로그램으로 실행되며 로그인, 서버, 계정이 필요 없다. R에서 실행한 뒤 데이터셋 (CSV 또는 Excel)을 가져오고, 모형을 설정한 다음 대화형으로 GAM을 적합하면 된다.

현재 배포판 기준으로 mgcvUI는 더 이상 별도로 설치하는 패키지가 아니다. mgcvUI는 earthUI 패키지 안에 GAM 후처리 모듈로 포함되어 배포되며 `earthUI::launch_mgcv()`로 시작한다. 응용 프로그램 자체와 본고에 설명된 모든 내용은 달라지지 않았고, 설치 및 실행 명령만 달라졌다. 세 루틴 (earth, glmnet, mgcv)은 하나의 프로젝트 트리, 하나의 설정 데이터베이스, 하나의 지원 인프라 사본을 공유한다.

이 응용 프로그램은 데이터 가져오기, 평활항 설정을 포함한 변수 구성, 백그라운드 처리 방식의 모형 적합, 진단 그림, 평활한 부분효과 곡선, 그리고 Word, PDF, HTML 형식으로 내려받을 수 있는 보고서에 이르는 완결된 작업 흐름을 제공한다.

일반화 가법 모형은 각 예측변수에 대해 선형 관계 βx 를 평활 함수 $f(x)$ 로 대체한다. 이는 다음을 의미한다:

- **함수 형태를 지정할 필요가 없다** — 각 관계의 형태는 데이터가 결정한다.
- **평활한 부분효과** — 각 예측변수의 기여는 직선이 아니라 평활한 곡선이므로, 변수가 서로 다른 값에서 반응변수에 어떤 영향을 주는지 쉽게 확인할 수 있다.
- **자동 평활도 선택** — mgcv는 REML 또는 GCV를 통해 각 곡선이 얼마나 굴곡져야 하는지를 결정하여, 적합도와 과적합 사이의 균형을 맞춘다.

1.3 세 가지 목적 모드

모든 프로젝트에는 목적 (purpose)이 있으며, 이는 사이드바 1 절 (**Project**)에서 프로젝트를 생성할 때 선택되고 이후에는 고정된다. 활성 프로젝트의 목적에 따라 어떤 도구와 인터페이스 요소가 나타나는지가 결정된다. 다른 모드로 작업하려면 해당 목적을 가진 프로젝트를 새로 만들거나 열어야 한다. 세 가지 모드는 다음과 같다:

- **General(범용)** — 모든 유형의 모집단 또는 데이터셋에 대한 GAM 회귀분석이다. 기본 모드이며, 분야별 추가 기능 없이 GAM 작업 흐름 전체를 제공한다.
- **For Appraisal(감정평가용)** — 부동산 감정평가에 맞춘 GAM 회귀분석이다. 대상 부동산 처리, 특수 열 지정, 잔차 제약 접근법 (Residual Constraint Approach, RCA) 등 단일 부동산 평가에 특화된 기능을 추가한다.
- **Market Area Analysis(시장지역 분석)** — 시장지역 연구에 맞춘 GAM 회귀분석이다. 정의된 시장 내 부동산 집단을 분석하기 위한 기능을 추가하며, 대상 행 처리는 선택 사항이다.

세 모드 모두에서 핵심 모델링 엔진은 동일하다. 언제나 `mgcv::gam()`을 통해 GAM을 적합한다. 목적 설정은 어떤 부가 도구와 인터페이스 요소를 사용할 수 있는지만 제어한다.

1.4 부동산 전용 기능

For Appraisal 또는 **Market Area Analysis**를 선택하면 mgcvUI는 부동산 분석을 위해 설계된 여러 기능을 활성화한다:

- **특수 열 지정** — 각 예측변수에는 `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions`, `display_only` 같은 특수 역할을 지정할 수 있다. 이 지정은 적합 및 출력 과정에서 해당 열을 어떻게 다룰지를 결정한다.
- **Sale Age 열** — 어떤 열이 `contract_date` 로 지정되고 기준일 (Effective Date) 이 제공 되면 `mgcvUI` 는 `sale_age` 열 (매매일과 기준일 사이의 일수) 을 계산하여 예측변수로 대체한다.
- **RCA 계산 (감정평가 모드 전용)** — 감정평가 모드에서는 모형 적합 후 `mgcvUI` 가 잔차 제약 접근법 (RCA) 출력을 계산할 수 있다. 이는 비교사례별 조정치, 순/총 조정 요약, 그리고 대상 부동산의 조정된 매매가격을 산출한다.

도움말: *General* 목적 모드는 금융, 과학, 공학, 부동산 등 어떤 데이터셋에도 사용할 수 있다. 감정평가 모드와 시장 모드는 부동산 전문가를 위한 편의 기능을 더할 뿐이다.

1.5 시작하기

`mgcvUI` 를 사용하려면:

1. R 에서 **earthUI** 를 설치한다: `install.packages("earthUI")` — `mgcv` 루틴이 함께 포함되어 있다. (개발 버전과 선택적 구성 요소에 관해서는 이번 호 `earthUI` 논문의 “Installing earthUI and vProlog” 절을 참조하라.)
2. **응용 프로그램을 실행한다:** R 에서 `earthUI::launch_mgcv()` 를 실행한다. 앱은 웹 브라우저에서 포트 7880 으로 열린다.
3. 사이드바 1 절에서 프로젝트를 만들거나 연다. 프로젝트는 목적 (General, For Appraisal, Market Area Analysis) 과 작업 위치 (국가, 주, 카운티, 도시) 및 입출력 폴더를 설정한다. 프로젝트는 공유 `regProj` 루트 아래에 저장되며 형제 앱인 `earthUI`, `glmnetUI` 와 공유된다. 다른 목적을 가진 프로젝트를 여는 것이 모드를 전환하는 방법이다.
4. 2 절에서 활성 프로젝트의 입력 폴더에 있는 CSV 또는 Excel 파일을 선택해 데이터를 가져온다. 해당 폴더에 파일을 추가한 뒤 **Refresh** 를 클릭하면 목록에 표시된다.
5. **earthUI** 에서 가져오기 (선택 사항) — 3 절에서는 `earthUI` 결과 `.rds` 파일을 가져와 `earth` 가 발견한 매듭점으로 GAM 을 초기화할 수 있다. `earthUI` 결과가 없다면 이 단계는 건너뛴다.
6. **변수를 구성한다** — 목표변수와 예측변수를 선택하고, 데이터 유형을 설정하며, 특수 열 역할을 지정하고, 변수를 요인 (factor) 또는 선형으로 표시한다.
7. **mgcv** 모수를 설정한다 — `family`, `method`, `gamma`, 기저 유형, `k`, 텐서 유형, 상호작용 및 기타 옵션을 지정한다.
8. **모형을 적합한다** — “Fit Mgcgv GAM Model” 을 클릭하고 메인 패널에서 결과를 검토한다.
9. **내보낸다** — 예측값을 Excel 로 내려받거나, Quarto 보고서를 생성하고 변환하거나, (감정평가 모드에서) RCA 조정치와 매매사례 그리드를 계산한다.

설정은 SQLite 데이터베이스에 자동으로 저장되며, 동일한 입력 파일을 다시 불러올 때 복원된다.

2 MLS 입력 데이터 요건

부동산 감정평가 및 시장분석 작업 흐름에서 입력 데이터는 보통 다중 매물 정보 서비스 (Multiple Listing Service, MLS) 내보내기 파일에서 나온다. 이 장에서는 예상되는 파일 구조와 mgcvUI 가 사용할 수 있는 열을 설명한다.

2.1 파일 형식과 구조

mgcvUI 는 CSV 와 Excel(.xlsx, .xls) 파일을 받아들인다. 가져오기 시 열 이름은 자동으로 snake_case 로 변환된다. 예를 들어 “Living SqFt” 는 living_sqft 로, “Contract Date” 는 contract_date 로, “Sale Price” 는 sale_price 로 바뀐다. 이 정규화는 작업 흐름 전반에서 열 참조를 일관되게 유지해 준다. 가져오기에 사용되는 CSV 구분자와 소수점 기호는 로케일 설정에 따라 결정된다 (3 장 “Locale & Regional Settings” 참조).

데이터 파일은 부동산 한 건이 한 행, 속성 하나가 한 열인 평면형 표여야 한다. 파일의 첫 행에는 반드시 열 머리글이 있어야 한다.

2.2 감정평가 모드에 필요한 열

mgcvUI 는 어떤 열 조합에서도 작동하지만, 완전한 감정평가 작업 흐름을 위해 다음 열이 있으면 유리하다:

열	특수 유형	용도
Sale Price	(목표변수)	모형의 반응변수
Contract Date	contract_date	sale_age(기준일로부터의 일수) 계산에 사용
Listing Date	listing_date	DOM 열이 없을 때 계약일과 함께 DOM 계산에 사용
Days on Market	dom	시장 노출 일수, 내보내기에 표시됨
Concessions	concessions	매매 양여 조건, 순매매가 = 매매가격 - 양여액
Living Area (SF)	living_area	단위면적당 잔차 (residual_sf, cqa_sf) 계산을 가능하게 함
Lot Size	lot_size	대지 규모 열
Site Dimensions	site_dimensions	대지 규모와 함께 묶임
Latitude	latitude	소수점 3 자리로 반올림
Longitude	longitude	소수점 3 자리로 반올림
Area ID	area	시장지역/근린 식별자
Actual Age	actual_age	부동산의 실제 연수
Effective Age	effective_age	부동산의 유효 연수
Address	display_only	내보내기에는 표시되고 모형에서는 제외됨

스프레드시트의 열 이름은 외국어여도 무방하다. “특수” 이름이 영어인 이유는 R 프로그램이 해당 열을 특별히 처리할 수 있도록 하기 위해서이다. 그 밖의 경우에는 사용자가 부여한 열 이름이 회귀 모형, 그래프, 그리고 감정평가를 수행할 때는 출력 보고서에 그대로 표시된다.

