glmnetUI: интерактивный интерфейс для пакета glmnet (лассо и эластичная сеть)

Русский перевод, подготовленный с использованием искусственного интеллекта и прошедший редакционную проверку; английская версия является основной.

William Bert Craytor

Том 1, Выпуск 1 · Лето 2026
Опубликовано: 1 июля 2026
DOI: 10.67330/ja8s2k39
© 2026 William Bert Craytor. Published by Pacific Vista Net.
Эта статья открытого доступа распространяется на условиях лицензии Creative Commons Attribution 4.0

Аннотация. glmnetUI — графический интерфейс пользователя для пакета R glmnet, который подгоняет регуляризованные обобщённые линейные модели с использованием штрафов лассо, гребневой регрессии и эластичной сети. Он предлагает три режима работы — общее прогнозное моделирование, оценку недвижимости и анализ рыночной зоны — и проводит пользователя через импорт данных, настройку модели, подгонку с перекрёстной проверкой, траектории коэффициентов и загружаемые отчёты. В статье описаны требования к формату данных, рабочий процесс моделирования, форматы вывода и полный справочник возможностей glmnetUI.

Ключевые слова: лассо, гребневая регрессия, эластичная сеть, регуляризованная регрессия, glmnet, R, графический интерфейс, оценка недвижимости

Yayıncı notu (Temmuz 2026): Baskıdan kısa süre önce glmnetUI, earthUI (sürüm 0.11.0 ve sonrası) içinde birleştirildi; earthUI burada betimlenen esnek ağ iş akışını artık MARS ve GAM iş akışlarının yanında sunmaktadır. glmnetUI, son yayımlanmış hâliyle (0.5.0) https://github.com/wcraytor/glmnetUI adresinde erişilebilir olmayı sürdürmektedir; geliştirme earthUI'de devam etmektedir. Bu makale, yayımlandığı hâliyle glmnetUI için geçerlidir.

Содержание

1 Введение 3
1.1 История и предыстория . . . 3
1.1.1 The Lasso . . . 3
1.1.2 Гребневая регрессия и эластичная сеть . . . 4
1.1.3 Пакет glmnet . . . 4
1.1.4 Сравнение с MARS (Земля) и GAM (mgcv) . . . 4
1.1.5 Рекомендуемый рабочий процесс . . . 6
1.2 Что такое glmnetUI? . . . 6
1.3 Три режима работы . . . 6
1.4 Недвижимость — специальные материалы . . . 7
1.5 Начало работы . . . 7

2 Требования к данным для ввода MLS 9
2.1 Формат файла & Структура . . . 9
2.2 Обязательные столбцы для оценки Режим . . . 9
2.3 Особые правила именования столбец . . . 10
2.4 Качество данных& Полнота данных . . . 10
2.5 Размещение объекта оценки в строке . . . 11

3 Режим общего назначения 12
3.1 Обзор . . . 12
3.1.1 Пропустить первую строку . . . 12
3.1.2 Настройки в зависимости от цели . . . 12
3.2 Рабочий процесс с боковой панелью . . . 12
3.3 Проекты . . . 13
3.4 Импорт из earthUI (Необязательно) . . . 14
3.5 Вкладки главной панели . . . 14
3.6 Сохранение настроек . . . 15
3.7 Темный режим . . . 16
3.8 Язык и регион & Региональные настройки . . . 16
3.8.1 Сохранение настроек по умолчанию . . . 16

4 Режим оценки 17
4.1 Обработка строк объекта оценки . . . 17
4.2 Дата оценки & Возраст, с которого разрешена продажа . . . 17
4.3 Специальный столбец Обозначения . . . 17
4.4 Обзор корректировок RCA . . . 17

5 Режим анализа рыночной зоны 18
5.1 Отличия от режима оценки . . . 18
5.2 Когда следует использовать режим «Рынок» . . . 18

6 Отбор переменных 19
6.1 Целевая переменная . . . 19
6.2 Таблица предикторов . . . 19
6.3 Определение типов данных& Переопределения . . . 19
6.4 Ожидаемые признаки& Обеспечение соблюдения . . . 19
6.5 Специальные типы столбцов Ссылка . . . 20

7 Выбор параметров 22
7.1 Alpha (параметр микширования) . . . 22
7.2 Lambda Selection . . . 22
7.3 Семья . . . 23
7.4 Стандартизировать . . . 23
7.5 Ограничения знаков . . . 23
7.6 «Расслабленное лассо» . . . 23
7.7 Начальные значения для генератора случайных чисел . . . 23
7.8 Матрица взаимодействий . . . 24
7.9 Блок от основного эффекта (только взаимодействие) . . . 24
7.10 Расширенные параметры . . . 25
7.11 Настройки по умолчанию . . . 25

8 Подгонка модели 26
8.1 Кнопка «Fit» . . . 26
8.2 Что происходит во время подгонки . . . 26

9 Вкладки «Результаты» 27
9.1 Предварительный просмотр данных . . . 27
9.2 Уравнение . . . 27
9.3 Корреляция . . . 27
9.4 Резюме . . . 27

9.5 Коэффициенты . . . 28
9.6 Важность переменных . . . 28
9.7 Материалы . . . 28
9.8 ANOVA . . . 28
9.9 Диагностика . . . 28
9.10 Вывод Glmnet . . . 29
9.11 Отчёт « . . . 29

10 Загрузка данных 30
10.1 Столбецы вывода . . . 30
10.2 Результаты CQA: балл . . . 30

11 Расчеты RCA& Скачивание 31
11.1 Открытие диалогового окна RCA . . . 31
11.2 Интерполяция баллов CQA . . . 31
11.3 Столбецы вывода . . . 31

12 Таблица сравнения продаж 32
12.1 Диалоговое окно выбора аналогов . . . 32
12.2 Макет книги . . . 32
12.3 Защита листа & Остатки ячейки . . . 32

13 Скачивание отчетов 33
13.1 «Двухэтапный кварто» Рабочий процесс . . . 33
13.2 Форматы отчётов . . . 33
13.3 Содержание отчёта . . . 33

14 Сравнение с earthUI 35
14.1 Основные различия . . . 35
14.2 Общие функции . . . 35
14.3 Когда использовать какой . . . 36

15 Демо-набор данных: Оценка _ 1.csv 37
15.1 Описание . . . 37
15.2 Столбецы . . . 38
15.3 Рекомендуемое руководство по быстрому запуску . . . 38

16 Системные требования& Устранение неполадок 40
16.1 Поддерживаемые платформы . . . 40
16.2 Дополнительные зависимости . . . 40
16.3 Примечания по платформе . . . 40
16.4 Плавная деградация . . . 40
16.5 Устранение неполадок в программе « . . . 40

1 Введение

1.1 История и предыстория

1.1.1 The Lasso

Метод Lasso (Least Absolute Сжатие и Отбор) был предложен Робертом Тибширани в 1996 году. Название представляет собой намеренную игру слов, связанную с ковбойским лассо

— верёвкой, которая ограничивает и отбирает. Тибширани выбрал его потому, что метод буквально «закидывает лассо» на коэффициенты, ограничивая их абсолютные значения и приводя некоторые из них к точно нулю, тем самым отбирая те переменные, которые останутся в модели.

Идея Тибширани заключалась в том, чтобы добавить штрафное условие «L_1» (сумма абсолютных значений коэффициентов) к целевой функции метода наименьших квадратов. В отличие от штрафа L_2, используемого в гребневой регрессии (Hoerl& Kennard, 1970), который сжимает коэффициенты в сторону нуля, но никогда не достигает его, штраф L_1 дает разреженные решения, в которых многие коэффициенты равны точно нулю. Это делает Lasso одновременно методом регуляризации и методом отбора переменных.

1.1.2 Гребневая регрессия и эластичная сеть

Гребневая регрессия была предложена Артуром Хёрлом и Робертом Кеннардом в 1970 году. В ней добавляется штраф по методу «L_2» (сумма квадратов коэффициентов), который пропорционально уменьшает все коэффициенты, но сохраняет все переменные в модели. Ридж-регрессия эффективна, когда предикторы коррелируют между собой (мультиколлинеарность), но не упрощает модель.

Метод «эластичной сети» (эластичная сеть), предложенный Хуи Цзоу и Тревором Хасти в 2005 году, сочетает в себе оба вида штрафов. Параметр смешивания α регулирует соотношение: $\alpha = 1$ — это чистый Lasso, $\alpha = 0$ — чистый Ridge, а промежуточные значения дают компромиссный вариант. Эластичная сеть преодолевает ограничение метода Lasso: когда предикторы сильно коррелируют, Lasso, как правило, выбирает один и игнорирует остальные, тогда как эластичная сеть объединяет коррелирующие предикторы в группы.

1.1.3 Пакет glmnet

Пакет R « glmnet» был разработан Джеромом Фридманом, Тревором Хэсти и Робертом Тибширани в Стэнфордском университете. Он реализует регуляризацию по методу «эластичная сеть» для обобщенных линейных моделей с использованием чрезвычайно эффективного алгоритма координатного спуска (Friedman, Hastie и Tibshirani 2010), который впоследствии был расширен на модели пропорциональных рисков Кокса (Simon и др. 2011) и на все семейства обобщенных линейных моделей (Tay, Narasimhan и Hastie 2023). Пакет поддерживает гауссовы (линейные), биномиальные (логистические), пуассоновские и другие семейства обобщённых линейных моделей.

К ключевым особенностям « glmnet» относятся:

- Метод регуляризации : подгонка модели по таблице значений λ за один проход, что значительно быстрее, чем отдельная подгонка для каждого значения λ.
- Перекрёстная проверка : cv.glmnet() автоматически выбирает оптимальную степень регуляризации с помощьюk -блока перекрёстной проверки.
- Границы коэффициентов : поддерживает верхние и нижние границы для отдельных коэффициентов (ограничения по знаку).
- Ослабленное Lasso : повторно подгоняет модель по выбранным переменным с уменьшенным штрафом или без него, что снижает смещение.

