На правах рукописи

Стрижов Вадим Викторович

ПОРОЖДЕНИЕ И ВЫБОР МОДЕЛЕЙ В ЗАДАЧАХ РЕГРЕССИИ И КЛАССИФИКАЦИИ

05.13.17 — теоретические основы информатики

Автореферат диссертации на соискание учёной степени
доктора физико-математических наук

Москва — 2014

Работа выполнена в Федеральном государственном бюджетном учреждении науки Вычислительный центр им. А. А. Дородницына Российской академии наук.

Официальные оппоненты:

Двоенко Сергей Данилович, доктор физико-математических наук, профессор, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования Тульский государственный университет, профессор Кафедры автоматики и телемеханики,

Сметанин Юрий Геннадиевич, доктор физико-математических наук, Российский фонд фундаментальных исследований, начальник Отдела инфокоммуникационных технологий и вычислительных систем,

Хачай Михаил Юрьевич, доктор физико-математических наук, профессор, Федеральное государственное бюджетное учреждение науки Институт математики и механики им. Н. Н. Красовского УрО РАН, заведующий Отделом математического программирования.

Ведущая организация:

Федеральное государственное бюджетное учреждение науки Институт системного анализа Российской академии наук.

Защита состоится *«13» ноября 2014 г. в 13:00* на заседании диссертационного совета *Д 002.017.02* при Федеральном государственном бюджетном учреждении науки Вычислительный центр им. А. А. Дородницына Российской академии наук, расположенном по адресу: *119333, г. Москва, ул. Вавилова, д. 40.*

С диссертацией можно ознакомиться в библиотеке Федерального государственного бюджетного учреждения науки Вычислительный центр им. А. А. Дородницына Российской академии наук и на сайте http://www.ccas.ru.

Автореферат разослан « ____ » _____ 2014 г.

Учёный секретарь диссертационного совета Д 002.017.02,

д.ф.-м.н., профессор В. В. Рязанов

Общая характеристика работы

Актуальность темы исследования. Диссертационная работа посвящена проблемам выбора моделей в задачах регрессионного анализа и классификации. Предлагается подход, согласно которому выбор производится из индуктивно-порождаемого множества моделей. Анализируется распределение параметров моделей. На основании этого анализа выбирается модель оптимальной сложности.

Модель, описывающая исследуемое явление, может быть получена двумя путями: во-первых, методами математического моделирования, во-вторых, методами анализа данных и информационного моделирования. Первый тип моделей интерпретируем экспертами в контексте моделируемого явления [Краснощеков: 2000]. Второй тип моделей, не всегда интерпретируем, но более точно приближает данные [Bishop: 2006]. Совмещение достоинств обоих подходов, результатом которого является получение интерпретируемых и достаточно точных моделей, является актуальной задачей теоретической информатики.

Центральным объектом исследования является проблема построения адекватных моделей регрессии и классификации при решении задач прогнозирования. Проблема заключается в отыскании моделей оптимальной сложности, которые описывают измеряемые данные с заданной точностью. Дополнительным ограничением является интерпретируемость моделей экспертами той предметной области, для решения задач которой создается модель.

Цель исследования заключается в создании и обосновании методов выбора моделей из индуктивно порождаемого множества, а также в исследовании свойств алгоритмов выбора моделей. Задача выбора моделей из счетного последовательно порождаемого множества поставлена впервые. При постановке задачи использовался обширный материал о способах выбора моделей и выбора признаков из конечного множества, наработанный ранее в области машинного обучения. Эта задача является одной из центральных проблем машинного обучения и интеллектуального анализа данных.

Основной задачей исследования является разработка методов последовательного порождения моделей и оценки ковариационных матриц параметров

моделей с целью управления процедурой выбора моделей. Основной сложностью такой задачи является необходимость выбора из значительного числа регрессионных моделей, либо необходимость оценки параметров структурно сложной, так называемой «универсальной» модели.