모든 열이 필수는 아니다. mgcvUI 는 상황에 맞추어 작동하며, 어떤 열이 없으면 해당 기능만 생략된다. 다만 부동산 가격 모형에서는 다음 열들을 적극 권장한다:

1. **Sale Age** — 계약 매매일과 기준일 사이의 일수. 여러 해에 걸친 매매 이력을 사용할 경우 sale_age 가 핵심 역할을 하는 경우가 많다.
2. **Living Area** — “Living Sqft”, “GLA”(gross living area, 총 거주면적) 등의 이름으로도 불린다.
3. **Total Bath Count** — 전체, 1/4, 1/2, 3/4 욕실을 모두 합한 개수.
4. **Garage Bays** 또는 **Garage Area** — 차고 주차 대수 또는 면적.
5. **Lot Size** — 부동산의 토지 면적.
6. 지리적 가격 변동을 반영하기 위한 **Longitude**, **Latitude**, 그리고 가능하다면 **Area ID**.

2.3 특수 열 명명 규칙

mgcvUI 는 열 이름이 아니라 특수 유형 지정으로 열을 식별한다. MLS 내보내기 파일에서 열 이름은 원하는 대로 지어도 된다. 중요한 것은 변수 구성 표 (6 장) 에서 올바른 특수 유형을 지정하는 일이다.

2.4 데이터 품질과 완전성

- **결측값 (NA):** 예측변수나 목표변수 열에 NA 가 있는 행은 적합 전에 자동으로 제거된다. mgcvUI 는 모형에 사용되는 열만 골라 데이터를 부분 선택하므로, 사용하지 않는 열의 NA 는 행 삭제를 일으키지 않는다.
- **수치형 열:** 매매가격, 거주면적, 대지 규모 등과 같은 항목은 반드시 수치형이어야 한다.
- **요인형 열:** 텍스트, TRUE/FALSE, R 요인 (factor) 열은 자동으로 범주형으로 처리되며, 모형에는 요인으로만 들어갈 수 있다. 수치형 열 (예: 수치로 된 지역 코드나 양식 코드) 은 **Factor** 체크박스를 선택했을 때에만 요인이 된다. mgcvUI 는 수치의 값 범위만 보고 범주형이라고 추론하지 않으므로, `bath_count` 같은 이산 수치형 예측변수는 연속형 (평활 처리) 으로 유지된다. 특수 역할 지정은 요인 여부 결정과 무관하다.

도움말: 가져오기 후 예측변수 표의 NA 열을 살펴보자. 결측값이 많은 열은 행 삭제를 유발할 수 있다. NA 가 50% 이상인 열은 빨간색으로 표시되고 자동으로 제외된다.

2.5 대상 행의 위치

Appraisal 모드에서는 1 행이 반드시 대상 부동산이어야 한다. 나머지 행은 모두 비교 매매사례이다. 대상 행은 모형 적합에서 제외된다. 적합 후에도 모형은 대상 행에 대한 예측값을 생성한다.

Market Area Analysis 모드에서는 대상을 1 행에 두는 것이 선택 사항이다.

General 모드에는 특별한 행 처리가 없으며 모든 행이 동등하게 취급된다.

3 범용 모드

3.1 개요

범용 (General Purpose) 모드는 mgcvUI 를 실행했을 때의 기본 모드이다. 이 모드는 부동산에 국한되지 않고 어떤 데이터셋에 대해서도 완전한 GAM 작업 흐름을 제공한다. 과학 데이터, 재무 분석, 공학 연구를 비롯해 평활한 비선형 관계가 예상되는 모든 회귀 문제에 mgcvUI 를 사용할 수 있다.

General 모드에서는 부동산 전용 기능 (특수 열, 매매 경과일, RCA) 이 인터페이스에서 생략된다. 사이드바는 변수 선택, 모수 설정, 모형 적합, 내보내기에 집중하도록 간소화된다.

3.2 사이드바 작업 흐름

사이드바는 번호가 매겨진 접이식 절들로 구성되어 프로젝트 선택부터 보고서 내보내기까지를 안내한다. 대부분의 절은 활성 프로젝트가 열려 있어야 나타나며, 몇몇 절은 프로젝트의 목적에 따라 달라진다. 출력 폴더는 활성 프로젝트에서 유도되며 별도의 출력 폴더 입력란은 없다.

1. Project(프로젝트) — 프로젝트를 만들거나 연다. 프로젝트는 목적과 작업 위치 (국가, 주, 카운티, 도시), 그리고 입력 폴더 (<os>_in/) 와 출력 폴더를 설정한다. 프로젝트는 공유 regProj 루트 (기본값 ~/regProj, REGPROJ_ROOT 환경변수나 Settings 드롭다운으로 변경 가능) 아래에 저장되고 형제 앱인 earthUI, glmnetUI 와 공유되므로, 세 앱의 모형이 나란히 놓이게 된다. 전환하려면 **Close Project** 를 클릭한다.

2. Import Data(데이터 가져오기) — 활성 프로젝트의 입력 폴더에서 CSV 또는 Excel 파일을 선택한다. Finder/탐색기로 폴더에 파일을 추가한 뒤 **Refresh** 를 클릭하면 목록에 표시된다. 시트가 여러 개인 Excel 파일에는 시트 선택기가 나타난다. 열 이름은 자동으로 snake_case 로 변환된다.

3. Import from earthUI (optional)(earthUI 에서 가져오기, 선택 사항) — earthUI 결과 .rds 파일을 가져와 earth 가 발견한 매듭점 위치 (및 변수 선택 상태) 로 GAM 을 초기화한다. 이 단계는 선택 사항이며, 가져올 earthUI 결과가 없다면 건너뛰는다.

4. Variable Configuration(변수 구성) — 목표변수 선택기, 반응 변환 (none, log, log10), 그리고 Include, Factor, Linear 체크박스가 있는 예측변수 표로 구성된다. Special 열은 Appraisal 과 Market 모드에서만 나타난다. 자세한 내용은 6 장을 참조하라.

5. MgcV Call Parameters(mgcV 호출 모수) — 모수 프리셋, family, method, gamma, 교차검증, select, 기저 유형, k, 텐서 유형, 상호작용 행렬, 고급 모수 등 모든 모형 설정을 다룬다. 전체 모수 참조는 7 장에 있다.

6. Fit MgcV GAM Model(mgcV GAM 모형 적합) — 모형을 실행하는 버튼이다. 결과는 오른쪽 탭에 표시된다.

7. Download Output(출력 내려받기) — 예측값, 잔차, CQA 점수, 변수별 기여도를 Excel 파일로 프로젝트의 출력 폴더에 내보낸다. 모형을 적합한 뒤에는 모든 목적 모드에서 사용할 수 있다.

8. Calculate RCA Adjustments & Download(RCA 조정치 계산 및 내려받기) — 각 비교사례에 대한 변수별 RCA 조정치를 계산하고 CQA 를 통해 대상의 잔차를 보간한다. Appraisal 과 Market 모드에서 적합 후에만 나타난다.

9. Generate Sales Grid & Download(매매사례 그리드 생성 및 내려받기) — 총조정률을 기준으로 비교사례를 추천하면서 거래사례비교 그리드를 생성한다. Appraisal 모드에서만 나타난다.

10. Generate Quarto Report(Quarto 보고서 생성) — 모든 그림과 표를 포함한 자립형 Quarto(.qmd) 보고서 묶음을 프로젝트의 출력 폴더에 작성한다. 모형을 적합한 뒤 사용할 수 있다.

11. Convert Quarto Report(Quarto 보고서 변환) — 생성된 .qmd 묶음을 HTML, Word, PDF 로 렌더링한다. PDF 는 LaTeX 설치가 필요하며, 설치가 감지되지 않으면 PDF 옵션이 숨겨진다. 13 장을 참조하라.

다른 모드에서는 목적별 단계가 생략되고 나머지 단계의 번호가 다시 매겨진다. Market 모드에서는 매매사례 그리드 단계가 없고, General 모드에서는 RCA 와 매매사례 그리드 단계가 모두 없다.

3.3 earthUI 에서 가져오기 (선택 사항)

사이드바 3 절 **Import from earthUI**에서는 earthUI 결과 .rds 파일을 가져올 수 있다. 이를 통해 earth-mgcv 파이프라인이 가능해진다. 즉 earth 가 데이터로부터 찾아낸 매듭점 위치가 GAM 평활항의 기준점이 된다.

earthUI 결과를 가져오면 mgcvUI 는 다음을 수행한다:

- 목표변수, 결정계수 (R-squared), 항의 개수, 변수별 매듭점 위치를 보여 주는 요약을 표시한다.
- “Earth Pipeline” 모수 프리셋 (삼차 회귀 스플라인 기저, $\gamma = 1.4$) 을 자동으로 선택한다.
- 목표변수를 earth 의 목표변수와 일치시킨다.
- earth 의 예측변수에 대해 Include 를 미리 선택한다.
- earth 의 범주형 변수에 대해 Factor 를, earth 의 선형 예측변수에 대해 Linear 를 미리 선택한다.
- earth 가 탐지한 상호작용을 미리 채운다 (상호작용 행렬에서 노란색으로 강조되며 수동 변경이 잠긴다).
- earth 의 매듭점을 CSV 파일로 내려받을 수 있는 **Export Knots CSV** 버튼을 제공한다.

그러면 변수 구성 표에는 변수 선택 (Include, Linear, Factor) 과 스플라인 매듭점이 earth 모형에서 초기값으로 설정되었다는 안내가 나타난다. glmnet 적합이 earth 의 기저에 묶여 있는 glmnetUI 와 달리, mgcvUI 는 현재 선택 상태를 바탕으로 GAM 을 적합한다. earth 는 출발점만 제공할 뿐이며 모든 제어 항목은 계속 편집할 수 있다.

Clear 버튼은 가져온 earth 데이터를 제거하고 독립 실행 모드로 되돌린다.