1.1.4 Сравнение с MARS (Земля) и GAM (mgcv)

Приложение «Earth» и его два модуля постобработки (glmnet и mgcv) используют разные механизмы моделирования. В приведённой ниже таблице приведены основные различия:

Особенности	glmnet (эластичная сеть)	earth (MARS)	mgcv (GAM)
Тип модели	Линейная (с регуляризацией)	Кусочно-линейная (адаптивные сплайны)	Сглаживание нелинейности (сплайны с штрафной пенализацией)
Нелинейность	Только через взаимодействия или разложение в базис	Автоматически через шарнирные функции с узлами, определяемыми данными	Автоматически через функции сглаживания (тонкая пластина, кубическая и т. д.)
Отбор переменных	Встроенный с помощью штрафаL_1 (Lasso)	Встроенный с помощью пошагового метода «вперед/назад» с обрезкой по GCV	Опционально с помощью сжатия с двойным штрафом (select=TRUE)
Взаимодействия	Попарные векторные произведения (сложно интерпретировать)	Произведения на основе шарниров (интерпретируемые, до степени 3)	Сглаженные тензорные произведения (te() , ti())
Интерпретируемость	Коэффициенты представляют собой линейные веса; их легко объяснить по отдельности	g-функции показывают кусочно-линейные частичные эффекты по каждой переменной	Сглаженные кривые частичных эффектов; очень интуитивно понятны
Лучше всего подходит для	Высокоразмерных данных, отбор переменных, небольшие выборки	Автоматическое обнаружение нелинейности и взаимодействий	Сглаживание нелинейных зависимостей, гибкое моделирование
Корректировки RCA	Линейные корректировки по каждой переменной	Кусочно-линейные корректировки с помощью g-функций	Сглаживание корректировок с помощью частичных эффектов
Контроль переобучения	Лямбда, проверенная с помощью перекрёстной проверки	Обрезка на основе GCV	Оценка параметров сглаживания REML/GCV

Метод MARS (Multivariate Adaptive Regression Splines) был предложен Джеромом Фридманом в 1991 году. Он позволяет строить кусочно-линейные модели путем адаптивного выбора функций-шарниров$\max(0, x-k)$ и положений их узлов на основе данных. Реализация в R представлена в пакете earth , разработанном Стивеном Милбороу. MARS отлично справляется с автоматическим обнаружением нелинейных зависимостей и взаимодействий, не требуя от пользователя их указания. Его g-функции (сгруппированные базисные слагаемые по переменным) обеспечивают наглядные кривые частичных эффектов.

Обобщенные аддитивные модели (GAM) были формализованы Хасти и Тибширани (1990). Реализация на языке R mgcv, разработанная Саймоном Вудом, предоставляет сплайны с штрафной регрессией и автоматическим выбором степени сглаживания по методам REML или GCV. GAM генерируют плавные кривые частичных эффектов, которые интуитивно понятны для визуализации и объяснения. Когда earthUI экспортирует положения

узлов в mgcvUI, узлы, полученные с помощью earth, служат отправными точками для сглаживающих слагаемых GAM, сочетая адаптивное размещение узлов MARS с сглаживанием GAM.

1.1.5 Рекомендуемый рабочий процесс

Для оценки недвижимости и аналогичных задач, требующих понятных и обоснованных моделей:

1. Начните с earthUI , чтобы выявить важные переменные, нелинейные зависимости и взаимодействия в ваших данных. Автоматический выбор базиса и g-функции earthUI позволяют получить отличные первоначальные выводы.
2. Уточните результаты с помощью mgcvUI , если вам нужны более плавные частичные эффекты. Импортируйте узлы из earthUI в mgcvUI для плавного перехода от кусочно-линейных моделей к моделям сглаживания.
3. Используйте glmnetUI , если вам требуется агрессивный отбор переменных (Lasso), когда количество предикторов приближается к числу наблюдений или превышает его, либо если вам нужна чисто линейная модель с регуляризацией для обеспечения обоснованности результатов.

1.2 Что такое glmnetUI?

glmnetUI — это графический интерфейс пользователя для пакета R « glmnet ». Он работает как локальное приложение Shiny — без входа в систему, без сервера и без учетных записей. Вы запускаете его из R, импортируете набор данных (CSV или Excel), настраиваете модель и выполняете ее подгонку в интерактивном режиме.

Начиная с текущей версии, glmnetUI больше не является отдельно устанавливаемым пакетом: он поставляется в составе пакета earthUI в качестве модуля постобработки для эластичной сети, запускаемого с помощью команды earthUI::launch_glmnet(). Само приложение — а также всё, что описано в данной статье, — осталось без изменений; отличаются лишь процедура установки и команда запуска. Эти три процедуры (earth, glmnet, mgcv) используют одно дерево проекта, одну базу данных настроек и одну копию вспомогательной инфраструктуры.

Приложение обеспечивает полный рабочий процесс: импорт данных, настройку переменных, подгонку модели, графики диагностики, оценку значимости переменных, уравнения модели, а также возможность скачивания отчетов в форматах Word, PDF или HTML.

Регрессия с эластичной сетью сочетает в себе два метода регуляризации:

- Гребневая регрессия (альфа = 0): пропорционально сжимает все коэффициенты к нулю. Подходит, если вы считаете, что все переменные являются значимыми.
- Регрессия Lasso (альфа = 1): может сжимать коэффициенты точно до нуля, выполняя автоматический отбор переменных. Позволяет получать более простые и более интерпретируемые модели.
- Эластичная сеть (0< альфа< 1): сочетание обоих подходов. Полезно, когда предикторы коррелируют между собой.

1.3 Три режима работы

Каждый проект glmnetUI имеет назначение, выбранное при его создании (раздел 1, «Проект»). Назначение активного проекта определяет, какие инструменты и элементы интерфейса отображаются; чтобы работать в другом режиме, создайте или откройте проект с соответствующим назначением. Существуют три режима:

- Общие настройки — Регрессия эластичной сети для любого типа совокупности или набора данных. Это режим по умолчанию. Он обеспечивает полный рабочий процесс регуляризованной регрессии без каких-либо дополнений, специфичных для конкретной области.
- Для оценки — Регрессия эластичной сети, адаптированная для оценки недвижимости. Добавляет функции, специфичные для оценки отдельных объектов недвижимости, включая обработку объекта оценки, специальные обозначения столбцов и Метод остаточных ограничений (RCA).
- Анализ рыночной зоны — Регрессия эластичная сеть, адаптированная для исследований рыночных зон. Добавляет функции для анализа групп объектов недвижимости на определённом рынке с возможностью исключения объекта оценки из ряда.

Во всех трёх режимах основной механизм моделирования одинаков — вы всегда проводите подгонку модели «эластичная сеть» (glmnet). Настройка «Цель» определяет, какие дополнительные инструменты и элементы интерфейса будут доступны.

1.4 Недвижимость — специальные материалы

При выборе опции «Для оценки» или «Анализ рыночной зоны» glmnetUI активирует ряд функций, предназначенных для анализа рынка недвижимости:

- Обозначения специальных столбцов — Каждому предиктору можно присвоить специальную роль, такую как « contract_date », « dom », « concessions », « latitude », « longitude », « living_area », « lot_size », « actual_age », « effective_age », « area », « site_dimensions » или « display_only ». Эти обозначения определяют, как столбец будет обрабатываться во время подгонки модели и при выводе результатов.
- Округление широты и долготы — Столбцы, обозначенные как широта или долгота, автоматически округляются до 3 знаков после запятой, чтобы предотвратить переобучение.
- Столбец «Срок продажи» — Когда столбец обозначен как contract_date и указана дата оценки, glmnetUI вычисляет столбец « sale_age » (количество дней между датой продажи и датой оценки) и использует его в качестве предиктора.
- Расчеты RCA (только для оценки) — В режиме оценки, после подгонки модели glmnetUI может вычислить результаты метода Residual Constraint (RCA), который генерирует корректировки для каждого аналога, сводные данные по чистым и валовым корректировкам, а также скорректированную цену продажи объекта оценки.

Совет: Режим «Общее назначение» подходит для любых наборов данных — финансовых, научных, инженерных или связанных с недвижимостью. Режимы «Оценка» и «Рынок» просто предоставляют дополнительные удобные функции для специалистов в сфере недвижимости.

1.5 Начало работы

Чтобы воспользоваться glmnetUI:

1 . Установка earthUI в R: install.packages("earthUI") — процедура glmnet включена. (См. раздел «Установка earthUI и vProlog» в статье об earthUI в этом выпуске, где описаны версия для разработчиков версии и дополнительных компонентов.)
2 . Запустите приложение : выполните команду earthUI::launch_glmnet() в R. Приложение откроется в вашем веб- браузере на порту 7879. Вы также можете получить доступ к приложению напрямую, перейдя по адресу http://localhost:7879 в своём браузере. Приложение запоминает последний использованный режим назначения и автоматически восстанавливает его.

3 . Создайте или откройте проект в разделе 1 боковой панели. В проекте задаются цель и место работы (страна, штат, округ, город), а также папки ввода и вывода. Проекты хранятся в общем корневом каталоге regProj и доступны для других приложений earthUI и mgcvUI. Чтобы переключить режимы, откройте проект с другим назначением.
4 . Импортируйте данные в разделе 2, выбрав файл CSV или Excel из папки вводных данных активного проекта. Добавьте файлы в эту папку, затем нажмите «Обновить» , чтобы отобразить их список.
5 . Импорт из earthUI (необязательно) — раздел 3 позволяет импортировать файл результатов earthUI .rds для использования базисных функций earth с регуляризацией glmnet. Пропустите этот шаг, если у вас нет результатов earthUI.
6 . Настройка переменных — выберите целевую переменную и предикторы, укажите типы данных, ожидаемые знаки и назначьте специальные роли столбцам.
7 . Настройка параметров glmnet — альфа, лямбда, семейство, ограничения по знаку, взаимодействия и другие параметры.
8 . Подгонка модели — нажмите «Fit Glmnet Model» и просмотрите результаты на главной панели.
9 . Экспорт — скачайте прогнозы в формате Excel, сгенерируйте и конвертируйте отчёт Quarto или (в режиме оценки) рассчитайте корректировки RCA и таблицу сравнения продаж.

Настройки автоматически сохраняются в локальном хранилище вашего браузера и восстанавливаются при повторной загрузке того же файла ввода.

2 Требования к данным для ввода MLS

В рамках рабочих процессов по оценке недвижимости и анализу рынка исходные данные обычно получают из экспорта из системы Multiple Listing Service (MLS). В этой главе описывается ожидаемая структура файла и столбцы, которые может использовать glmnetUI.

2.1 Формат файла & Структура

glmnetUI поддерживает файлы в форматах CSV и Excel (.xlsx, .xls). При импорте названия столбцов автоматически преобразуются в формат snake_case — например, «Living SqFt» превращается в « living_sqft», «Contract Date» — в « contract_date», а «Цена продажи» — в « sale_price». Такая нормализация обеспечивает единообразные ссылки на столбцы на протяжении всего рабочего процесса. Разделитель CSV и знак десятичной запятой, используемые при импорте, определяются настройками локали (см. главу 3, «& ные региональные настройки»).