Взаимосвязь задачи порождения и задачи выбора регрессионных моделей была освещена в начале 1980-х годов А. Г. Ивахненко. Согласно предложенному им методу группового учета аргументов [Ивахненко: 1981, Madala: 1994], модель оптимальной структуры может быть найдена путем последовательного порождения линейных моделей, в которых компоненты являются мономами полинома Колмогорова-Габора от набора независимых переменных. Критерий оптимальности структуры модели задается с помощью скользящего контроля.

В отличие от этого метода, метод символьной регрессии [Koza: 2005, Зелинка: 2008] рассматривает порождение произвольных нелинейных суперпозиций базовых функций. В последние годы тема анализа сложности моделей, получаемых с помощью этого метода, стала распространенным предметом исследований [Hazan: 2006, Владиславлева: 2009].

Первоначально принципы индуктивного порождения моделей были предложены в методе группового учета аргументов. Структура суперпозиций задавалась при этом внешними критериями качества модели. Впоследствии эти критерии были обоснованы в рамках гипотезы порождения данных с помощью связанного байесовского вывода. При последовательном порождении моделей необходимо оценивать информативность элементов суперпозиции. В рамках метода байесовской регрессии [Bishop: 2000] для этого предложено использовать функцию плотности распределения параметров модели. Эта функция является параметрической и ее параметры были названы гиперпараметрами [Bishop: 2006]. Было предложено использовать гиперпараметры моделей для оценки информативности элементов суперпозиции, что сделало анализ гиперпараметров одним из способов выбора моделей.

Для модификации суперпозиций нелинейных моделей был предложен метод оптимального прореживания [LeCun: 1990]. Согласно этому методу, элемент суперпозиции можно отсечь как неинформативный, если значение выпуклости

функции ошибки от параметров модели не превосходит относительный заданный порог.

Задача выбора модели является одной из самых актуальных в регрессионном анализе. В современной зарубежной литературе для ее решения используется принцип минимальной длины описания. Он предлагает использовать для описания данных наиболее простую и одновременно наиболее точную модель [Grunwald: 2005].

Задача сравнения моделей детально разработана [MacKay: 1994–2003]. Как альтернатива информационным критериям [Burnham: 2002, Lehmann: 2005] был предложен метод двухуровневого байесовского вывода. На первом уровне вывода настраиваются параметры моделей. На втором уровне настраиваются их гиперпараметры. Согласно этому методу, вероятность выбора более сложной модели ниже вероятности выбора простой модели при сравнимом значении функции ошибки на регрессионных остатках. Принципы байесовского подхода для выбора линейных моделей регрессии и классификации предложены авторами [Celeux: 2006, Massart: 2008, Fleury: 2006].

В то же время, в упомянутых публикациях и подходах остается открытым ряд важных проблем, решение которых определяет актуальность представляемой диссертации. Поэтому представляется целесообразным создать и развить теорию порождения и выбора регрессионных моделей. Она заключается в следующем. Множество моделей заданного класса индуктивно порождается набором параметрических базовых функций, заданных экспертами. Каждая модель является допустимой суперпозицией таких функций.

Интерпретируемость моделей обеспечена тем, что каждая из порождаемых моделей является суперпозицией базовых функций, заданных экспертами. Класс моделей задается правилами порождения суперпозиций. Точность моделей обеспечивается тем, что рассматривается достаточно большой набор моделей-претендентов, из которого выбирается оптимальная модель. Критерий оптимальности включает в себя понятия сложности и точности модели. При построении критерия учитывается гипотеза порождения данных — предположение о распределении регрессионных остатков.

Одновременно с оценкой параметров вычисляются и гиперпараметры (параметры распределения параметров) модели. На основе гиперпараметров оценивается информативность элементов суперпозиции и оптимизируется её структура. Оптимальные модели выбираются согласно критерию, заданному гипотезой порождения данных.