3.4 메인 패널 탭

데이터를 가져오면 메인 패널에 다음 탭들이 나타난다 (모형에 의존하는 탭은 적합 후에 채워진다):

탭	내용
Data	가져온 데이터의 대화형 DataTable. 감정평가 및 시장 모드에서는 대상 부동산 표와 비교 매매사례 표로 분리된다.
Equation	MathJax 로 렌더링된 전체 GAM 식과 평활 함수 정의 표.
Summary	주요 지표 (R^2 , $CV R^2$, 설명된 이탈도, AIC, n, 평활항 개수) 와 평활항/모수항 표.
Variable Importance	F-통계량 (평활항) 과 $ t $ 값 (모수항) 의 가로 막대그래프 및 순위 표.
Contribution	신뢰 구간 띠와 earth 매듭점 표식을 포함한 대화형 plotly 평활 부분효과 곡선.
Correlation	수치형 예측변수 간 상관 히트맵 (적합 전에도 사용 가능).
Diagnostics	잔차 대 적합값, Q-Q 그림, 히스토그램, 반응변수 대 적합값 (gratia 사용). 실측값 대 예측값 산점도.
RCA Adjustments	RCA 분석 히스토그램 (감정평가/시장 모드에서 계산 후).
Anova	모수항과 평활항을 함께 담은 ANOVA 표.
Mgcv Output	원시 텍스트: 소요 시간, 식, 모형 요약, family, 공곡선성.
Sign Check	earth 와 GAM 의 방향 일치 여부 점검 (earth 를 가져온 경우).
Concurvity	색상으로 구분된 전체 및 쌍별 공곡선성 행렬.
Basis Check	gam.check() 의 기저 차원 적정성 검증.

3.5 설정의 유지

mgcvUI 는 입력 파일 이름을 키로 하여 설정을 SQLite 데이터베이스에 자동 저장한다. 같은 파일을 다시 불러오면 목표변수 선택, 예측변수 체크박스, 데이터 유형, mgcv 모수, 상호작용 행렬 등 모든 설정이 복원된다. 목적은 전역 설정이 아니라 활성 프로젝트의 속성이며, 마지막으로 사용한 입력 파일은 프로젝트별로 기억되어 프로젝트를 다시 열 때 자동으로 불러와진다.

3.6 다크 모드

오른쪽 위 모서리의 달/해 아이콘을 클릭하면 Nord Light 와 Nord Dark 테마를 전환할 수 있다. 테마 설정은 localStorage 에 저장되어 세션이 바뀌어도 유지된다. 모든 UI 요소가 선택한 테마에 맞추어 바뀐다.

3.7 로케일과 지역 설정

mgcvUI 는 국가 기반 로케일 체계를 통해 국제적인 숫자 및 CSV 형식 관례를 지원한다. 제목 표 시줄의 **Settings** 드롭다운에는 30 개국 이상을 지원하는 **Country** 와 **Paper** 선택기가 있다. 각 프리셋은 다음을 설정한다:

- **CSV 구분자** — 미국/영국/일본은 쉼표 (,), 유럽 대부분은 세미콜론 (;).
- **소수점 기호** — 마침표 (.) 또는 쉼표 (,).
- **천 단위 구분자** — 쉼표 (미국/영국/일본), 마침표 (독일/이탈리아/스페인), 공백 (핀란드/프랑스/폴란드/발트 3 국), 또는 아포스트로피 (스위스).
- **용지 크기** — Letter(미국/캐나다/멕시코) 또는 A4(그 외 지역).

Save 를 클릭하면 로케일 설정이 전역으로 저장된다. Import Data 절에도 관례가 다른 일회성 가져오기를 위한 파일별 **Locale** 선택기가 있다. 같은 Settings 드롭다운에는 **regProj** 루트 폴더 입력란이 있으며, 이는 mgcvUI, earthUI, glmnetUI 가 함께 사용하는 공유 프로젝트 트리의 위치를 바꾼다.

도움말: 그림 축과 기울기 라벨의 숫자 형식은 로케일에 맞추어 자동으로 조정된다. 독일 로케일은 천 단위에 마침표 (200.000), 핀란드는 공백 (200 000), 스위스는 아포스트로피 (200'000) 를 사용한다. 통화 기호는 표시되지 않으며, *mgcvUI* 는 통화에 중립적이다.

4 감정평가 모드

활성 프로젝트의 목적이 **For Appraisal** 일 때 mgcvUI 는 단일 부동산 평가에 맞추어 스스로를 구성한다. 3 장에서 설명한 모든 기능은 그대로 사용할 수 있으며, 이 장에서는 감정평가 전용으로 추가되는 부분만 다룬다.

4.1 대상 행의 처리

감정평가 모드에서는 데이터셋의 1 행이 대상 부동산이고 나머지 행은 모두 비교 매매사례이다. 입력 파일은 그에 맞게 구성되어야 한다 (2 장 참조). 대상의 매매가격은 비워 두거나 아무 값이나 넣어도 된다. mgcvUI 가 적합 과정에서 이를 자동으로 NA 로 처리하기 때문이다.

적합 후 모형은 대상 행에 대한 예측값을 생성하며, 이것이 RCA 조정 작업 흐름의 출발점이 된다.

4.2 기준일과 매매 경과일

감정평가 및 시장 모드에서는 변수 구성 절에 **Effective Date**(기준일) 입력란이 나타난다 (기본값은 오늘 날짜). Special 드롭다운에서 어떤 열을 `contract_date` 로 지정하면 mgcvUI 는 `sale_age` 열, 즉 각 매매의 계약일과 기준일 사이의 정수 일수를 계산한다. 이 열은 예측변수로 추가된다.

기준일이 바뀌면 `sale_age` 는 자동으로 다시 계산된다.

4.3 특수 열 지정

감정평가 및 시장 모드에서는 변수 구성 표의 각 예측변수마다 **Special** 드롭다운이 나타난다. 특수 유형과 그 효과에 관한 전체 참조는 6 장에 있다.

4.4 RCA 조정 개요

Calculate RCA Adjustments & Download 버튼 (감정평가 및 시장 모드에서 적합 후 표시됨) 은 대상 대비 각 비교사례에 대해 시장에서 도출된 조정치를 계산한다. 전체 RCA 작업 흐름은 11 장에서 설명한다.

도움말: 대상에 부여하는 *CQA* 점수는 잔차 분포 중 얼마만큼을 대상에 귀속시킬지를 좌우한다. 점수 5.00 은 대상을 비교사례들의 중앙값에 위치시킨다.

5 시장지역 분석 모드

활성 프로젝트의 목적이 **Market Area Analysis** 일 때 mgcvUI 는 감정평가 모드와 동일한 부동산 전용 기능 (특수 열, 매매 경과일) 을 제공하지만, 단일 대상의 평가보다는 부동산 집단의 분석에 초점을 둔다.

5.1 감정평가 모드와의 차이

- 대상 행 처리가 유연하다 — 시장 모드에서는 1 행이 자동으로 대상 부동산으로 취급되지 않는다.
- 매매사례 그리드가 없다 — “Generate Sales Grid & Download” 단계를 사용할 수 없다.

5.2 시장 모드를 사용하는 경우

시장지역 분석 모드는 다음과 같은 경우에 적합하다:

- 비선형적인 가치 결정 요인을 파악하기 위해 근린이나 시장지역에 대한 GAM 모형을 구축할 때
- 면적, 연수, 대지 규모, 위치 같은 변수가 부동산 집단 전반의 매매가격에 어떤 영향을 주는지 분석할 때
- 시장 상황 분석이나 근린 확정을 뒷받침할 근거를 준비할 때
- 특수 열 기능 (매매 경과일) 은 사용하되 매매사례 그리드 작업 흐름은 필요하지 않은 데이터셋을 다룰 때

도움말: 시장 모드는 특수 열 기능은 쓰고 싶지만 단일 부동산 평가 작업 흐름은 필요 없는 일반적인 부동산 회귀분석에도 유용하다.

6 변수 선택

사이드바 4 절 **Variable Configuration** 은 어떤 열을 모형에 참여시킬 것인지, 그리고 그 열을 어떻게 다룰 것인지를 정하는 곳이다.

6.1 목표변수

4 절 상단의 **Target (response) variable** 드롭다운에는 데이터셋의 모든 수치형 열이 나열된다. 그중 하나를 반응변수 (예: 매매가격) 로 선택한다. 목표변수로 지정된 열은 예측변수 목록에서 자동으로 제외된다.

6.2 반응변수 변환

목표변수 선택기 아래의 **Response Transform** 드롭다운은 세 가지 선택지를 제공한다:

- **None (raw)** — 반응변수를 원래대로 사용한다. 예측변수와 반응변수의 관계가 원래 척도에서 대체로 가법적일 때 적절하다.
- **Log (natural)** — 적합 전에 자연로그 변환을 적용한다. 역변환 시 조정치가 절대액 (\$) 이 아니라 비율 (%) 이 된다. 1 만 달러의 조정이 가격 수준에 따라 전혀 다른 의미를 갖는 매매가격에 특히 유용하다.
- **Log10** — 밑이 10 인 로그 변환을 적용한다.

도움말: 로그 변환은 시점 조정을 절대액 (\$) 대신 비율 (%) 로 만든다. 가격대가 넓게 분포한 부동산 모형에서는 로그 변환이 더 안정적인 잔차를 낳는 경우가 많다.

로그 변환을 선택하면 반응변수의 ≤ 0 값은 자동으로 걸러진다. 모든 출력 (예측값, 기여도, 잔차) 은 자동으로 원래 척도로 역변환된다.

6.3 예측변수 표

목표변수 선택기 아래에는 나머지 모든 열을 나열하는 스크롤 가능한 표가 있다. 체크박스 열은 공간 절약을 위해 세로로 회전된 머리글을 사용한다. 열 순서는 다음과 같다:

열	설명
Variable	열 이름 (고정폭 글꼴, 길면 말줄임표로 잘림). 테두리가 있는 칸.
Type	데이터 유형 드롭다운: numeric, integer, character, factor, logical, Date, POSIXct. 테두리가 있는 드롭다운.
Include	체크박스 — 이 열을 모형의 예측변수로 포함한다. 세로 머리글.
Factor	체크박스 — 범주형 요인으로 취급한다 (수준마다 별도의 계수를 만든다). 세로 머리글.
Linear	체크박스 — 선형 관계를 강제한다 (평활 없이 βx 만). 세로 머리글.
Special	드롭다운 (감정평가/시장 모드 전용) — 아래의 특수 열 유형 참조를 보라. 테두리가 있는 드롭다운.
NAs	결측값의 개수와 비율. 색상으로 구분됨: 빨강 ($\geq 50\%$), 주황 ($\geq 20\%$), 회색 (> 0). 테두리가 있는 칸.