Ваш файл с данными должен представлять собой плоскую таблицу, в которой на каждую характеристику приходится одна строка, а на каждую характеристику — один столбец. Первая строка файла должна содержать заголовки столбцов.

2.2 Обязательные столбцы для оценки Режим

Хотя glmnetUI работает с любым набором столбцов, для полноценного проведения оценки рекомендуется использовать следующие столбцы:

Столбец	Специальный тип	Назначение
Цена продажи	(целевая)	Переменная отклика для модели
Дата заключения договора	contract_date	Используется для расчета sale_age (количество дней с даты оценки)
Дата размещения объявления	listing_date	Используется вместе с датой заключения договора для расчета DOM, если столбец DOM отсутствует
Дни на рынке	dom	Дни на рынке; отображается в экспортируемых данных
Уступки	concessions	Уступки при продаже; чистая цена продажи (Net SP) = цена продажи− Уступки
Жилая площадь (SF)	living_area	Позволяет рассчитывать остатки на квадратный фут (residual_sf , cqa_sf)
Размер участка	lot_size	Столбец «Размер участка»
Размеры участка	site_dimensions	Группируется по размеру участка
Широта	latitude	Округляется до 3 знаков после запятой
Долгота	longitude	Округляется до 3 знаков после запятой
Идентификатор района	area	Идентификатор рыночного района / микрорайона
Фактический возраст	actual_age	Возраст объекта недвижимости
Эффективный возраст	effective_age	Эффективный возраст объекта недвижимости
Адрес	display_only	Отображается в экспортируемых данных; исключается из модели

Названия столбцов в таблице могут быть на иностранном языке — «специальные» названия указываются на английском языке, чтобы программа R могла обрабатывать их особым образом. В противном случае указанные названия столбцов будут отображаться в регрессионных моделях, графиках и, при проведении оценок, в выводах.

Не все столбцы являются обязательными. glmnetUI адаптируется — если какой-либо столбец отсутствует, соответствующая характеристика просто опускается. Однако для моделей оценки стоимости недвижимости настоятельно рекомендуется использовать определенные столбцы, чтобы добиться приемлемой качества подгонки:

1 . Срок с момента продажи — количество дней между датой заключения договора купли-продажи и датой оценки или анализа. Если используется многолетняя история продаж, особенно за периоды свыше 5 лет, показатель « sale_age » часто играет ключевую роль в оценке цены продажи.
2 . Жилая площадь — также известна под такими названиями, как «Living Sqft», «GLA» (общая жилая площадь) и т. д. Это ещё один основной фактор, определяющий цену продажи.
3 . Общее количество ванных комнат — общее количество полных, четвертных, половинных и 3/4 ванных комнат. Например, две полные ванные комнаты и одна половинная ванная комната будут иметь значение 2,5.
4 . Гаражные места или площадь гаража — количество гаражных мест или площадь гаража в квадратных футах.
5 . Площадь участка (недвижимость) — площадь земельного участка, обычно в квадратных футах или акрах.
6 . Долгота, широта и , если доступно, идентификатор района . Эти географические переменные помогают модели учитывать географические различия в ценах.

2.3 Особые правила именования столбец

glmnetUI идентифицирует столбцы по их специальным обозначениям типов, а не по названиям столбцов. В экспорте в формат MLS вы можете называть столбцы как угодно — главное, чтобы в таблице «Настройка переменных» (глава 6) был назначен правильный специальный тип.

Например, ваша система MLS может экспортировать данные о жилой площади как «GLA», «Living SqFt», «liv _ area» или «gross _ living _ area». После импорта (когда она преобразуется в формат snake_case) вам достаточно просто указать её как « living_area» в раскрывающемся списке «Special». glmnetUI будет затем использовать её для расчёта остатков на квадратный фут независимо от её исходного названия.

2.4 Качество данных& Полнота данных

- Отсутствующие значения (NA) : Строки, содержащие значения NA в любом столбце предиктора или целевого столбца, автоматически удаляются перед подгонкой модели.
- Столбцы с датами : Символьные столбцы, соответствующие распространенным форматам даты, автоматически анализируются при импорте. При отсутствии 4-значного года приоритет отдается 2-значным форматам года.
- Числовые столбцы : Цена продажи, жилая площадь, площадь участка, уступки и аналогичные поля должны быть числовыми. Если ваша система MLS экспортирует цены с символами валют или разделителями тысяч (например, «$ 350,000» или «350.000,00»), вам может потребоваться очистить их перед импортом.
- Столбцы-факторы : текстовые столбцы, столбцы TRUE/FALSE и столбцы R-факторов автоматически обрабатываются как категориальные — регрессия может обрабатывать их только как факторы. Числовой столбец (например, числовое значение площади или код стиля) становится фактором только при установке флажка «Factor» ; флажок

glmnetUI никогда не определяет категориальность на основе диапазона значений числового колонка, поэтому дискретные числовые предикторы, такие как bath_count , остаются непрерывными. Роль «Special» не зависит от выбора фактора.

Совет: После импорта проверьте столбец «NA» в таблице предикторов. Столбцы с большим количеством пропущенных значений могут привести к исключению строк. Рекомендуется исключить столбцы с высоким процентом пропущенных значений или очистить данные перед импортом.

2.5 Размещение объекта оценки в строке

В режиме оценки в первой строке должна быть указана недвижимость, которая подлежит оценке. Все остальные строки представляют собой аналоги. Строка с объектом оценки исключается из процесса подгонки модели. После подгонки модели она по-прежнему генерирует прогнозы для этой строки.

В режиме «Анализ рыночной зоны» размещение объекта оценки в строке 1 является необязательным.

В режиме «Общий» специальной обработки строк не предусмотрено — все строки обрабатываются одинаково.

3 Режим общего назначения

3.1 Обзор

Режим «General Purpose» (Общего назначения) является режимом по умолчанию при запуске glmnetUI. Он обеспечивает полный рабочий процесс регрессии с использованием метода эластичной сети для любых наборов данных — не только для данных о недвижимости. Вы можете использовать glmnetUI для работы с научными данными, финансового анализа, инженерных исследований или решения любых задач регрессии.

В общем режиме в интерфейсе отсутствуют функции, специфичные для рынка недвижимости (специальные столбцы, срок продажи, округление координат, RCA). Боковая панель упрощена, чтобы сосредоточить внимание на отборе переменных, настройке параметров, подгонке модели и экспорте.

3.1.1 Пропустить первую строку

В верхней части боковой панели появляется флажок «Пропустить первую строку», как только открывается проект типа «Общий» или «Рынок». Если флажок установлен, первая строка исключается из процесса подгонки модели. Это полезно, когда первая строка содержит целевое или эталонное наблюдение, которое не должно влиять на модель. В режиме «Оценка» строка 1 (объект оценки) всегда исключается автоматически.

3.1.2 Настройки в зависимости от цели

Настройки (выбор предикторов, параметры модели, взаимодействия) сохраняются отдельно для каждой комбинации входного файла и режима назначения. При переключении между режимами «Общий», «Оценка» и «Рынок» настройки каждого режима сохраняются независимо друг от друга.

3.2 Рабочий процесс с боковой панелью

Боковая панель состоит из пронумерованных сворачиваемых разделов, которые помогают вам пройти весь путь от выбора проекта до экспорта отчёта. Большинство разделов отображаются только при открытии активного проекта, а некоторые из них зависят от цели проекта. Папка для вывода данных определяется на основе активного проекта — отдельного поля для указания папки вывода данных нет.

1. Проект — создание или открытие проекта. В проекте задаются цель и место проведения работ (страна, штат, округ, город), а также папка ввода (<os>_in/) и папка вывода. Проекты хранятся в общей корневой папке regProj (по умолчанию ~/regProj, значение можно переопределить с помощью переменной среды REGPROJ_ROOT) и используются совместно с родственными приложениями earthUI и mgcvUI, поэтому модели всех трёх приложений находятся в одной папке. Нажмите «Закрыть проект», чтобы переключиться.

2. Импорт данных — Выберите файл CSV или Excel из папки ввода текущего проекта. Добавьте файлы в папку с помощью Finder/Проводника, затем нажмите «Обновить», чтобы отобразить их список. Для файлов Excel с несколькими листами появится окно выбора листа. Названия столбцов автоматически преобразуются в формат «snake _».

3. Импорт из earthUI (необязательно) — Импортируйте файл earthUI с результатами в формате .rds, чтобы использовать базисные функции «hinge» из earthUI в сочетании с регуляризацией «эластичной сети» из glmnet. Кнопка «Обзор» оформлена в стиле, соответствующем окну выбора файлов из раздела 2. Этот шаг является необязательным — пропустите его, если у вас нет результата earthUI для импорта.

4. Настройка переменных — селектор целевой переменной, таблица предикторов с флажками «Включить», «Фактор» и «Принудительно», а также раскрывающиеся списки «Знак». Столбец «Специальное» (для указания дат заключения договоров, координат и т. д.) отображается только в режимах «Оценка» и «Рынок». Подробную информацию см. в главе 6.

5. Параметры вызова glmnet — все настройки модели: alpha, lambda, family, standardize, sign constraints, relaxed lasso, начальное значение секретного ключа, матрица взаимодействий, а также расширенные параметры (lambda.min.ratio, nlambda, метрика потерь CV, порог сходимости, максимальное количество итераций, пересечение). Полный справочник по параметрам см. в главе 7.

6. Подгонка модели Glmnet — кнопка, запускающая модель. Результаты отображаются на вкладках справа.

7. Скачать результаты — экспортирует прогнозы, остатки, балы CQA и вклады по каждой переменной в виде файла Excel в папку с выводами проекта. Доступно во всех режимах после подгонки модели.

8. Расчет корректировок RCA & Скачать — Вычисляет корректировки RCA по каждой переменной для каждого аналога и интерполирует остаток по объекту оценки с помощью CQA. Отображается только в режимах «Оценка» и «Рынок» после подгонки модели. См. главу 11.

9. Сгенерировать таблицу продаж & Скачать — Генерирует промежуточную таблицу сравнения продаж, в которой аналоги ранжируются по проценту валовой корректировки. Отображается только в режиме оценки после вычисления корректировок RCA. См. главу 12.