Таким образом, предложен новый подход к решению поставленной задачи. Множество моделей индуктивно порождается из набора базовых функций, заданных экспертами. Каждая модель является допустимой суперпозицией базовых функций. Одновременно с оценкой параметров моделей выполняется также и оценка гиперпараметров функции распределения параметров моделей. На основе этих параметров оценивается информативность элементов суперпозиции и принимается решение об оптимизации ее структуры. Оптимальные модели выбирается согласно критерию, заданному гипотезой порождения данных.

В связи с вышеизложенным, решение крупной задачи теории распознавания, в рамках которой будут предложены новые способы порождения и выбора моделей регрессии и классификации, является актуальной темой.

Цель диссертационной работы — создание нового математического подхода для решения задачи последовательного выбора регрессионных моделей. Цель работы находится в рамках направления «создание и исследование информационных моделей, моделей данных и знаний, методов машинного обучения и обнаружения новых знаний».

В частности, цель работы включает в себя:

1) создание и обоснование методов выбора индуктивно порождаемых моделей для решения задач регрессии и классификации,

2) исследование ограничений, накладываемых на структуру суперпозиции различными алгоритмами выбора моделей,

3) исследование структуры последовательно порождаемых суперпозиций и свойств параметров моделей.

Эти цели соответствуют направлению области исследования специальности 05.13.17 «разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных а также создание техники, которая предоставляет, во-первых, совокупность методов разработки математических моделей и, во-вторых, возможность интерпретации моделей в той прикладной области знаний, в рамках которой эти модели создаются» (пп. 5, 12).

На защиту выносятся следующие результаты.

1. *Формализованы и исследованы методы выбора моделей* для основных классов моделей: линейных, обобщенно-линейных и существенно нелинейных. Предложенные методы позволяют анализировать также информативность отдельных элементов суперпозиций.

2. *Предложен способ оценки информативности элементов суперпозиций* путем анализа пространства параметров моделей. Каждому элементу суперпозиции ставится в соответствие вектор параметров, который рассматривается как многомерная случайная величина. При заданной гипотезе порождения данных выполняется приближение эмпирического распределения параметров модельной параметрической функцией распределения.

3. *Предложены алгоритмы оптимизации параметров и гиперпараметров* — параметров функций распределения параметров моделей. Данная оценка является информативностью элемента суперпозиции.

4. *Исследованы ограничения, накладываемые на множество суперпозиций*, при которых порождаемые суперпозиции являются допустимыми, предложены методы порождения допустимых суперпозиций.

5. *Предложен метод последовательного порождения и выбора моделей.* Он заключается в том, что на каждом шаге анализируется информативность элементов порождаемых моделей, после чего модель модифицируется таким образом, чтобы доставить наибольшее увеличение значению критерия выбора модели на данном шаге.

6. *Предложен метод анализа ковариационной матрицы параметров нелинейных моделей.* Предложен критерий отыскания мультиколлинеарности для рассматриваемых моделей. Поставлена и решена оптимизационная задача последовательного исключения элементов модели. Полученное решение позволяет получать устойчивые модели.

Научная новизна. Выносимые на защиту результаты (1–6) являются новыми; также новыми являются следующие результаты, ранее опубликованные автором в рецензируемых журналах: 1) метод индуктивного порождения регрессионных моделей как суперпозиций гладких функций из заданного множества; 2) алгоритм выбора наиболее информативных элементов суперпозиции с помощью вектора гиперпараметров; 3) метод выбора опорного множества объектов как альтернатива процедурам регуляризации при построении интегральных индикаторов; 4) алгоритм согласования экспертных оценок в ранговых шкалах: используется линейная комбинация конусов экспертных оценок в пространстве интегральных индикаторов и в пространстве весов показателей, 5) алгоритм решения прямой и обратной задачи при нахождении оптимального управления.

Методика исследования: методы алгебраического подхода к решению задач распознавания; методы вычислительной линейной алгебры, многомерной статистики и теории машинного обучения; методы теории категорий. В рамках машинного обучения используются такие методы как связный байесовский вывод, метод минимальной длины описания, устойчивое оценивание параметров, аппроксимация Лапласа в пространстве параметров. Все эти методы являются новыми и активно обсуждаются в научных публикациях в течение последних лет.