Factor 대 평활 대 Linear: 기본적으로 포함된 수치형 변수에는 평활항 $f(x)$ 가 부여된다. **Factor** 를 선택하면 범주형 항 (수준마다 계수 하나) 이 만들어진다. **Linear** 를 선택하면 평활 대신 단순 선형항 βx 가 만들어진다. Factor 와 Linear 를 모두 표시한 변수는 Factor 로 취급된다.

6.4 데이터 유형 탐지와 재지정

mgcvUI 는 가져오기 시 데이터 유형을 자동으로 탐지한다. 수치형, 정수형, 논리형, 요인형, 날짜형 열이 인식된다. **Type** 드롭다운을 바꾸면 탐지 결과를 임의로 재지정할 수 있다. 유형을 바꾸면 해당 열이 모형에서 다루어지는 방식도 달라진다.

6.5 특수 열 유형 참조

감정평가 및 시장 모드에서 **Special** 드롭다운은 다음 선택지를 제공한다:

가중:

- **weight** — 관측치 가중치 열 (하나만 지정 가능. 가중치가 0 인 행은 적합에서 제외됨)

날짜 및 시간 유형:

- **contract_date** — 기준일로부터 **sale_age** 를 자동으로 계산하도록 한다
- **sale_age** — 이미 계산되어 있는 매매 경과일 열 (기준일로부터의 일수) 을 식별하며, 그대로 사용된다
- **listing_date** — 시장 노출 일수를 계산할 때의 대체 수단으로 사용된다
- **dom** — 시장 노출 일수 (Days on Market) 열을 식별한다

레코드 유형:

- **sale_type** — 대상과 비교 매매사례를 구분하는 레코드 유형 열을 식별한다

금액 유형:

- **concessions** — 매매 양여 조건 (매도인 크레딧, 매수인 유인책 등) 을 식별한다

규모 및 위치 유형:

- **latitude** — 값이 자동으로 소수점 3 자리로 반올림된다
- **longitude** — 위도와 동일한 반올림 처리
- **area** — 시장지역 또는 근린 식별자
- **living_area** — 단위면적당 잔차 계산 (**residual_sf**, **cqa_sf**) 을 가능하게 한다
- **lot_size** — 대지 규모 열
- **site_dimensions** — 대지 규모와 함께 묶인다

연수 유형:

- **actual_age** — 부동산의 실제 연수 열
- **effective_age** — 부동산의 유효 연수

표시 유형:

- **display_only** — 해당 열은 Excel 내보내기에는 포함되지만 모형 적합에서는 완전히 제외된다. 주소 항목, MLS 번호, 필지 ID 등의 참조 데이터에 사용한다.

도움말: *display_only* 로 지정된 열은 데이터셋에 그대로 남아 *Excel* 내보내기에 나타나지만 예측변수 표에서는 완전히 제외된다. *ID* 열, 주소, 기타 참조 항목에 사용하라.

7 모수 선택

사이드바 5 절 **Mgcv Call Parameters** 는 GAM 의 모든 설정 항목에 접근할 수 있게 해 준다. 각 모수에는 설명 풍선말이 달린 파란색 도움말 아이콘 (?) 이 있다.

7.1 모수 프리셋

상단의 드롭다운에는 흔한 작업 흐름에 맞는 합리적 기본값을 설정해 주는 두 가지 프리셋이 있다:

프리셋	기저	설정
Standalone (탐색)	tp (박판)	k = 자동 (10), gamma = 1.2, select = TRUE
Earth Pipeline (정교화)	cr (삼차 회귀)	k = 자동 (10), gamma = 1.4, select = TRUE

earthUI 매듭점을 가져오면 **Earth Pipeline** 프리셋이 자동으로 선택된다. earth 매듭점을 통합하려면 정확한 매듭점 위치를 지정할 수 있는 삼차 회귀 스플라인 (**cr**) 기저가 필요하다.

7.2 Family(분포족)

반응변수의 분포족을 선택한다:

Family	용도
gaussian	연속형 반응변수 (예: 매매가격). 가장 흔하다.
Gamma	분산이 평균에 비례하는 양의 연속형 반응변수.
poisson	계수 자료 (예: 매매 건수).
binomial	이항 결과.
inverse.gaussian	오른쪽으로 크게 치우친 양의 자료.

7.3 Method(방법)

평활 모수 추정 방법이다:

Method	설명
REML	제한 최대가능도 (기본값, 권장). 과적합에 가장 강건하다.
GCV.Cp	일반화 교차검증. 다소 더 복잡한 모형을 고르는 경향이 있다.
ML	최대가능도. REML 과 비슷하지만 평활이 부족해질 수 있다.
P-REML	피어슨 REML 변형.
P-ML	피어슨 ML 변형.
GACV.Cp	일반화 근사 교차검증.

7.4 Gamma(평활 승수)

gamma 모수는 평활 기준에서 유효자유도에 곱해져 더 매끄러운 (덜 굴곡진) 적합을 유도한다. 기본값은 1.2 이다.

값	효과
1.0	표준 평활

값	효과
1.2-1.4	기본보다 약간 더 매끄러움 (권장)
2.0 이상	훨씬 더 매끄러움 — 잡음이 많은 자료나 소표본에 적합

gamma 값이 클수록 복잡도에 더 무거운 벌점을 부과하여 과적합을 막아 준다. Earth Pipeline 프리셋은 earth 의 매듭점을 정교화할 때 추가 평활을 위해 1.4 를 사용한다.

7.5 교차검증

체크하면 (기본값) mgcvUI 는 적합 후 10 겹 교차검증을 실행하여 CV 결정계수를 계산한다. 이는 표본 외 예측력에 대한 정직한 추정치를 제공한다.

7.6 Select(수축)

체크하면 (기본값) 각 평활항에 그 항을 0 으로 수축시킬 수 있는 추가 벌점이 더해진다. 이로써 자동 변수 선택이 가능해지며, 중요하지 않은 평활항은 사실상 모형에서 제거된다. gam() 의 select = TRUE 로 구현된다.

7.7 기본 기저

평활항에 사용할 스플라인 기저 함수 유형이다:

기저	전체 명칭	설명
tp	박판 회귀 스플라인	기본값. 등방적이며 매듭점 배치가 필요 없다. 범용으로 무난한 선택이다.
cr	삼차 회귀 스플라인	매듭점을 명시적으로 배치할 수 있다. earth 매듭점 통합에 필수적이다.
ps	P-스플라인	차분 벌점을 갖는 등간격 B-스플라인 기저. 계산 효율이 좋다.
bs	B-스플라인	유연한 B-스플라인 기저.

7.8 기본 k(기저 차원)

기저 차원 k 는 각 평활항이 가질 수 있는 최대 굴곡 정도를 제어한다. 0(기본값) 은 “자동” 을 뜻하며, mgcvUI 는 상황에 따라 $k = 10$ 또는 earth 매듭점의 개수를 사용한다.

- 작은 $k(3-5)$: 거의 직선에 가까운 매우 매끄러운 곡선.
- 중간 $k(8-15)$: 적당한 유연성 (대개 이 정도면 충분하다).
- 큰 $k(20 이상)$: 매우 유연함 — 강한 정칙화가 없으면 과적합 위험이 있다.

도움말: Summary 탭의 유효자유도 (EDF) 는 각 평활항이 실제로 사용하는 자유도가 얼마인지를 알려 준다. EDF 가 $k - 1$ 에 가깝다면 기저가 너무 작을 수 있으므로 k 를 늘려라. EDF 가 k 보다 훨씬 작다면 평활항에 벌점이 잘 작동하고 있는 것이다.

7.9 텐서 유형

연속형 변수 간 상호작용에는 두 가지 텐서곱 유형을 사용할 수 있다:

유형	함수	설명
ti	ti(x1, x2)	텐서 상호작용 — 상호작용 효과만 모형화하고 주효과는 일변량 $s()$ 항이 따로 담당한다. 해석 가능성 면에서 권장된다.
te	te(x1, x2)	텐서곱 — 주효과를 포함한 결합 효과 전체를 모형화한다. RCA 조정을 위해 분해하기가 더 어렵다.

7.10 허용 상호작용

접이식 **Allowed Interactions** 절에는 포함된 모든 평활 (비선형, 비요인) 예측변수에 대한 삼삼각 체크박스 행렬이 표시된다. 각 체크박스는 해당 변수 쌍 사이의 텐서곱 평활을 활성화한다.

- **Allow All / Clear All** — 일괄 전환 버튼.
- 변수 이름 클릭 — 그 변수에 관한 모든 상호작용을 전환한다.
- 변수 이름 오른쪽 클릭 — “Block from main effect” — 해당 변수를 상호작용 전용 (일변량 평활 없음) 으로 표시한다.

earthUI 매듭점을 가져오면 earth 가 탐지한 상호작용이 미리 선택되고 잠긴다 (노란 배경으로 강조됨). earth 모형이 degree = 1(상호작용 없음) 을 사용했다면 안내 메시지가 표시된다.

주의

텐서 상호작용을 활성화하면 모형 복잡도가 크게 증가한다. 해당 분야의 지식이나 earth 가 탐지한 근거가 뒷받침될 때에만 상호작용을 사용하라. 그 결과로 나오는 조정치가 합리적이고 설명 가능한지 반드시 확인하라.

7.11 요인별 평활 상호작용

요인 변수와 평활 변수가 모두 포함되면 별도의 **Factor-by-Smooth Interactions** 행렬이 나타난다. 각 체크박스는 $s(x, by = factor)$ 항, 즉 요인 수준마다 별개의 평활 곡선을 만든다.