10. Сгенерировать отчёт Quarto — создает автономный набор отчета Quarto (.qmd) — со всеми графиками и таблицами — в папку вывода проекта. Функция становится доступной после подгонки модели.

11. Конвертировать отчёт Quarto — преобразует сгенерированный набор .qmd в формат HTML, Word или PDF. Для создания PDF-файла требуется установка LaTeX; если она не обнаружена, опция PDF будет скрыта. См. главу 13.

В режимах «Рынок» и «Общий» скрытые этапы, предназначенные исключительно для оценки, пропускаются, и два раздела отчёта перенумеровываются соответствующим образом.

3.3 Проекты

glmnetUI организует работу в рамках проектов, расположенных в общей корневой папке regProj (в том же дереве, что и earthUI и mgcvUI). Каждый проект имеет собственную папку для вводных данных (<os>_in) и отдельные для каждого движка папки для вывода (например, <os>_out_glmnet).

Чтобы создать проект, нажмите кнопку «+ Новый» (. . .) в разделе 1 и укажите:

- Цель — «Общая», «Оценка» или «Рыночная зона».
- Местоположение — политико-административная иерархия проекта. Набор уровней, которые необходимо указать, зависит от выбранной вами страны, поэтому иностранные системы наименований поддерживаются автоматически — в проекте в США используются штат/округ/город, во французском — регион/департамент/коммуна, в японском — префектура/город или район/округ и т. д.
- Название проекта — короткое название, длиной не более 8 символов (буквы, цифры, _ и -).

Приложение не использует введенное вами имя дословно: оно формирует уникальное имя папки с отметкой времени, добавляя в начало коды местоположения и дату-время создания в формате <country>_<admin level codes>_<YYYYMMDD-HHMMSS>_<your name> —

например, проект в США с именем « burl» становится « us_ca_081_burlin_20260604-131500_burl». Ограничение в 8 символов применяется только к той части, которую вы вводите. Префикс с отметкой времени добавляется только к вновь созданным проектам; проекты, уже находящиеся на диске, сохраняют свои исходные имена и остаются полностью доступными.

3.4 Импорт из earthUI (Необязательно)

Раздел 3 боковой панели — «Импорт из earthUI» — позволяет импортировать файл результатов earthUI .rds. Благодаря этому glmnetUI может использовать базисные функции earth типа «hinge» (кусочно-линейные слагаемые) в сочетании с регуляризацией «эластичной сети» из glmnet, объединяя автоматическое обнаружение нелинейности earth с отбором переменных и ограничениями на коэффициенты glmnet.

При импорте результатов earthUI подгонка модели glmnet определяется предикторами модели earth — а не флажками «Включить», — поскольку glmnet выполняет подгонку на основе базисного разложения earth. Таблица «Конфигурация переменных» явно отражает это: поля «Включить» и «Тип» заблокированы, строки для предикторов, не использованных моделью earth, отключены, а в примечании поясняется, что набор предикторов берется из модели earth. Поля «Сила» и «Знак» остаются доступными для редактирования для предикторов earth, поскольку они по-прежнему применяются к базису earth во время подгонки.

Кнопка «Обзор» оформлена в стиле, соответствующем выбору файлов из раздела 2. Этот шаг является полностью необязательным — если у вас нет файла результатов earthUI, просто пропустите раздел 3 и перейдите к разделу 4.

3.5 Вкладки главной панели

После подгонки на главной панели появляются следующие вкладки:

Вкладка	Содержание
Предварительный просмотр данных	Предварительный просмотр импортированных данных. В режиме оценки данные разбиты на таблицы «Объект оценки» и «Аналоги».
Уравнение	Уравнение построенной модели, отображаемое в формате LaTeX/MathJax, с указанием каждого ненулевого коэффициента и членов взаимодействия.
Корреляция	Тепловая карта корреляций числовых предикторов (до подгонки модели).
Сводка	Ключевые метрики (R^2 , Adj R^2 , GR^2 , CV R^2 , RMSE, MAE) и метрики оценки (COD, PRD, Median Ratio).
Коэффициенты	Таблица ненулевых коэффициентов с предупреждениями о нарушении знака.
Важность переменных	Гистограмма и таблица ранжирования важности предикторов ($\|\beta\| \times \text{sd}(x)$). Интерактивный Plotly-ховер, если доступен.
Вклады	Графики частичных эффектов по каждой переменной с метками наклона. Точечный график + линия подгонки для числовых предикторов, бокс-плоты для факторов, 3D-поверхность/точечный график/тепловая карта для взаимодействий.
ANOVA	Таблица декомпозиции дисперсии: SS по каждой переменной,% SS модели, коэффициент.
Диагностика	Четыре графика: путь коэффициента, ошибка CV, фактические значения против прогнозируемых, остатки против подгонки.
Результаты Glmnet	Исходные результаты модели: использованное начальное значение случайного числа, вывод модели, выбранное значение lambda, lambda.min/1se, gamma (если ослаблено) и полный вектор коэффициентов.
Отчёт	Настройки отчёта и поля экспорта.
Корректировки RCA	Результаты анализа RCA (режим оценки, после вычислений).

3.6 Сохранение настроек

glmnetUI автоматически сохраняет вашу конфигурацию в локальном хранилище браузера, используя в качестве ключа имя файла для ввода. При повторной загрузке того же файла все настройки восстанавливаются: выбор целевой переменной, флажки предикторов, типы данных, ожидаемые знаки, параметры glmnet и матрица взаимодействий. Режим назначения, использовавшийся последним, также сохраняется глобально и восстанавливается при повторном запуске приложения. Доступны три варианта:

- Использовать последние настройки для файла ввода — восстановить сохраненные настройки для этого конкретного файла
- Использовать настройки по умолчанию — применить сохраненные глобальные настройки по умолчанию
- Настройки glmnet по умолчанию — восстановить заводские настройки по умолчанию (alpha = 1, CV lambda.1se, 10 блоков, гауссово распределение, стандартизация включена)

Кнопка «Сохранить текущие настройки в качестве стандартных» позволяет сохранить все текущие настройки в качестве глобальных настроек по умолчанию.

3.7 Темный режим

Нажмите на значок луны/солнца в правом верхнем углу, чтобы переключаться между светлой и тёмной темами. Настройки темы сохраняются в локальном хранилище и сохраняются между сеансами. Все элементы интерфейса (таблицы, графики, карточки, кнопки) адаптируются к выбранной теме с помощью переменных CSS.

3.8 Язык и регион & Региональные настройки

glmnetUI поддерживает международные стандарты формата чисел и CSV-файлов благодаря системе локалей, основанной на странах. В раскрывающемся списке «Настройки» в заголовке окна доступны переключатели «Страна» и «Формат» для 31 поддерживаемой страны. Каждая предустановка настраивает:

- Разделитель CSV — запятая (,) для США, Великобритании и Японии или точка с запятой (;) для большинства стран Европы, где запятая используется в качестве десятичного знака.
- Десятичный знак — точка (.) или запятая (,).
- Разделитель тысяч — запятая (США/Великобритания/Япония), точка (Германия/Италия/Испания), пробел (Финляндия/Франция/Польша/Прибалтика/Украина/Россия) или апостроф (Швейцария).
- Формат бумаги — Letter (США/Канада/Мексика) или A4 (везде else).

3.8.1 Сохранение настроек по умолчанию

Нажмите «Сохранить как значение по умолчанию», чтобы сохранить настройки локали глобально. Эти настройки по умолчанию будут применяться ко всем будущим сеансам независимо от того, какой файл данных вы загрузите.

Совет: формат чисел на осях графика и в метках наклона автоматически адаптируется к вашей языковой настройке. В немецкой языковой настройке для разделителя тысяч используются точки (200.000), в финской — пробелы (200 000), а в швейцарской — апострофы (200’000). Символы валют не отображаются — glmnetUI не зависит от валюты.

4 Режим оценки

При выборе цели « Для оценки» glmnetUI настраивается на оценку одного объекта недвижимости. Все функции, описанные в главе 3, остаются доступными; в данной главе рассматриваются только дополнительные возможности, специфичные для оценки.

4.1 Обработка строк объекта оценки

В режиме оценки первая строка вашего набора данных представляет собой объект оценки, а все остальные строки — аналогичные сделки. Ваш вводный файл должен быть организован соответствующим образом (см. главу 2). Цена продажи объекта может быть оставлена пустой или установлена на любое значение — glmnetUI автоматически рассматривает ее как NA во время подгонки модели.

После импорта вкладка «Предварительный просмотр данных» разделяется на два раздела: «Объект оценки» (строка 1) и «Сопоставимые сделки» (строки 2 и далее). Строка 1 всегда исключается из подгонки модели. После подгонки модели модель по-прежнему генерирует прогнозы для строки, относящейся к данному объекту оценки.

4.2 Дата оценки & Возраст, с которого разрешена продажа

В режимах «Оценка» и «Рынок» в разделе «Настройка переменных» появляется поле «Дата оценки» (по умолчанию установлена сегодняшняя дата). Если вы выберете для столбца значение « contract_date» в раскрывающемся списке «Специальный столбец», glmnetUI рассчитает столбец « sale_age» — количество целых дней между датой заключения каждого договора купли-продажи и датой оценки. Этот столбец добавляется в качестве предиктора.

При изменении даты оценки показатель « sale_age» автоматически пересчитывается.

4.3 Специальный столбец Обозначения

В режимах оценки и рынка для каждого предиктора в таблице «Конфигурация переменных» появляется специальный раскрывающийся список. Полную справку по специальным типам и их эффектам см. в главе 6.

4.4 Обзор корректировок RCA

Кнопка «Рассчитать корректировки RCA» & (боковая панель, раздел 8, отображается в режимах «Оценка» и «Рынок» после подгонки) позволяет рассчитать корректировки, основанные на рыночных данных, для каждого аналога по отношению к объекту оценки. Полный рабочий процесс RCA описан в главе 11.

Совет: Балл CQA, присваиваемый объекту оценки, определяет, какая доля остатков распределения относится именно к этому объекту. Балл 5,00 означает, что объект оценки находится на медиане среди аналогов.

5 Режим анализа рыночной зоны

Если в качестве цели выбрать «Анализ рыночной зоны», glmnetUI предоставляет те же функции, характерные для рынка недвижимости, что и в режиме оценки (специальные столбцы, срок с момента продажи, округление координат), но ориентирован на анализ группы объектов недвижимости, а не на оценку отдельного объекта.