Достоверность и обоснованность результатов подтверждена строгостью и корректностью математических высказываний и доказательств. Была выполнена экспериментальная проверка полученных результатов на задачах с модельными и реальными данными. Результаты исследований неоднократно обсуждались на российских и международных научных конференциях. Результаты

исследования опубликованы в рецензируемых научных изданиях из числа рекомендованных ВАК РФ.

Теоретическая значимость. Впервые связаны методы порождения и методы выбора моделей. При этом снята проблема оценки параметров и их ковариационных матриц моделей большой структурной сложности, так как для этой оценки параметров последующих моделей используются результаты анализа ранее порожденных моделей. Такой подход позволяет получать устойчивые оценки параметров в условиях большого числа мультикоррелирующих и шумовых признаков. Для выбора конкурирующих моделей используется байесовский подход, что позволяет получить модель оптимальной статистической сложности.

Практическая значимость. Работа носит преимущественно теоретический характер. Для иллюстрации возможных практических применений в последней главе работы приведены математические постановки и анализ прикладных задач, при решении которых были использованы полученные результаты.

Апробация работы. Основные результаты работы и отдельные её части докладывались на конференциях:

- международная конференция «Conference of the International Federation of Operational Research Societies», Барселона — 2014 г.;
- международная конференция «European Conference on Operational Research», Бонн — 2009 г.; Лиссабон — 2010 г.; Вильнюс — 2012 г.; Рим — 2013 г.;
- международная конференция «Operational Research: Mastering Complexity», Бонн — 2010 г.; Цюрих — 2011 г.;
- всероссийская конференция «Математические методы распознавания образов», Москва — 2003, 2005, 2007, 2009 гг.;

– международная конференция «Интеллектуализация обработки информации», Симферополь — 2006, 2008 гг.;

– международная конференция «Математика. Компьютер. Образование», Дубна — 2005, 2006, 2008, 2009 гг.;

– международная конференция «SIAM Conference on Computational Science and Engineering», Майами — 2009 г.;

– международный форум «Quo Vadis Energy in Times Of Climate Change», Загреб — 2009 г.;

– международная конференция «Citizens and Governance for Sustainable Development», Вильнюс — 2003, 2006 гг.

Личный вклад. Все результаты, выносимые на защиту, получены автором лично и не имеют пересечений с результатами его кандидатской диссертации.

Полный текст диссертации размещен на официальном сайте Федерального государственного бюджетного учреждения науки Вычислительный центр им. А. А. Дородницына Российской академии наук, `http://www.ccas.ru`.

Публикации. Результаты диссертации описаны в 28-ми статьях в журналах, рекомендованных ВАК.

Описания отдельных результатов работы включались в научные отчёты по проектам РФФИ 04-01-00103-а, 04-01-00401-а, 04-01-00401-а, 05-01-08030-офи, 07-01-00064-а, 07-01-12076-офи, 07-07-00181-а, 07-07-00372-а, 08-01-12022-офи, 10-07-00422-а, 10-07-00673-а, 12-07-13118-офи, 13-07-00709-а.

Структура и объем работы. Диссертация состоит из оглавления, введения, шести глав, разбитых на параграфы, заключения, списка основных обозначений, предметного указателя, списка иллюстраций (99 пунктов), списка таблиц (26 пунктов) и списка литературы из 404-х наименований. Основной текст занимает 320 страниц.

Содержание работы

Во введении обоснована актуальность диссертационной работы, поставлены цели и задачи исследования, аргументирована научная новизна, показана практическая значимость результатов, представлены выносимые выносимые на защиту научные положения.

1. Постановка задачи выбора моделей

Определение 1. *Регрессионная выборка* $\mathfrak{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^m$ *— множество* m *пар, состоящих из вектора* $\mathbf{x}_i = [x_{ij}]_{j=1}^n$ *значений* n *свободных переменных и соотвествующего этому вектору значения зависимой переменной* y_i.