7.12 고급 모수

접이식 “Advanced” 절에서는 추가 설정을 노출한다:

- **Optimizer** — 평활 모수 추정 알고리즘: `outer/newton`(기본값), `outer/bfgs`, 또는 `efs`(확장 Fellner-Schall).
- **Scale** — 척도 모수. 0(기본값) 은 데이터로부터 추정함을 뜻한다. 척도가 고정된 모형에서는 알려진 값을 지정한다.
- **Discrete covariate method** — 대용량 데이터셋을 위한 빠른 이산화 적합 (`discrete = TRUE`).
- **Threads** — 병렬 스레드 수 (1-32). `discrete = TRUE` 일 때에만 효과가 있다.

8 모형 적합

8.1 Fit 버튼

사이드바 6 절에는 **Fit Mgcgv GAM Model** 버튼이 있다. 이를 클릭하면 현재 설정으로 모형이 실행된다.

8.2 적합 중에 일어나는 일

Fit 을 클릭하면 mgcvUI 는 다음을 수행한다:

1. 설정을 검증한다 — 예측변수가 하나 이상 있고 완전한 행이 10 개 이상인지 확인한다. 완전한 행이 50% 미만이면 경고하고 NA 를 가장 많이 유발하는 열들을 알려 준다.
2. 데이터를 준비한다 — 모형에 필요한 열만 남기고, NA 가 있는 행을 제거하며, (지정된 경우) 가중치를 적용하고, (감정평가 모드에서) 대상 행을 제외하며, 반응변수 변환을 적용한다.
3. GAM 식을 구성한다 — 수치형 예측변수에는 평활항 $s(x, bs="cr", k=10)$ 을, 범주형에는 요인항을, Linear 로 표시된 변수에는 선형항을, 선택된 상호작용에는 텐서곱을 만든다. earth 매듭점은 cr 기저 항의 명시적 매듭점 위치로 삽입된다.
4. 예외 상황을 처리한다 — 상수 변수를 제거하고, 이항 예측변수를 요인으로 변환하며, k 를 고유값 개수에서 1 을 뺀 값으로 제한하고, 매듭점이 더 필요할 때에는 데이터 분위수로 earth 매듭점을 보강한다.
5. GAM 을 적합한다 — 구성된 식, family, method, gamma, select 및 기타 모수와 함께 `mgcv::gam()` 을 호출한다.
6. 교차검증을 실행한다 (활성화된 경우) — 표본 외 결정계수를 계산하는 10 겹 CV.
7. 모든 결과 탭을 갱신한다 — Summary, Equation, Contribution 그림, Diagnostics 를 비롯한 모든 탭이 채워진다.

8.3 백그라운드 적합과 진행 상황

`callr` 패키지를 사용할 수 있으면 모형 적합이 백그라운드 프로세스에서 실행되어 앱이 계속 반응한다. 어두운 테마의 터미널 형태 진행 오버레이에는 다음이 표시된다:

- 경과 초/분을 보여 주는 타이머.
- 색상으로 구분된 추적 로그: 상태는 파랑, CV 겹 진행은 노랑, 오류는 빨강, 결과는 초록.
- 필요할 때 적합 프로세스를 중단하는 **Abort** 버튼.
- 완료 후 나타나는 **Close** 버튼.

`callr` 를 사용할 수 없으면 적합이 동기적으로 실행된다 (적합이 끝날 때까지 앱이 멈춘다).

적합에 성공하면 Fit 버튼에 흰색 체크 표시가 나타나고 상태 표시줄에 “R-sq = X, CV R-sq = X, Dev = X%, AIC = X, n = X” 가 표시된다.

9 결과 탭

9.1 Data(데이터)

가져온 데이터를 가로 스크롤이 가능하고 한 페이지에 15 행씩 표시되는 대화형 DataTable 로 보여 준다. 감정평가 및 시장 모드에서는 탭이 **Subject Property** 표 (1 행) 와 **Comparable Sales** 표 (나머지 행) 로 나뉘며, General 모드에서는 하나의 표만 표시된다.

각 칸은 한 줄로 유지되고 열 너비에 맞추어 잘리므로, 부동산 비고란처럼 긴 자유 서술 항목 이 행을 아래로 늘어뜨리지 않는다. 한 번 클릭하면 행이 선택되고, 어떤 칸이든 두 번 클릭하면 잘리지 않은 전체 내용을 보여 주는 팝업이 열린다.

9.2 Equation(식)

적합된 GAM 모형을 두 부분으로 나누어 표시한다:

모형 식 (MathJax 로 렌더링됨):

- 좌변에는 해당하는 변환 (log, log10) 과 연결함수를 적용한 반응변수가 표시된다.
- 우변에는 절편값, f_i (변수) 형태의 평활항, 수준 개수와 함께 변수별로 묶인 요인항, 그리고 계수값을 포함한 선형항이 표시된다.
- family 와 method 는 식 위에 표시된다.

평활 함수 정의 (Smooth Function Definitions) 표:

- 항의 식 문자열 (예: `s(living_sqft, bs="cr", k=8)`)
- 변수 이름
- 유형 (s, te, ti 또는 linear)
- 기저 함수 유형
- k 값
- earth 매듭점 위치 (가져온 경우)

9.3 Summary(요약)

상단에 여섯 개의 지표 카드가 표시된다:

- **R-squared** — 설명된 이탈도의 비율
- **CV R-squared** — 교차검증 결정계수 (CV 를 활성화한 경우)
- **Dev. Explained %** — 설명된 이탈도의 백분율
- **AIC** — 아카이케 정보기준
- **n** — 사용된 관측치 수
- **Smooths** — 평활항의 개수

카드 아래에는 다음이 표시된다:

- **평활항 표** — 항, EDF(유효자유도), Ref.df, F 통계량, p-값
- **모수항 표** — 항, 추정값, 표준오차, t-값, p-값

9.4 Variable Importance(변수 중요도)

각 모형 항의 상대적 중요도를 보여 주는 가로 막대그래프이다:

- **평활항**: F-통계량 순위 (파란 막대)
- **모수항**: $|t|$ 값 순위 (초록 막대)

그래프 아래의 DataTable 에는 항, 유형, EDF, 통계량, p-값이 통계량 내림차순으로 정렬되어 표시된다.

9.5 기여도 그림

예측값에 기여하는 모든 모형 항에는 대화형 **plotly** 기여도 그림이 있으며, 테두리가 있는 카드 안에 표시된다. 이 탭은 모든 유형의 항에 대한 그림을 보여 준다:

9.5.1 일변량 평활항 — $s(x)$

- 95% 신뢰 구간 띠를 두른 파란 선
- 부분 잔차를 보여 주는 회색 산점 (로그 변환하지 않은 모형에서)
- earth 매듭점 위치의 빨간 파선 수직선 (earth 를 가져온 경우)
- 마우스오버 풍선말: 변수값, 기여도 (\$), 95% 신뢰구간, 변화율 (단위당 기울기)

로그 변환 모형에서는 y 축이 $\bar{y} \times (e^{f(x)} - 1)$ 공식을 통해 달러 기여도로 역변환되어, 곡선을 곧바로 금액 단위로 해석할 수 있다.

위도/경도 변수에서는 기울기가 1 도당이 아니라 0.001 도당으로 표시되는데, 지리적으로 그 편이 더 의미 있는 증분이기 때문이다.

9.5.2 요인별 평활항 — $s(x, \text{by} = \text{factor})$

요인 수준마다 색이 다른 선 하나씩을 한 그림에 겹쳐 표시한다. 예측변수의 효과가 집단 (예: 서로 다른 근린) 에 따라 어떻게 달라지는지 확인하는 데 유용하다.

9.5.3 상호작용항 — $ti(x_1, x_2)$ 및 $te(x_1, x_2)$

2 차원 기여도 곡면을 보여 주는 히트맵이다. 색 척도는 빨강 (음) → 흰색 (0) → 파랑 (양) 이다. 마우스를 올리면 두 변수의 값과 기여량이 함께 표시된다.

9.5.4 모수항 — 선형항과 요인항

- 수치형 선형항: 적합선을 겹쳐 그린 데이터 산점도
- 요인항: 요인 수준별 평균 기여도를 보여 주는 막대그래프

9.6 Correlation(상관)

모든 수치형 예측변수 간의 피어슨 상관을 보여 주는 히트맵 행렬이다. 데이터를 가져온 직후 바로 사용할 수 있으며 모형 적합이 필요 없다.

- 색 그라데이션: 파랑 (-1) → 흰색 (0) → 빨강 (+1)
- 각 칸에 상관계수 값이 표시됨
- 적합 전에 공공선성 (다중공선성의 평활 대응 개념) 을 확인하는 데 유용함

9.7 Diagnostics(진단)

두 가지 진단 시각화를 제공한다:

진단 패널 (`gratia::appraise` 사용): 잔차 대 적합값, Q-Q 그림, 잔차 히스토그램, 반응변수 대 적합값의 네 가지 그림.

실측값 대 예측값 산점도: 45 도 기준 파선과 함께 관측값 대 예측값을 보여 준다. 로그 변환 모형에서는 값이 원래 척도로 역변환된다.

9.8 RCA Adjustments(RCA 조정)

RCA 계산 후에 표시되는 세 개의 히스토그램이다 (감정평가/시장 모드 전용):

- **Residual Adjustment %** — 매매가격 대비 잔차 조정의 백분을 분포
- **Net Adjustment %** — 순조정치의 분포
- **Gross Adjustment %** — 총조정치의 분포

각 히스토그램은 부제에 평균, 중앙값, 표준편차를 표시하고 평균과 중앙값 위치에 파선 기준선을 그린다.

9.9 Anova(분산분석)

GAM 요약의 모수항과 평활항을 합친 통합 ANOVA 표이다. 항, 유형 (모수항 또는 평활항), 그리고 관련된 모든 통계량을 보여 준다.