5.1 Отличия от режима оценки

- Обработка строк с объектами оценки осуществляется гибко — в режиме рынка строка 1 не рассматривается автоматически как объект оценки.
- Отсутствует таблица сравнительных продаж — шаг «Расчет корректировок RCA & Скачать» (раздел 8) доступен в рыночном режиме, но шаг «Таблица продаж» (раздел 9) доступен только в режиме оценки.

5.2 Когда следует использовать режим «Рынок»

Режим «Анализ рыночной зоны» подходит в следующих случаях:

- Построение регрессионной модели для района или рыночной зоны с целью понимания факторов, влияющих на стоимость
- Анализ того, как такие переменные, как площадь, возраст, площадь участка и расположение, влияют на цены продажи в группе объектов недвижимости
- Подготовка материалов для анализа рыночной конъюнктуры или
- Работа с набором данных, в котором требуются специальные свойства столбцов (срок продажи, округление координат), но не требуется рабочий процесс корректировки RCA

Совет: Режим «Market» также полезен для общей регрессионного анализа рынка недвижимости, когда требуются специальные характеристики столбцов, но не нужен рабочий процесс корректировки RCA.

6 Отбор переменных

В разделе 4 боковой панели — «Настройка переменных» — вы можете выбрать, какие столбцы будут включены в модель и как с ними будет производиться работа.

6.1 Целевая переменная

В раскрывающемся списке «Целевая (реакционная) переменная» в верхней части раздела 4 перечислены все столбцы вашего набора данных. Выберите один столбец в качестве целевой переменной (например, «Цена продажи»). Целевой столбец автоматически исключается из списка предикторов.

6.2 Таблица предикторов

Ниже селектора целей в таблице перечислены все оставшиеся столбцы со следующими полями:

Столбец	Описание
Переменная	Название столбеца
Тип	Раскрывающийся список типов данных: numeric , integer , character , Date , POSIXct
Включить	Флажок — включить этот столбец в качестве предиктора в модель
Фактор	Флажок — рассматривать переменную как категориальный фактор (создает фиктивные столбецы в матрице модели)
Флажок «Force»	— принудительно включить в модель (коэффициент штрафа = 0, никогда не исключается методом Lasso)
«Special»	Раскрывающийся список (только для оценки/рынка) — см. справочник по специальным типам столбцов ниже
«Sign»	Знак ожидаемого коэффициента: положительный, отрицательный или любой
«NAs»	Количество отсутствующих значений

6.3 Определение типов данных& Переопределения

glmnetUI автоматически определяет типы данных при импорте. Распознаются числовые, целочисленные и даты столбцы. Символьные столбцы, которые выглядят как даты (типичные форматы дат), классифицируются как « Date».

Вы можете переопределить любое распознавание, изменив значение в раскрывающемся списке «Тип». Изменение типа влияет на то, как столбец кодируется в матрице модели.

Столбцы «Текст», «TRUE/FALSE» и «R-фактор» автоматически рассматриваются как категориальные. Числовой столбец становится фактором только в том случае, если установить флажок «Фактор» — glmnetUI никогда не определяет категориальность на основе диапазона значений числового столбца, поэтому дискретные числовые предикторы, такие как количество ванн, остаются непрерывными, если вы не укажете иное. Настройка «Фактор» сохраняется в localStorage.

6.4 Ожидаемые признаки& Обеспечение соблюдения

Для каждого предиктора в раскрывающемся списке «Знак» указывается, ожидается ли, что его коэффициент будет положительным, отрицательным или может принимать значения

любого знака. После подгонки модели:

- Коэффициенты с неожиданными знаками помечаются в таблице «Коэффициенты»
- Предупреждения о знаках отображаются на вкладке «Сводка»

Если в параметрах glmnet (глава 7) установлен флажок «Enforce Sign Constraints», модель вынуждена вычислять коэффициенты, соответствующие ожидаемым знакам. Это реализовано с помощью функций upper.limits и lower.limits в пакете glmnet.

6.5 Специальные типы столбцов Ссылка

В режимах «Оценка» и «Рынок» в раскрывающемся списке «Специальное» доступны следующие параметры:

Взвешивание:

- weight — Столбец с весами наблюдений (доступна во всех режимах; допускается только одна; строки с весом = 0 исключаются из подгонки)

Дата& . Типы времени:

- contract_date — Запускает автоматический расчет показателя « sale_age » с даты оценки
- listing_date — Используется в качестве резервного варианта для расчета показателя «Days on на рынке
- dom — Определяет столбец «Дни на рынке»
- sale_age — Указывает существующий столбец «Срок продажи», благодаря чему любой столбец может служить показателем срока продажи вместо того, чтобы glmnetUI вычислял его на основе contract_date

Типы денежных единиц:

- concessions — Определяет условия скидок при продаже (кредиты продавца, поощрения для покупателя и т. д.)

Типы транзакций:

- sale_type — Обозначает столбец с типом продажи/транзакции

Размер& . Типы мест:

- latitude — Значения автоматически округляются до 3 знаков после запятой
- longitude — Округление аналогично широте
- area — Идентификатор рыночной зоны или района (например, area_ id, район, рынок_ area)
- living_area — Включает расчёт остаточной стоимости на квадратный фут (residual_sf и cqa_sf)
- lot_size — Столбец «Размер участка»
- site_dimensions — Группируется по площади участка (например, «75x120»)

Возрастные группы:

- actual_age — Столбец «Возраст объекта недвижимости»
- effective_age — Фактический возраст объекта недвижимости

Типы дисплеев:

- display_only — Столбец включается в экспортные файлы Excel, но полностью исключается из процесса подгонки модели. Используйте эту настройку для полей адреса, номеров MLS, идентификаторов земельных участков или других справочных данных. Эту настройку можно применить к нескольким столбцам.

Совет: Столбцы, помеченные как « display_only» (не используемые в модели), остаются в наборе данных и отображаются при экспорте в Excel, но полностью исключаются из таблицы предикторов. Используйте эту функцию для столбцов с идентификаторами, адресами или другими справочными полями.

7 Выбор параметров

Раздел 5 боковой панели — «Параметры вызова glmnet» — предоставляет доступ ко всем параметрам настройки модели эластичной сети. У каждого параметра есть синий значок справки (?) с всплывающей подсказкой, содержащей пояснение.

7.1 Alpha (параметр микширования)

Параметр alpha определяет тип регуляризации:

Значение альфа	Тип	Поведение
0	Ридж	Пропорционально уменьшает все коэффициенты; никогда не устанавливает их значение равным нулю
1	Лассо	Может устанавливать коэффициенты равными нулю (отбор переменных)
от 0 до 1	Эластичная сеть	Сочетание методов «Ридж» и «Лассо»

Доступны два метода выбора альфа-значений:

- Фиксированное значение — Установите одно значение альфа с помощью ползунка (по умолчанию: 1,0)
- Поиск по таблице — Проведите тестирование нескольких значений альфа и выберите то, которое дает наименьшую ошибку CV. Настройте диапазон и количество значений.

7.2 Lambda Selection

Параметр lambda определяет степень регуляризации:

Метод	Описание
Перекрёстная проверка	Рекомендуется. Проводит тестирование множества значений lambda с использованием k-кратной перекрёстной проверки.
Руководство	Укажите конкретное значение lambda, если у вас есть на то основание.

При использовании перекрёстной проверки доступны два варианта выбора лямбды:

- lambda.1se (по умолчанию) — Наибольшее значение lambda в пределах 1 стандартной ошибки от минимального значения. Позволяет получить более простые и лучше регуляризованные модели. Рекомендуется для работ по оценке.
- lambda.min — Значение лямбды, минимизирующее ошибку перекрёстной проверки. Позволяет получить модель, наилучшим образом подходиющую к данным, но потенциально более сложную.

Параметр «Количество блоков CV» определяет, как данные будут разделены для перекрёстной проверки (по умолчанию: 10 блоков).

7.3 Семья

Выберите семейство распределений для вашей зависимой переменной:

Семейство	Пример применения
gaussian	Непрерывные значения (например, цена продажи). Наиболее распространенный случай.
binomial	Бинарные исходы (например, продано/не продано).
poisson	Данные, выраженные в виде количества (например, количество продаж).

7.4 Стандартизировать

Если этот флажок установлен (по умолчанию), перед подгонкой все предикторы нормируются так, чтобы их среднее значение равнялось 0, а стандартное отклонение — 1. Это гарантирует, что штрафная функция будет одинаково учитывать все предикторы независимо от их исходного масштаба. Коэффициенты возвращаются в исходном масштабе. Обычно этот флажок следует оставлять установленным.

7.5 Ограничения знаков

Если включена опция «Принудительное соблюдение ограничений по знакам», коэффициенты ограничиваются таким образом, чтобы они соответствовали ожидаемым знакам, заданным в таблице переменных (глава 6):

- Знак «положительный» принудительно устанавливает коэффициент≥ 0
- Знак «отрицательный» принудительно устанавливает коэффициент≤ 0
- Переменные, для которых установлено значение «либо», не имеют ограничений

Это реализуется с помощью параметров upper.limits и lower.limits в файле glmnet().

7.6 «Расслабленное лассо»

Обычный метод Lasso применяет одинаковый штраф как при отборе переменных, так и при оценке коэффициентов, что может привести к чрезмерному сжатию важных коэффициентов в сторону нуля. Метод Relaxed Lasso разделяет эти этапы:

1 . Сначала выберите переменные с помощью инструмента «Лассо» (штраф определяет, какие коэффициенты останутся в модели)
2 . Затем выполните повторную подгонку оставшихся переменных с меньшим штрафом или без него для получения менее смещенных оценок

Ползунок «Gamma» регулирует степень смягчения:

Гамма-эффект	Эффект
0	Полностью релаксированная подгонка (переподгонка по OLS, без какого-либо сжатия)
1	Без релаксации (так же, как в обычном glmnet)
Межгрупповой	Сочетает релаксированную и штрафную подгонки

При перекрёстной проверке коэффициент гамма выбирается автоматически для обеспечения оптимальной производительности.

7.7 Начальные значения для генератора случайных чисел

Ввод текста «Random seed», предварительно заполненный случайным целым числом, обеспечивает воспроизводимость распределения блоков при перекрёстной проверке — перед

подгонкой вызывается функция set.seed(), а использованное начальное значение отображается на вкладке «Glmnet Output» и в сообщении о статусе подгонки. После каждой подгонки автоматически генерируется новое начальное значение (с использованием системного времени, а не генератора случайных чисел R). Последние 5 начальных значений, использованных для текущего файла, отображаются в виде активных ссылок для удобного доступа; история начальных значений сбрасывается при загрузке нового файла .