9.10 MgcV Output(mgcV 출력)

다음은 보여 주는 원시 텍스트 출력이다:

- 적합 소요 시간
- GAM 식
- 모형 출력 (print)
- 전체 `summary(model)` 출력
- family 와 연결함수
- 전체 공곡선성 측도

9.11 Sign Check(부호 점검)

earthUI 매듭점을 가져온 경우, 이 탭은 earth 모형에서의 각 변수 방향과 GAM 에서의 방향을 비교한다:

- **earth_direction**: 증가, 감소 또는 혼합 (한지 함수의 부호에 근거)
- **gam_direction**: 증가, 감소 또는 혼합 (100 개 격자점에서 평활항의 도함수로 추정)
- **consistent**: TRUE/FALSE
- **warning**: 불일치가 있을 경우 그 내용

상태 표시줄에는 “All directions consistent” 또는 불일치가 있는 변수의 개수가 표시된다.

9.12 Concurvity(공곡선성)

공곡선성은 다중공선성의 평활 대응 개념이다. 이 탭은 다음을 보여 준다:

- 전체 공곡선성 표: 행은 worst/observed/estimate 측도, 열은 각 평활항이다. 색으로 구분된다: 흰색 (<0.5), 노랑 (0.5–0.8), 빨강 (>0.8).
- 쌍별 공곡선성 표: 동일한 색 구분을 적용한 최악의 경우 쌍별 행렬.

“worst” 행의 값이 약 0.8 을 넘으면 해당 평활항이 다른 평활항들로 잘 근사될 수 있다는 뜻이며, 중복 가능성을 시사한다.

9.13 Basis Check(기저 점검)

`mgcv::gam.check()` 의 출력으로, 각 평활항에 대한 기저 차원 적정성 검정이다. 어떤 평활항의 k 가 너무 작으면 이 검정이 낮은 p-값으로 이를 표시하며, k 를 늘릴 것을 시사한다.

10 데이터 내려받기

적합 후 **Download Output (Excel)** 버튼 (사이드바 7 절) 은 예측값과 진단 결과를 담은 Excel 파일을 활성 프로젝트의 mgcv 출력 폴더로 내보낸다.

10.1 출력 열

열	설명
basis	역변환된 절편값 (모든 행에서 동일)
<var>_contribution	평활항별 예측 기여도. 로그 모형에서는 금액 단위로 비례 배분된다.
est_<target>	모형 예측값 (예: est_sale_price). 해당하는 경우 로그 척도에서 역변환됨
residual	실측값 - 예측값
calc_residual	검증용: 실측값 - 기여도 합계 (역변환됨)
cqa	상태-품질-선호도 (Condition-Quality-Appeal, CQA) 점수 (0-10 척도)
residual_sf	잔차 / 거주면적 (living_area 가 지정된 경우)
cqa_sf	residual_sf 에 의한 순위로 계산한 CQA

감정평가 모드에서 1 행 (대상) 의 residual/cqa/cqa_sf 는 NA 로 설정된다. living_area 열이 지정된 경우 행은 residual_sf 내림차순으로 정렬된다.

성공적으로 완료되면 내려받기 버튼에 흰색 체크 표시가 나타난다.

10.2 CQA 점수

CQA 는 각 행의 잔차를 다른 모든 행과 견주어 0-10 척도로 순위를 매긴다:

- **높은 CQA(약 9-10):** 예측보다 훨씬 높은 가격에 매매됨 — 상태/품질/선호도가 우수할 가능성이 크다
- **낮은 CQA(약 0-1):** 예측보다 훨씬 낮은 가격에 매매됨 — 상태가 열악하거나 급매일 가능성이 크다
- **CQA 약 5:** 중앙값 부근

11 RCA 계산과 내려받기

RCA(잔차 제약 접근법, Residual Constraint Approach) 작업 흐름은 **Appraisal** 과 **Market Area Analysis** 모드에서 모형을 적합한 뒤 사용할 수 있다. 대상 부동산은 반드시 1 행에 있어야 한다.

11.1 RCA 대화상자 열기

사이드바 8 절의 **Calculate RCA Adjustments & Download** 버튼을 클릭한다. 다음 항목이 있는 모달 대화상자가 나타난다:

- **Score type:** CQA 또는 CQA per SF(living_area 가 지정된 경우). 단위면적당 옵션은 대상의 잔차를 거주면적으로 조정한다.
- **Subject CQA Score:** 0.00 에서 10.00 사이의 값을 받는다. 처음에는 빈칸이며, earthUI 모형을 가져온 경우 earthUI 에서 사용한 점수 유형과 값이 공유 프로젝트 설정을 통해 그대로 이어진다.

Generate 를 클릭하면 계산과 내려받기가 이루어진다. Excel 파일은 활성 프로젝트의 mgcv 출력 폴더 (<os>_out_mgcv) 에 기록된다.

11.2 CQA 점수 보간

1. 비교사례들의 CQA 점수와 잔차를 정렬한다
2. 선형 보간 (approx()) 으로 입력한 CQA 값을 잔차에 대응시킨다
3. 대상값 = 모형 예측값 + 보간된 잔차
4. 가중치가 0 인 행은 대상과 동일한 잔차를 가진 것처럼 취급된다

11.3 출력 열

열	설명
subject_value	모형 예측값 + 보간된 잔차
subject_cqa	사용자가 입력한 CQA 점수
<var>_adjustment	대상 기여도 - 비교사례 기여도 (금액 단위)
residual_adjustment	대상 잔차 - 비교사례 잔차
net_adjustments	모든 조정치의 합
gross_adjustments	조정치 절댓값의 합
residual_adj_pct	매매가격 대비 잔차 조정%
net_adj_pct	매매가격 대비 순조정%
gross_adj_pct	매매가격 대비 총조정%
adjusted_sale_price	실제 매매가격 + 순조정치

계산이 성공적으로 끝나면 버튼에 흰색 체크 표시가 나타난다.

12 거래사례비교 그리드

거래사례비교 그리드 (Sales Comparison Grid) 는 **Appraisal** 모드에서만, RCA 조정치를 계산한 뒤에 사용할 수 있다. 이 기능은 감정평가 보고서에 넣기 적합한 서식이 적용된 Excel 통합문서를 생성한다.

12.1 비교사례 선택 대화상자

Generate Sales Grid & Download 버튼을 클릭하면 모든 비교 매매사례를 나열하는 모달 대화상자가 열린다. 비교사례는 두 무리로 나뉜다:

- 추천 비교사례 (미리 선택됨): 총조정액이 매매가격의 $< 25\%$ 이며 매매 경과일 순 (최근 순) 으로 정렬된다. 최대 30 건.
- 기타 비교사례 (기본적으로 선택 해제): 나머지 사례로, 총조정률 순으로 정렬된다.

비교사례를 선택한 뒤 **Generate Sales Grid** 를 클릭하면 통합문서가 만들어진다.

필수 특수 유형: `contract_date` 와 `living_area` 가 반드시 지정되어야 한다. 권장 특수 유형: `latitude`, `longitude`, `lot_size` — 이들이 없어도 그리드는 작동하지만 일부 항목이 비게 된다.

12.2 통합문서 구성

생성된 통합문서에는 여러 시트가 있으며, 각 시트에는 비교사례 3 건이 나란히 배치된다:

- 머리글 행: 대상과 비교사례의 주소, 매매가격, 매매일, 시장 노출 일수
- 묶음 행: 위치 (위도, 경도, 지역), 대지 규모 (대지 면적, 대지 치수), 연수 (실제 연수, 유효 연수) 를 묶음 머리글과 하위 항목으로 표시
- 모형 변수 행: 평활항마다 한 행씩, 대상 기여도, 비교사례 기여도, 조정액을 보여 준다
- 잔차 특성 행: 모형이 포착하지 못한 특성 (예: 상태, 품질, 조망) 에 대해 감정평가사가 입력할 수 있는 빈 잠금 해제 칸
- 요약 행: 순조정%, 총조정%, 조정된 매매가격을 실제 Excel 수식으로 표시

12.3 시트 보호와 잔차 칸

각 시트는 실수로 편집되는 것을 막기 위해 보호되지만, 잔차 특성의 값 칸만은 명시적으로 잠금이 해제되어 있다. 이로써 감정평가사는 수식 기반의 조정 계산을 유지한 채 모형에 없는 특성의 값을 입력할 수 있다.

13 보고서 내려받기

보고서 생성은 이전 배포판의 단일 단계 보고서를 대체하는 2 단계 Quarto 작업 흐름이다.

13.1 1 단계: Quarto 보고서 생성

Generate Quarto Report 버튼 (모형을 적합한 뒤 사용 가능) 은 자립형 .qmd 묶음을 활성 프로젝트의 mgcv 출력 폴더에 기록한다. 이 묶음에는 Quarto 원본, 미리 렌더링된 그림 파일 (PNG 와 PDF), 모든 표와 모형 메타데이터를 담은 `report_data.rds` 페이로드, 그리고 Word 출력을 위한 `reference.docx` 서식 템플릿이 들어 있다. 묶음이 자립형이므로 렌더링 전에 .qmd 원본을 편집하거나 다른 Quarto 문서와 결합할 수 있다.

그림은 묶음을 생성할 때 미리 렌더링된다. 개별 그림 생성이 실패하면 자리표시 이미지가 기록되고 묶음의 나머지는 정상적으로 생성된다.

13.2 2 단계: Quarto 보고서 변환

Convert Quarto Report 절은 `quarto::quarto_render()` 를 통해 .qmd 파일을 하나 이상의 출력 형식으로 렌더링한다. 기본값은 가장 최근에 생성된 묶음이지만, **Browse** 버튼으로 임의의 .qmd 파일을 선택할 수도 있다. 세 가지 형식을 사용할 수 있다:

- **HTML** — 자립형 HTML 문서. 추가 소프트웨어 없이 모든 플랫폼에서 작동한다. 권장 기본값이다.
- **Word (.docx)** — 함께 제공되는 `reference.docx` 서식 템플릿으로 렌더링된다. 편집과 배포에 적합하다.
- **PDF** — **LaTeX** 설치가 필요하다 (아래 시스템 요건 참조). LaTeX 이 감지되지 않으면 형식 체크박스에서 PDF 옵션이 자동으로 숨겨진다.

이전 방식의 단일 단계 렌더러인 `render_gam_report()` 와 `export_gam_docx()` 는 프로그래밍 방식으로 계속 사용할 수 있다.