7.8 Матрица взаимодействий

Интерактивная верхнетреугольная матрица позволяет управлять тем, какие пары переменных могут взаимодействовать. В каждой ячейке находится флажок — если он установлен, это означает, что член взаимодействия x1:x2 включен в матрицу модели.

- «Разрешить всё» — Установить флажки для всех пар взаимодействий
- «Снять всё» — Снять флажки со всех пар (по умолчанию в новом файле все флажки сняты)
- Щелкните по имени переменной — Включить/выключить все взаимодействия для этой переменной

Настройки взаимодействия сохраняются в локальном хранилище браузера (localStorage) для каждого файла ввода и восстанавливаются при повторной загрузке того же файла.

Если какое-либо взаимодействие включено, экспортированный отчёт содержит матрицу «Включенные взаимодействия» (предикторы по обеим осям, галочка в тех случаях, где был включен термин продукта « x1:x2»), отображаемую в альбомной ориентации и повторяющую матрицу «Допустимые взаимодействия» из earthUI. Обратите внимание, что коэффициент включенного термина все же мог быть сведён к нулю с помощью инструмента «Лассо».

Совет: При большом количестве предикторов метод Lasso автоматически обнуляет ненужные термы взаимодействия. Разумным подходом будет начать с параметра «Allow All» и позволить модели самостоятельно выбрать нужные термы.

Warning

Включение взаимодействий в модели glmnet приведет к корректировкам, которые может быть сложно интерпретировать и объяснить в суде или в ходе аудита. Используйте взаимодействия с осторожностью и убедитесь, что полученные корректировки являются обоснованными.

7.9 Блок от основного эффекта (только взаимодействие)

Щелкните правой кнопкой мыши по имени переменной в матрице взаимодействий, чтобы заблокировать её включение в модель в качестве основного эффекта. После имени переменной появится выделенный жирным шрифтом индикатор « 1». При блокировке:

- Переменная удаляется из слагаемых основной формулы
- Переменная по-прежнему может фигурировать в слагаемых взаимодействия (x1:x2)
- Снова щелкните правой кнопкой мыши, чтобы разблокировать
- Состояние сохраняется в localStorage для каждого файла ввода

Типичный пример применения: при моделировании временных корректировок, которые должны зависеть от размера недвижимости, следует заблокировать влияние « sale_age» в качестве основного эффекта и допустить его только во взаимодействии с « living_area». В результате получается взаимодействующий член « sale_age:living_area» вместо постоянного коэффициента « sale_age» — темпы роста стоимости с течением времени варьируются в зависимости от размера дома.

7.10 Расширенные параметры

Сворачиваемый раздел «Advanced» в нижней части страницы «Call Parameters» в glmnet открывает доступ к дополнительным настройкам. Для них используются разумные значения по умолчанию, но они отображаются на экране, чтобы все настройки модели можно было задокументировать для судебного разбирательства или аудита.

- Коэффициент Lambda Min — Отношение наименьшего значения лямбды к наибольшему в траектории регуляризации. По умолчанию: 0,00001. Для небольших наборов данных (<50 наблюдений) более высокое значение (например, 0,01) может улучшить результаты.
- Количество значений лямбды — Количество значений лямбды в траектории. По умолчанию: 100.
- Метрика потерь при перекрёстной проверке — Функция потерь для перекрёстной проверки: MSE (по умолчанию для гауссового распределения), MAE (устойчивая к выбросам) или Deviance.
- Порог сходимости — Порог сходимости метода координированного спуска . По умолчанию: 1e-07.
- Максимальное количество итераций — Максимальное количество итераций метода координированного спуска. Значение по умолчанию: 100 000. Увеличьте, если появляются предупреждения о сходимости.
- Включить подгонку — включить член подгонки (константу). Значение по умолчанию: включено. Член подгонки является «базовым значением» в оценке RCA.

7.11 Настройки по умолчанию

Существует три параметра, определяющих порядок инициализации параметров при загрузке файла:

- Использовать последние настройки для файла ввода — восстановить точные настройки, использовавшиеся в последний раз с этим файлом
- Использовать настройки по умолчанию — применить сохраненные глобальные настройки по умолчанию
- Настройки по умолчанию glmnet — сбросить до заводских настроек (alpha = 1, CV lambda.1se, 10 блоков, гауссово распределение, стандартизация включена)

Нажмите «Сохранить текущие настройки в качестве стандартных», чтобы сохранить все текущие настройки в качестве глобальных настроек по умолчанию для будущих файлов.

8 Подгонка модели

8.1 Кнопка «Fit»

В разделе 6 боковой панели находится кнопка «Fit Glmnet Model» (Построить модель Glmnet). При нажатии на эту кнопку модель будет запущена с текущими настройками.

8.2 Что происходит во время подгонки

При нажатии кнопки «Подгонка» glmnetUI:

1 . Подготовка данных — удаление строк с значениями «NA», применение весов (если указаны), исключение строки «объект оценки» (в режиме оценки) и кодирование факторных переменных в виде фиктивных столбцов.
2 . Постраивает матрицу модели — создает матрицу дизайна, включая все допустимые термы взаимодействия из матрицы взаимодействий.
3 . Применяет ограничения — если включено принудительное соблюдение знаков, устанавливает upper.limits и lower.limits для коэффициентов.
4 . Выполняет перекрёстную проверку — если выбрана CV, вызывает cv.glmnet() для нахождения оптимального значения lambda. Если включен метод «Relaxed Lasso», вызывается функция cv.glmnet(..., relax = TRUE) .
5 . Подгоняет окончательную модель — сохраняет подгоняемую модель, коэффициенты, прогнозы и остатки.
6 . Обновляет все вкладки результатов — «Сводка», «Коэффициенты», «Уравнение», «Важность переменных», «Вклады», «ANOVA» и «Диагностика» заполняются данными.

В сообщении о состоянии под кнопкой «Fit» отображается количество наблюдений, исключенных строк и выбранное значение лямбды.

9 Вкладки «Результаты»

9.1 Предварительный просмотр данных

Отображает импортированные данные в виде интерактивной таблицы DataTable. В режимах «Оценка» и «Рынок» предварительный просмотр разбит на две таблицы: «Объект оценки» (строка 1) и «Сопоставимые сделки» (строки 2 и далее).

Каждая ячейка отображается в одной строке и обрезается до ширины столбца, поэтому широкие поля свободного ввода (например, примечания к объектам недвижимости) не растягивают строку вниз по странице. Дважды щелкните любую ячейку, чтобы открыть всплывающее окно с её полным, необрезанным содержимым.

9.2 Уравнение

Отображает уравнение модели, созданной в процессе подгонки, сгенерированное в LaTeX с помощью MathJax. Показывает все ненулевые коэффициенты с надлежащим математическим форматированием:

- Положительные коэффициенты обозначаются знаком «+
- Отрицательные коэффициенты обозначаются знаком «−
- Слагаемые взаимодействия отображаются как «$x_1 \times x_2$
- Числа форматируются с запятыми и соответствующим количеством знаков после запятой

9.3 Корреляция

Матрица тепловой карты, отображающая корреляции Пирсона между всеми числовыми предикторами и зависимой переменной. Доступна сразу после импорта данных — не требуется подгонка модели.

- Цветовой градиент: синий (-1)→ белый (0)→ красный ($+1$)
- Значения корреляции отображаются в каждой ячейке
- Размер текста и осей адаптируется в зависимости от количества переменных
- Полезно для выявления мультиколлинеарности перед подгонкой

9.4 Резюме

Статистика подгонки модели, отображаемая в виде карточек:

- R^2 — доля объясненной дисперсии
- Adj R^2 — корректированное значение R с учетом количества предикторов
- GR^2 — обобщенное R^2 , полученное на основе коэффициента отклонения из glmnet при выбранном значении лямбды
- CV R^2 — перекрёстная проверка R^2 (при использовании CV)
- RMSE — среднеквадратичная ошибка
- MAE — средняя абсолютная ошибка

В режимах оценки и рынка отображаются дополнительные карточки:

- Медианное соотношение — медиана отношения (прогнозное / фактическое)
- COD — коэффициент дисперсии (чем ниже, тем лучше)
- PRD — ценовой дифференциал

Предупреждение о переобучении появляется, когда при обучении R^2 превышает CV R^2 более чем на 0,1.

Под карточками для удобства расположена таблица коэффициентов с указанием знаков, которая дублирует вкладку «Коэффициенты».

9.5 Коэффициенты

Таблица всех ненулевых коэффициентов, в которой указано:

- Имя переменной
- Значение коэффициента
- Ожидаемый знак (из конфигурации переменной)
- Флаг предупреждения о знаке, если направление коэффициента не соответствует ожиданиям

9.6 Важность переменных

Величина стандартизированного коэффициента: $|\beta| \times \mathrm{sd}(x)$ для каждого предиктора, агрегированная по фиктивным столбцам для факторных переменных. Два режима отображения:

- Интерактивный Plotly (при наличии установленного пакета plotly): Горизонтальная гистограмма с всплывающими подсказками, отображающими точные значения значимости, коэффициента и относительной доли.
- Статический ggplot2 (резервный вариант): Горизонтальная гистограмма с подписями на осях, отформатированными с запятыми.

Под диаграммой в таблице DataTable приведены значения переменных, их значимость, коэффициенты и относительные значения %.

9.7 Материалы

Графики частичных эффектов по каждой переменной, отражающие вклад каждого предиктора в прогнозы модели. Выберите переменную из раскрывающегося списка.

Для числовых предикторов: диаграмма рассеяния, отражающая зависимость значений переменной по оси x от её вклада ($\beta \times x$), с красным отрезком линии и подписью, указывающей наклон. Метка наклона показывает маргинальный эффект на единицу (например, +1,234.56/unit) с адаптивными единицами для переменных с небольшим диапазоном значений, таких как широта/долгота (например, +0.12/0.001). Подзаголовок показывает значение пересечения с осью (базис).

Для факторных переменных: бокс-плат, отражающий распределение значений вклада по уровням факторов.

Для взаимодействий между двумя переменными: три наложенные друг на друга визуализации — трехмерный график поверхности (интерактивный Plotly, если доступен; в противном случае — статический persp()) совместного вклада, диаграмма рассеяния с точками, окрашенными в зависимости от вклада, и тепловая карта среднего вклада по всем диапазонам значений обеих переменных.

9.8 ANOVA

Таблица декомпозиции дисперсии, в которой для каждого предиктора указано:

- SS — Сумма квадратов: $\sum(\mathrm{contribution}_j - \overline{\mathrm{contribution}_j})^2$
- %модели SS — Доля от общей суммы квадратов модели
- Коэффициент — Доминирующий коэффициент для данного предиктора

Включает строку пересечения и строку «TOTAL MODEL». Термы взаимодействия (содержащие «:») отображаются в виде отдельных групп.

9.9 Диагностика

Четыре графика для диагностики с крупным шрифтом (базовый размер 16pt), делениями по осям с шагом 15 и метками в формате с запятой (без научной нотации):

1 . Траектория коэффициентов — Показывает, как изменяются все коэффициенты при увеличенииlog(λ) . Линии окрашены в цвета, соответствующие переменным. Помогает визуализировать траекторию регуляризации и отбор переменных.
2 . Ошибка перекрёстной проверки — Ошибка перекрёстной проверки как функция log(λ) . Точки отображают среднюю ошибку перекрёстной проверки; полосы погрешности показывают ± 1 стандартную ошибку. Пунктирные вертикальные линии обозначают $\lambda_{\min}$ (синий) и$\lambda_{1\text{se}}$ (зеленый).
3 . Фактические значения против прогнозируемых — Диаграмма рассеяния наблюдаемых значений против прогнозируемых. Точки должны группироваться вокруг 45-градусной пунктирной линии отсчета. Более широкое рассеяние указывает на большую ошибку прогнозирования.
4 . Остатки по сравнению с результатами подгонки — Диаграмма рассеяния остатков по сравнению с результатами подгонки. Должно наблюдаться случайное рассеяние вокруг нулевой линии (пунктирной). Определённые закономерности указывают на некорректную спецификацию модели (например, нелинейность, гетероскедастичность).

На каждом графике есть кнопка «Скачать PNG» для сохранения изображения с разрешением 150 DPI.

9.10 Вывод Glmnet

Исходный вывод модели, отображающий: использованное начальное значение генератора случайных чисел, вывод модели, выбранное значение lambda, lambda.min/lambda.1se, gamma (если включен режим Relaxed Lasso) и полный вектор коэффициентов.

9.11 Отчёт «

»

Настройте поля отчета (имя оценщика, адрес объекта недвижимости, дату отчета, номер дела) и экспортируйте его в формат Word (.docx) или PDF прямо с этой вкладки. Полную информацию о рабочем процессе подготовки отчетов на базе Quarto (HTML, Word или PDF) см. в главе 13.

10 Загрузка данных

После подгонки кнопка «Скачать вывод (Excel)» (боковая панель, раздел 7, доступна во всех режимах) позволяет экспортировать файл Excel с прогнозами и диагностикой в папку вывода проекта. Заголовок раздела звучит как «Скачать расчетные цены продажи & остатки» в режимах «Оценка» и «Рынок» и как «Скачать расчетные целевые переменные & остатки» в режиме «Общий».

10.1 Столбцы вывода

Колонка	Описание
est_<target>	Прогноз по модели (например, est_sale_price)
residual	Фактическое значение− Прогнозируемое значение
cqa	Оценка «Состояние-Качество-Привлекательность» (по шкале от 0 до 10)
residual_sf	Остаточная площадь / жилая площадь (если указано living_area)
cqa_sf	Показатель CQA, рассчитанный на основе рейтинга по residual_sf
<var>_contribution	Вклад каждой переменной в прогноз
basis	Значение пересечения (одинаковое для всех строк)

После успешного завершения на кнопке загрузки появляется белая галочка.

10.2 Результаты CQA: балл

балл

CQA оценивает остаток каждой строки по сравнению со всеми остальными по шкале от 0 до 10:

- Высокий показатель CQA ($\sim$ 9–10) : продан значительно дороже прогнозируемой цены — вероятно, отличное состояние/качество/привлекательность
- Низкий показатель CQA ($\sim$ 0–1) : продан значительно дешевле прогнозируемой цены — вероятно, неудовлетворительное состояние или продажа в связи с финансовыми затруднениями
- Показатель CQA $\sim$ 5 : близок к медиане

В режимах «Оценка» и «Рынок» строки сортируются по параметру « residual_sf» в порядке убывания, если указан столбец « living_area».

11 Расчеты RCA& Скачивание

Рабочий процесс RCA (Метод остаточных ограничений) доступен в режимах «Оценка» и «Анализ рыночной зоны» после подгонки модели. Последующая таблица сравнительных продаж (глава 12) предназначена исключительно для оценки.

11.1 Открытие диалогового окна RCA

Нажмите кнопку «Рассчитать корректировки RCA» & «Скачать» в разделе 8 боковой панели. Выберите:

- Тип оценки CQA или «CQA на SF ». Если выбран вариант «CQA на SF», остаточный балл объекта оценки нормируется с учётом жилой площади.
- Введите балл CQA объекта оценки (0,00–9,99, по умолчанию 5,00)

11.2 Интерполяция баллов CQA

1 . Баллы CQA и остатки для аналогов отсортированы
2 . Линейная интерполяция сопоставляет значение CQA с остатком
3 . Значение для объекта оценки = прогноз модели + интерполированный остаток

11.3 Столбецы вывода

Колонка	Описание
subject_value »	Прогноз по модели + интерполированные остатки
<var>_adjustment	Вклад объекта оценки− Вклад «comp»
residual_adjustment	Остатки объекта оценки− Остатки «comp»
net_adjustments	Сумма всех корректировок
gross_adjustments	Сумма абсолютных корректировок
adjusted_sale_price	Цена продажи «comp» + чистые корректировки

После успешного выполнения вычислений на кнопке появляется белая галочка.

12 Таблица сравнения продаж

Таблица сравнения продаж доступна только в режиме оценки, после расчета корректировок RCA (шаг 8). Она формирует отформатированную книгу Excel, подходящую для включения в отчёты об оценке.

12.1 Диалоговое окно выбора аналогов

При нажатии кнопки «Сгенерировать таблицу продаж» & «Скачать» в боковой панели в разделе 9 открывается модальное диалоговое окно со списком всех аналогов из вывода анализа RCA. Аналоги разделены на две группы:

- Рекомендуемые аналоги (уже отмечены): Валовая корректировка$<$ 25% от цены продажи, отсортированные по дате продажи (сначала самые последние).
- Другие аналоги (по умолчанию не отмечены): Остальные аналоги, отсортированные по дате продажи.

Выберите до 30 аналогов, а затем нажмите «Сформировать таблицу продаж», чтобы создать книгу.

12.2 Макет книги

Сгенерированная книга содержит до 10 листов, на каждом из которых размещены 3 аналога рядом друг с другом:

- Строки заголовков : адреса объекта оценки и покупателя, цены продажи, даты продажи и количество дней на рынке
- Сгруппированные строки : местоположение (широта, долгота, площадь), размер участка (площадь участка, габариты участка) и возраст (фактический возраст, эффективный возраст) с заголовками групп и подпунктами
- Строки с переменными модели : по одной строке на каждый ненулевой коэффициент модели, показывающие вклад объекта оценки, вклад аналогов и корректировки
- Строки с остаточными характеристиками : пустые, незаблокированные ячейки для ввода оценщиком характеристик, не учтенных моделью (например, состояние, качество, вид)
- Сводные строки : чистая корректировка% , валовая корректировка% и скорректированная цена продажи с активными формулами Excel

12.3 Защита листа & Остатки ячейки

Каждый лист защищен паролем для предотвращения случайных изменений, однако ячейки с остаточными значениями характеристик явно разблокированы. Это позволяет оценщикам вводить значения для характеристик, не включенных в модель, при этом сохраняя расчеты корректировок, основанные на формулах.

13 Скачивание отчетов

13.1 «Двухэтапный кварто» Рабочий процесс

Экспорт отчёта из боковой панели представляет собой двухэтапный рабочий процесс в Quarto:

1 . «Генерация отчёта Quarto» (раздел 10 боковой панели) — создаёт самодостаточный набор Quarto (исходный файл .qmd , а также графики, report_data.rds и reference.docx) в папке вывода glmnet проекта.
2 . «Convert Quarto отчёт» (раздел 11 в боковой панели) — преобразует файл .qmd в выбранные вами форматы вывода. По умолчанию указывается путь к самому последнему созданному набору, но вы можете выбрать любой .qmd .

13.2 Форматы отчётов

На этапе «Преобразование» доступны три формата:

- HTML — Автономный HTML-документ. Работает на всех платформах без дополнительного программного обеспечения, кроме pandoc. Рекомендуемый вариант по умолчанию.
- Word (.docx) — Отформатированный документ. Подходит для редактирования и распространения. Работает на всех платформах.
- PDF — Сгенерирован с помощью Quarto на основе LaTeX. Размер бумаги соответствует настройкам локали (Letter или A4). Требует установки LaTeX (см. приложение «Системные требования»). Если LaTeX не обнаружен, опция PDF автоматически скрывается из списка флажков форматов.

Отчёты формируются с помощью Quarto, если это возможно, а в противном случае — с помощью rmarkdown. На вкладке «Отчёт» также доступны кнопки «Экспорт в Word» и «Экспорт в PDF».

13.3 Содержание отчёта

Отчёты включают всё содержимое вкладок:

- Описание набора данных (имя файла, наблюдения, зависимая переменная, предикторы, цель)
- Спецификация модели (альфа, лямбда, семейство, стандартизация, ослабленные ограничения, ненулевые коэффициенты)
- Сводка результатов (R^2 , скорректированное R^2 , обобщённое R^2 , CV R^2 , RMSE, MAE)
- Уравнение модели (с отображением в LaTeX)
- Таблица коэффициентов
- Важность переменных (гистограмма и таблица)
- Графики вклада по каждой переменной, соответствующие вкладке «Вклады»:
  - Числовые предикторы : диаграмма рассеяния с линией линейной подгонки
  - Факторные переменные : бокс-плот по уровням
  - Взаимодействия : диаграмма рассеяния с цветовой маркировкой по вкладу, тепловая карта среднего вклада и статический снимок 3D-поверхности (в перспективе)
- Матрица включённых взаимодействий (когда включено любое взаимодействие) — предикторы на обеих осях, с галочкой там, где был включён слагаемое x1:x2 , отображается в альбомной ориентации

- Тепловая карта матрицы корреляций
- Таблица декомпозиции ANOVA
- Диагностические графики:
  - Путь коэффициентов
  - Ошибка перекрёстной проверки
  - Фактические и прогнозируемые значения
  - Остатки и результаты подгонки
  - Нормальный график Q-Q
- Вывод модели

Во всех подписях к осям и легендах цветов используются числа с форматом с запятыми (без научной нотации).