13.3 보고서 내용

생성되는 보고서에는 다음이 포함된다:

- **데이터셋과 모형:** 데이터 파일, 표본 크기, family, method, 적합 정보
- **결과 요약:** R^2 , CV R^2 , 설명된 이탈도, AIC 및 관련 적합 지표
- **모형 식:** 평활항은 f_i 표기로, 선형항은 계수를 명시한 모형 식과 평활 함수 정의. 긴 식은 PDF 에서 여러 줄로 줄바꿈된다.
- **요약 표:** 평활항 (EDF, F, p-값) 과 모수항 (추정값, 표준오차, t, p-값)
- **활성화된 상호작용:** 모형에 포함된 텐서 및 요인별 평활 상호작용 항의 행렬 (상호작용이 있을 때만 표시)
- **변수 중요도:** F-통계량 (평활항) 또는 $|t|$ (모수항) 순위의 막대그래프
- **평활 효과** (모든 항 유형):
 - 일변량 평활 그림 (대화형 그림의 정적 ggplot2 버전)
 - 요인별 평활 다중선 그림
 - 상호작용 히트맵 (t_i 및 t_e 항)
 - 모수항 그림 (수치형은 산점도, 요인형은 막대그래프)
- **예측변수 상관 히트맵**
- **진단:** 잔차 진단과 실측값 대 예측값 산점도
- **mgcv 요약:** 적합된 모형의 전체 텍스트 출력
- **표기법 부록:** 보고서에서 사용한 가치 기여 함수 표기법을 설명하는 번호 없는 마무리 절

생성이 성공하면 Generate Quarto Report 버튼에 흰색 체크 표시가 나타난다.

14 모형 함수 내보내기

mgcvUI 는 적합된 GAM 의 평활 함수를 여러 언어의 독립 코드로 내보낼 수 있다. 각 평활항은 200 개 격자점에서 평가되어 선형 보간을 사용하는 조회 표 (lookup table) 로 내보내지므로, mgcv 에 의존하지 않고 R 밖에서도 모형을 사용할 수 있다.

현재 배포판에서 함수 내보내는 것은 프로그래밍 방식으로 이루어진다. 내보내기 보조 함수들은 (fit_gam() 이 생성한) 적합 결과 객체를 받아 아래에 설명한 파일들을 기록한다:

- generate_r_code(), generate_python_code(), generate_cpp_code() — 언어별 소스 파일
- export_functions_sqlite() — SQLite 조회 표 데이터베이스
- export_functions_zip() — 네 가지 형식을 임의로 조합해 zip 압축 파일로 묶는다

14.1 R 함수

파일: gam_functions.R

- 반응변수, R^2 , $CV R^2$, 식을 담은 머리말 주석.
- INTERCEPT 상수.
- 평활 변수마다 미리 계산된 x/y 벡터를 갖는 approxfun() 하나씩.
- 선형항 계수를 이름 있는 상수로 제공.
- 사용법: 절편, 각 g_var(value), 각 coef_var * value 를 합산한다.

14.2 Python 함수

파일: gam_functions.py

- numpy 가져오기.
- INTERCEPT 상수.
- 변수별: _x_var 및 _fx_var 배열과 np.interp() 를 사용하는 g_var(x) 함수.
- 선형항 계수를 COEF_* 상수로 제공.

14.3 C++ 헤더

파일: gam_functions.hpp

- #pragma once 를 사용하는 헤더 전용 파일.
- gam_functions 네임스페이스.
- std::lower_bound 를 사용하는 interp() 보조 함수.
- 변수별: 정적 상수 배열을 갖는 인라인 함수.
- 선형항 계수를 constexpr double 로 제공.

14.4 SQLite 데이터베이스

파일: gam_functions.sqlite

- smooth_grids 테이블: (variable, x, fx) 열 — 200 점 조회 표.
- model_info 테이블: response, r_squared, cv_rsqr, dev_explained, aic, n_obs, family, method, formula.
- linear_terms 테이블: variable, coefficient.
- intercept 테이블: 절편값.

15 earthUI 및 glmnetUI 와의 비교

earth 응용 프로그램과 그 후처리 모듈들은 회귀 모델링을 위한 하나의 도구 모음이다. 이들은 동일한 데이터 형식, 특수 열 유형, RCA 작업 흐름, 시연용 데이터셋을 공유하지만 모델링 엔진은 서로 다르다.

15.1 주요 차이

특성	mgcvUI	earthUI	glmnetUI
방법	GAM (mgcv)	MARS/Earth (earth)	엘라스틱 넷 (glmnet)
관계	평활 비선형	구간별 선형	선형
변수 선택	<code>select=TRUE</code> 수축	단계적 가지치기	라쏘 별점
상호작용	텐서곱 (te/ti)	힌지 곱 (2-3 차)	쌍별 곱
부분효과	신뢰구간을 갖는 평활 곡선	g-함수 (구간별)	선형 기울기
earth 파이프라인	earth 에서 매듭점을 가져옴	매듭점의 출처	힌지 기저를 가져옴
코드 내보내기	R, Python, C++, SQLite	R, Python, C++, SQLite	—
보고서 형식	Word, PDF, HTML	Word, PDF, HTML	Word, PDF, HTML

earthUI 에만 있는 차별점 하나를 언급할 만하다. earthUI 에는 자유 서술 형태의 MLS 비교란에서 구조화된 부동산 속성을 추출하는 실험적 Prolog/DCG 기능이 포함되어 있다.

15.2 공통 기능

세 도구 모두 다음을 제공한다:

- 공유 `regProj` 프로젝트 트리 (기본값 `~/regProj`) — 프로젝트, 입력 파일, 방법별 출력 형제 앱들 사이에서 공유된다
- `snake_case` 열 변환을 동반한 CSV 및 Excel 가져오기
- CSV 구분자, 소수점 기호, 용지 크기를 포함한 30 개국 이상의 로케일 지원
- 특수 열 지정 (`contract_date`, `living_area`, `area`, `latitude`, `longitude` 등)
- 기준일로부터의 `sale_age` 자동 계산
- 다크/라이트 모드 전환 (Nord 테마)
- 설정 유지 (SQLite 및/또는 `localStorage`)
- 상관 히트맵, 변수 중요도, 진단
- CQA 점수 산정과 RCA 조정 작업 흐름
- 비교사례 선택과 서식이 적용된 Excel 출력을 갖춘 거래사례비교 그리드
- 동일한 시연용 데이터셋 (`Appraisal_1.csv`) 호환

15.3 권장 작업 흐름

- 중요한 변수, 비선형 관계, 상호작용을 발견하려면 **earthUI** 로 시작하라. earth 의 자동 모형 구축이 가장 좋은 출발점을 제공한다.
- 부분효과를 더 매끄럽게 만들고 싶다면 **mgcvUI** 로 정교화하라. earthUI 의 매듭점을 가져오면 매끄럽게 이행할 수 있다.
- 공격적인 변수 선택이나 계수 부호 제약이 필요한 순수 선형 모형이 필요하다면 **glmnetUI** 를 사용하라.
- 비선형 모형의 추가적인 복잡도가 정당한지 판단하려면 동일한 데이터에 세 도구를 모두 적용해 비교하라.

16 시연용 데이터셋: Appraisal_1.csv

16.1 설명

시연용 데이터셋은 earthUI, glmnetUI 와 공유된다. 이 파일은 mgcvUI 소스의 demo_mls/ 폴더에 있으며, earthUI 가 설치되어 있다면 다음과 같이 찾을 수 있다:

```
demo_file <- system.file("extdata", "Appraisal_1.csv", package = "earthUI")
```

이 파일에는 모의 MLS 내보내기에서 만든 약 1,500 건의 주거용 매매 기록이 들어 있으며, 1 행이 대상 부동산이고 나머지 행이 비교 매매사례이다. 데이터는 부동산의 규모, 연수, 위치가 다양하게 분포한 다중 지역 시장의 단독주택 매매를 나타낸다.

이는 실제 데이터가 아니며, 북부 캘리포니아의 실제와 유사한 근린을 바탕으로 만든 것이다. 모든 식별 정보는 변경되거나 삭제되었다.

16.2 열

열	유형	특수 유형	설명
sale_type	character	sale_type	대상과 비교 매매사례를 구분하는 레코드 유형 코드
weight	numeric	weight	관측치 가중치 (0 = 적합에서 제외)
id	numeric	display_only	내부 레코드 ID
property_id	numeric	display_only	MLS 부동산 식별자
listing_id	character	display_only	MLS 매물 번호
parcel_number	character	display_only	카운티 과세평가관의 필지 번호 (APN)
street_address	character	display_only	부동산 주소
city_name	character	display_only	도시
postal_code	character	display_only	우편번호
county_name	character	display_only	카운티
contract_date	Date	contract_date	매매 계약일 (sale_age 를 계산함)
sale_age	numeric	—	계약일부터 기준일까지의 일수 (미리 계산됨)
coe_date	Date	display_only	에스크로 종결일
listing_status	character	display_only	매물 상태 (예: "Sold")
sale_price	numeric	(목표변수)	매매가격 — 반응변수
rent	numeric	—	월 임대료 (다중 목표변수 모형용)
list_price	numeric	display_only	호가
original_list_price	numeric	display_only	최초 호가
living_sqft	numeric	living_area	총 거주면적 (제곱피트)
ppsf	numeric	display_only	제곱피트당 가격
beds_total	integer	—	침실 수
baths_total	numeric	—	총 욕실 수 (예: 2.5 = 전체 2 + 반 1)
lot_size	numeric	lot_size	대지 면적 (제곱피트)
area_id	integer	area	MLS 지역 식별자
area_text	character	display_only	지역명
age	numeric	actual_age	부동산의 연수 (년)
year_built	integer	display_only	건축 연도
latitude	numeric	latitude	위도 (모형에서는 소수점 3 자리로 반올림)
longitude	numeric	longitude	경도 (모형에서는 소수점 3 자리로 반올림)
latitude6	numeric	display_only	전체 정밀도 위도
longitude6	numeric	display_only	전체 정밀도 경도
garage_spaces	integer	—	차고 주차 대수
fp_count	integer	—	벽난로 개수
no_of_stories	numeric	—	층수
style	character	—	건축 양식
view	character	—	조망 유형 (예: "Neighborhood", "Hills")
days_on_market	integer	dom	시장 노출 일수
listing_date	Date	listing_date	매물 등록일