После успешного создания отчёта на кнопке отчёта появляется белая галочка, а во время оформления отчёта отображается таймер, показывающий время, прошедшее с начала процесса.

14 Сравнение с earthUI

Приложение «Earth» и его модуль постобработки glmnet представляют собой единый набор инструментов для регрессионного моделирования. Они используют общий формат данных, специальные типы столбцов, рабочий процесс RCA и демонстрационные наборы данных, но построены на разных механизмах моделирования.

14.1 Основные различия

Функции	glmnetUI	earthUI
Методы	Эластичная сеть (glmnet)	MARS/Earth (earth)
Отбор переменных	Автоматический отбор с использованием штрафа Lasso	Автоматический отбор с использованием метода «forward/backward pruning»
Зависимости коэффициентов	Только линейные эффекты	Кусочно-линейные (функции-шарниры)
Взаимодействия	Попарные через матрицу взаимодействий	До 3-й степени
Знаки коэффициентов	Можно задать с помощью ограничений	Невозможно задать напрямую
Смягчённая подгонка	Доступна смягчённая Lasso-подгонка	Неприменимо
Настройка альфа-коэффициента	Доступен поиск по таблице	Неприменимо (нет параметра смешивания)
Таблица сравнения продаж	Доступно (шаг 9)	Доступно (шаг 8)
Графики g-функций	Не применимо	Визуализация сгруппированных слагаемых
Форматы отчетов	HTML, Word, PDF	HTML, Word, PDF

14.2 Общие функции

Оба инструмента предоставляют:

- Импорт из CSV и Excel с помощью библиотеки snake_ Преобразование регистра в столбцах
- Поддержка локалей 31 страны с учетом разделителя в CSV, знака десятичной запятой и формата бумаги
- Специальные обозначения столбцов (контракт_ дата, жилая_ площадь, площадь, широта, долгота и т. д.)
- Автоматический расчёт срока действия договора_ Расчёт срока действия договора с даты оценки
- Переключение между тёмным и светлым режимами
- Сохранение настроек с помощью localStorage
- Тепловая карта корреляций, значимость переменных, вклады, ANOVA, диагностика
- Рабочий процесс оценки CQA и корректировки RCA
- Таблица сравнения продаж с выбором аналогов для сравнения и отформатированным выводом в Excel
- Совместимость с теми же демонстрационными наборами данных (Оценка_ 1.csv)

14.3 Когда использовать какой

- Используйте glmnetUI , если вам нужна простая, интерпретируемая линейная модель с автоматическим отбором переменных. Модель эластичной сети особенно хорошо подходит в тех случаях, когда у вас много предикторов и вы хотите, чтобы модель сама выбрала наиболее значимые из них.
- Используйте earthUI , если вы ожидаете нелинейных зависимостей (например, влияние жилой площади на изменение цены при разных размерах) или вам нужны функции-шарниры для учета пороговых эффектов. Модели earth также могут учитывать взаимодействия высших порядков (2-й и 3-й степени). earthUI дополнительно предлагает экспериментальную функцию извлечения комментариев на основе Prolog/DCG, которой нет аналога в glmnetUI.
- Используйте обе модели для сравнения линейных и нелинейных моделей на одних и тех же данных. Если обе модели дают схожие значения R^2 , более простая модель glmnet может быть предпочтительнее с точки зрения интерпретируемости.

15 Демо-набор данных: Оценка _ 1.csv

15.1 Описание

Демонстрационный набор данных входит в состав glmnetUI (и используется совместно с earthUI). Найти его можно следующим образом:

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

Файл содержит 1 502 сделки по продаже жилой недвижимости (плюс 1 объект оценки, указанный в 1-й строке) из смоделированного экспорта данных MLS. Данные отражают продажи односемейных домов на рынке, охватывающем несколько районов, с различными размерами, возрастом и местоположением объектов недвижимости.

Это не реальные данные, но они основаны на реальном районе в Северной Калифорнии. Вся идентифицирующая информация была изменена или удалена.

15.2 Столбецы

Колонка	Тип	Специальный тип	Описание
weight	числовой	вес	Вес наблюдения (0 = исключить из подбора)
id	числовой	display_only	Внутренний идентификатор записи
property_id	числовой	display_only	Идентификатор объекта недвижимости в системе MLS
listing_id	символьный	display_only	Номер объявления в системе MLS
parcel_number	символьный	display_only	Номер участка по оценке округа (APN)
street_address	символьный	display_only	Адрес объекта недвижимости
city_name	символьный	display_only	Город
postal_code	символьный	display_only	Почтовый индекс
county_name	символьный	display_only	Округ
contract_date	Дата	contract_date	Дата заключения договора купли-продажи (рассчитывается sale_age)
sale_age	числовой	—	Количество дней от даты заключения договора до даты оценки
sale_price	числовой	(целевая переменная)	Цена продажи — переменная ответа
living_sqft	числовой	living_area	Общая жилая площадь в квадратных футах
beds_total	целое число	—	Количество спален
baths_total	числовой	—	Общее количество ванных комнат (например, 2,5 = 2 полные + 1 половина)
lot_size	числовой	lot_size	Площадь участка в квадратных футах
area_id	целое число	area	Идентификатор района в системе MLS
age	числовой	actual_age	Возраст объекта недвижимости в годах
latitude	числовой	latitude	Широта (округленная до 3 знаков после запятой для модели)
longitude	числовой	longitude	Долгота (округленная до 3 знаков после запятой для модели)
garage_spaces	целое число	—	Количество парковочных мест
days_on_market	целое число	dom	Количество дней на рынке
listing_date	Дата	listing_date	Дата размещения объявления
sale_concessions	числовой тип	concessions	Уступки продавца

15.3 Рекомендуемое руководство по быстрому запуску

1 . Запустите процедуру glmnet: earthUI::launch_glmnet()
2 . Создайте новый проект (раздел 1) с целью «For оценка»
3 . Поместите файл Appraisal_1.csv в папку вводных данных проекта и выберите его в разделе 2 (при необходимости нажмите «Refresh»)
4 . Выберите в качестве цели « sale_price
5 . Назначьте специальные типы, как показано в таблице выше
6 . Включите предикторы: sale_age , living_sqft , baths_total , lot_size , area_id (в качестве фактора), age , latitude , longitude , garage_spaces
7 . Оставьте альфа = 1 (Lasso), CV с lambda.1se
8 . Нажмите «Fit Glmnet Model» (Построить модель Glmnet)

9 . Просмотрите вкладки «Summary» (Сводка), «Коэффициенты» (Коэффициенты) и «Диагностика» (Диагностика)
10 . Скачайте вывод (Шаг 7), просмотрите рейтинг CQA
11 . Рассчитайте корректировки RCA (Шаг 8) с баллом CQA ~5,00
12 . Сгенерируйте таблицу сравнения продаж (Шаг 9) с рекомендуемыми аналогами

16 Системные требования& Устранение неполадок

16.1 Поддерживаемые платформы

glmnetUI работает на macOS, Windows и Linux. Настоятельно рекомендуется использовать RStudio Desktop (версия 2023.06 и выше) — в него входит набор pandoc, необходимый для создания отчётов.

16.2 Дополнительные зависимости

- LaTeX (TinyTeX, MiKTeX или MacTeX): только для создания отчетов в формате PDF. Если программа не обнаружена, опция PDF автоматически скрывается. Установите с помощью: tinytex::install_tinytex()
- sysfonts + showtext : шрифт Roboto Condensed для графиков. При недоступности или отсутствии подключения к сети автоматически используется системный шрифт sans-serif.
- earth ($\geq$ 5.3.0): Требуется только для конвейера импорта earthUI .

16.3 Примечания по платформе

macOS: наименьшее количество проблем. Установите TinyTeX для работы с PDF: tinytex::install_tinytex()

Windows: Хорошо работает с RStudio. На корпоративных/заблокированных компьютерах могут действовать ограничения на использование временных каталогов — приложение выдаст предупреждение в консоли, но продолжит работу без сохранения настроек.

Linux: Возможно, потребуются системные библиотеки: sudo apt install libcurl4-openssl-dev libssl-dev libxml2-dev libsqlite3-dev libfontconfig1-dev

16.4 Плавная деградация

Отсутствующий компонент	Поведение
Отсутствует LaTeX	Параметр PDF скрыт. HTML и Word по-прежнему доступны.
Отсутствует интернет / ошибка шрифтов	Используется системный шрифт без засечек. Сообщение в консоли зарегистрировано.
Отсутствует RSQLite / файловая система только для чтения	Настройки не сохраняются. Приложение работает нормально.
Временный каталог недоступен для записи	Сбой генерации отчёта с явной ошибкой. Укажите TMPDIR в качестве пути к каталогу, доступному для записи.

16.5 Устранение неполадок в программе «

»

«Опция PDF недоступна»: выполните команду tinytex::install_tinytex()в R, перезапустите приложение.

«Настройки не сохраняются»: проверьте права доступа к каталогу пользовательских данных (macOS: ~/Library/Application Support/R/glmnetUI/, Linux: ~/.local/share/R/glmnetUI/, Windows: %APPDATA%/R/data/glmnetUI/).

Порт 7879 уже занят : запустите команду ‘ lsof -ti:7879 | xargs kill ‘ (macOS/Linux) или ‘ Stop-Process -Id (Get-NetTCPConnection -LocalPort 7879).OwningProcess ‘ (Windows PowerShell).

Список литературы

Friedman, Jerome, Trevor Hastie и Robert Tibshirani (2010). «Regularization Paths for Generalized Linear Models via Coordinate Descent». В: Journal of Statistical Software 33.1, с. 1—22. DOI: 10.18637/jss.v033.i01.

Simon, Noah и др. (2011). «Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent». В: Journal of Statistical Software 39.5, с. 1—13. DOI: 10.18637/jss.v039.i05.

Tay, J. Kenneth, Balasubramanian Narasimhan и Trevor Hastie (2023). «Elastic Net Regularization Paths for All Generalized Linear Models». В: Journal of Statistical Software 106.1, с. 1—31. DOI: 10.18637/jss.v106.i01.