열	유형	특수 유형	설명
sale_concessions	numeric	concessions	매도인 양여 조건

16.3 빠른 시작 예시

1. mgcv 루틴을 실행한다: `earthUI::launch_mgcv()`
2. 목적을 **For Appraisal** 로 하여 새 프로젝트를 만든다 (1 절)
3. `Appraisal_1.csv` 를 프로젝트의 입력 폴더에 넣고 2 절에서 선택한다 (필요하면 **Refresh** 를 클릭한다)
4. `sale_price` 를 목표변수로 선택한다
5. 비율 조정을 원한다면 Response Transform 을 **Log (natural)** 로 설정한다 (선택 사항)
6. 위의 표에 나온 대로 특수 유형을 지정한다
7. 예측변수를 포함시킨다: `sale_age, living_sqft, baths_total, lot_size, area_id`(요인으로), `age, latitude, longitude, garage_spaces`
8. 기본값을 유지한다: `REML, gamma = 1.2, tp 기저, k = 자동, select = TRUE`
9. **Fit Mgcvm GAM Model** 을 클릭한다
10. Summary, Contribution(평활 그림), Diagnostics 탭을 검토한다
11. 출력을 내려받고 (7 단계) CQA 순위를 살펴본다
12. CQA 점수 ~5.00 으로 RCA 조정치를 계산한다 (8 단계)
13. 추천 비교사례로 거래사례비교 그리드를 생성한다 (9 단계)

17 시스템 조건과 문제 해결

17.1 지원 플랫폼

mgcvUI 는 **macOS, Windows, Linux** 에서 실행된다. 이 응용 프로그램은 RStudio 를 사용하는 macOS 에서 개발되고 주로 시험되었다. 플랫폼별 유의 사항은 아래에 정리했다.

17.2 R 버전

R $\geq 4.1.0$ 이 필요하다. RStudio Desktop(2023.06 이상) 을 적극 권장한다. RStudio 에는 보고서 렌더링에 필요한 **Quarto** 와 **pandoc** 이 함께 들어 있으며 앱을 실행하기에 편리한 환경을 제공한다.

17.3 필수 R 패키지

다음 패키지들은 의존성으로 자동 설치된다:

패키지	용도
shiny	웹 응용 프로그램 프레임워크
mgcv	GAM 모형 적합
valengrCore	공유 감정평가 코어 (특수 역할, earthUI 설정 이어받기)
ggplot2, plotly	정적 및 대화형 기여도 그림
gratia	평활항 추출과 진단
DT	대화형 데이터 표
quarto, rmarkdown	Quarto 보고서 생성과 렌더링
readxl	Excel 파일 가져오기
shinyFiles	파일 및 폴더 탐색기
RSQLite, DBI	설정 및 프로젝트 데이터베이스 (SQLite)
jsonlite	변수 상태 직렬화

17.4 선택적 의존성

구성 요소	필요한 경우
LaTeX (TinyTeX, MiKTeX 또는 MacTeX)	PDF 보고서 생성 시에만 필요하다. LaTeX 설치가 감지되지 않으면 형식 드롭다운에서 PDF 옵션이 자동으로 숨겨진다. 설치 방법: <code>tinytex::install_tinytex()</code>
sysfonts + showtext	ggplot2 그림에 쓰이는 Roboto Condensed 글꼴용. 사용할 수 없거나 오프라인이면 앱이 자동으로 시스템 산세리프 글꼴로 대체한다.
earth ($\geq 5.3.0$)	earthUI 가져오기 파이프라인에만 필요하다. 단독 사용에는 필요하지 않다.
callr	백그라운드 모형 적합 (앱의 반응성을 유지한다). 없으면 적합이 동기적으로 실행된다.
writexl, openxlsx	Excel 내려받기 (모형 출력, RCA, 매매사례 그리드). 없으면 앱이 안내한다.
bslib	Bootstrap 5 테마 적용 (Nord Light/Dark).
officer	이전 방식의 프로그래밍 Word 내보내기 (<code>export_gam_docx()</code>).

17.5 플랫폼별 유의 사항

17.5.1 macOS

문제가 가장 적다. 시스템 라이브러리에는 Homebrew 를 권장한다. PDF 보고서가 필요하다면:

```
# In R:
install.packages("tinytex")
tinytex::install_tinytex()
```

17.5.2 Windows

RStudio Desktop 과 잘 작동한다. 주요 유의 사항은 다음과 같다:

- 기업용/보안이 강화된 컴퓨터: 임시 디렉터리나 사용자 데이터 디렉터리가 제한되어 있으면 설정 유지와 보고서 생성이 실패할 수 있다. 앱은 R 콘솔에 경고를 표시하지만 설정 유지 없이 계속 작동한다.
- **LaTeX**: R 에서 TinyTeX 를 설치하거나 (위 macOS 와 동일한 명령) miktex.org 에서 MiKTeX 를 설치한다.
- 긴 파일 경로: Windows 에는 260 자 경로 제한이 있다. 작업 디렉터리 경로를 짧게 유지하라.

17.5.3 Linux

배포판에 따라 편차가 가장 크다. Ubuntu/Debian 사용자는 R 패키지를 컴파일 하기 전에 시스템 라이브러리가 필요할 수 있다:

```
sudo apt install libcurl4-openssl-dev libssl-dev \
  libxml2-dev libsqlite3-dev libfontconfig1-dev
```

PDF 보고서를 위해서는 `tinytex::install_tinytex()` 를 실행하거나 시스템 패키지 관리자로 `texlive-full` 을 설치한다.

디스플레이가 없는 헤드리스 서버에서는 plotly 대화형 그림이 보고서에서 정적 대체 이미지로 렌더링된다. Shiny 앱 자체는 웹 브라우저 연결이 필요하다.

17.6 점진적 기능 축소

mgcvUI 는 선택적 구성 요소가 없어도 작동하도록 설계되었다:

누락된 구성 요소	동작
LaTeX 없음	보고서 형식 드롭다운에서 PDF 옵션이 숨겨진다. HTML 과 Word 는 계속 사용할 수 있다.
인터넷 없음 / 글꼴 실패	Roboto Condensed 글꼴이 시스템 산세리프로 대체된다. 콘솔에 메시지가 기록된다.
RSQLite 없음 / 읽기 전용 파일 시스템	세션 사이에 설정이 유지되지 않는다. 콘솔에 경고가 표시된다. 그 밖에는 앱이 정상 작동한다.
gratia::appraise 실패	Diagnostics 탭과 보고서에 자리표시 메시지가 표시된다. 대안으로 콘솔에서 <code>mgcv::gam.check()</code> 를 사용하라.
임시 디렉터리에 쓸 수 없음	보고서 생성이 명확한 오류 메시지와 함께 실패한다. TMPDIR 환경변수를 쓰기 가능한 위치로 지정하여 해결한다.

17.7 문제 해결

17.7.1 “PDF option is not available”

LaTeX 설치가 감지되지 않은 경우이다. R 에서 다음을 실행한다:

```
install.packages("tinytex")
tinytex::install_tinytex()
```

그런 다음 앱을 다시 시작하면 PDF 옵션이 나타난다.

17.7.2 “Settings will not persist”

SQLite 데이터베이스 디렉터리를 만들거나 여기에 쓸 수 없는 경우이다. 다음 경로의 권한을 확인하라:

- macOS/Linux: ~/Library/Application Support/R/mgcvUI/ 또는 ~/.local/share/R/mgcvUI/
- Windows: %APPDATA%/R/data/mgcvUI/

17.7.3 “Diagnostics plot unavailable”

gratia::appraise() 함수는 복잡한 모형 (텐서 상호작용이 많거나 비표준 family 인 경우) 에서 이따금 실패한다. 이는 상위 패키지의 알려진 문제이다. 우회 방법: R 콘솔에서 mgcv::gam.check(model) 을 실행해 진단 그림을 직접 확인하라.

17.7.4 “Report error: number of items to replace...”

이는 gratia 패키지와 관련하여 이전 버전에서 알려졌던 문제이다. 최신 버전의 mgcvUI 가 설치되어 있는지 확인하라. 현재 버전은 이 오류를 포착하여 진단 절에 자리표시를 넣은 보고서를 생성한다.

17.7.5 포트 7880 이 이미 사용 중

이전 세션 때문에 포트 7880 이 점유되어 앱이 시작되지 않는다면 다음을 실행한다:

```
# macOS/Linux:
lsof -ti:7880 | xargs kill
```

```
# Windows (PowerShell):
Stop-Process -Id (Get-NetTCPConnection -LocalPort 7880).OwningProcess
```

17.7.6 Linux 에서 패키지 설치 실패

시스템 라이브러리가 없는 경우이다. 위 Linux 절에 나열된 개발용 헤더를 설치한 뒤 install.packages() 를 다시 시도하라.

참고문헌

- Wood, S. N. (2003). “Thin-plate regression splines”. In: *Journal of the Royal Statistical Society (B)* 65.1, pp. 95–114. DOI: [10.1111/1467-9868.00374](https://doi.org/10.1111/1467-9868.00374).
- (2004). “Stable and efficient multiple smoothing parameter estimation for generalized additive models”. In: *Journal of the American Statistical Association* 99.467, pp. 673–686. DOI: [10.1198/016214504000000980](https://doi.org/10.1198/016214504000000980).
- (2011). “Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models”. In: *Journal of the Royal Statistical Society (B)* 73.1, pp. 3–36. DOI: [10.1111/j.1467-9868.2010.00749.x](https://doi.org/10.1111/j.1467-9868.2010.00749.x).

- Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC.
- Wood, S. N., N. Pya, and B. Säfken (2016). “Smoothing parameter and model selection for general smooth models (with discussion)”. In: *Journal of the American Statistical Association* 111, pp. 1548–1575. DOI: [10.1080/01621459.2016.1180986](https://doi.org/10.1080/01621459.2016.1180986).