Предполагается, что переменные принадлежат множеству действительных чисел, либо его подмножеству: $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n$ и $y \in \mathcal{Y} \subseteq \mathbb{R}^1$. Индекс i элемента выборки и индекс j свободной переменной рассматриваются как элементы конечных множеств $i \in \mathcal{I} = \{1, \ldots, m\}$ и $j \in \mathcal{J} = \{1, \ldots, n\}$.

Предполагается, что элементы выборки связаны соотношением $y_i = f(\mathbf{w}, \mathbf{x}_i) + \varepsilon(\mathbf{x}_i)$, которое аддитивно включает случайную величину $\varepsilon = \varepsilon(\mathbf{x})$. Предположение о том, что то зависимая переменная есть сумма значений модели и некоторой случайной величины, сохраняется и ниже.

Для нахождения функции регрессии f используются регрессионные модели.

Определение 2. *Регрессионная модель — параметрические семейство функций, отображение*

$$f : \mathcal{W} \times \mathcal{X} \to \mathcal{Y}$$

декартова произведения области допустимых значений $\mathcal{W}$ *параметров модели и области допустимых значений* $\mathcal{X}$ *свободных переменных в область значения* $\mathcal{Y}$ *зависимой переменной. Иначе, регрессионная модель есть поэлементное отображение*

$$f : (\mathbf{w}, \mathbf{x}) \mapsto y,$$

в котором вектор параметров $\mathbf{w} \in \mathcal{W}$*, свободная переменная* $\mathbf{x} \in \mathcal{X}$ *и зависимая переменная* $y \in \mathcal{Y}$.

Считаем вектор зависимых переменных $\mathbf{y}$ и вектор параметров $\mathbf{w}$ многомерными нормально распределенными случайными величинами с ковариационными матрицами $\mathbf{A}^{-1}$ и $\mathbf{B}^{-1}$ соответственно. Чтобы получить оценки гиперпараметров $\mathbf{A}, \mathbf{B}, \mathbf{w}$, введем ограничения на вид распределений $p(\mathfrak{D}|\mathbf{w}, \mathbf{B})$ и $p(\mathbf{w}|\mathbf{A})$.

Для нахождения *наиболее вероятных параметров* модели $f(\mathbf{w}, \mathbf{x})$ используем Байесовский вывод. При заданной модели f и заданных значениях $\mathbf{A}$ и $\mathbf{B}$ максимизируется выражение

$$p(\mathbf{w}|\mathfrak{D}, \mathbf{A}, \mathbf{B}, f) = \frac{p(\mathfrak{D}|\mathbf{w}, \mathbf{B}, f)p(\mathbf{w}|\mathbf{A}, f)}{p(\mathfrak{D}|\mathbf{A}, \mathbf{B}, f)}. \tag{1}$$

Элементы этого выражения и соответствующие им параметры:

$p(\mathbf{w}|\mathfrak{D}, \mathbf{A}, \mathbf{B}, f)$ — апостериорное распределение параметров,

$\mathbf{w}_{\text{MP}} = \arg\max p(\mathbf{w}|\mathfrak{D}, \mathbf{A}, \mathbf{B}, f)$ — наиболее вероятные параметры,

$\mathbf{w}_{\text{ML}} = \arg\max p(\mathfrak{D}|\mathbf{w}, \mathbf{B}, f)$ — наиболее правдоподобные параметры,

$p(\mathfrak{D}|\mathbf{w}, \mathbf{B}, f)$ — функция правдоподобия данных,

$p(\mathbf{w}|\mathbf{A}, f)$ — априорное распределение параметров,

$p(\mathfrak{D}|\mathbf{A}, \mathbf{B}, f)$ — функция правдоподобия модели f.

Записывая функцию ошибки $S = E_{\mathbf{w}} + E_{\mathfrak{D}}$ в виде

$$S(\mathbf{w}) = \frac{1}{2}(\mathbf{w} - \mathbf{w}_{\text{ML}})^{\intercal}\mathbf{A}(\mathbf{w} - \mathbf{w}_{\text{ML}}) + \frac{1}{2}(\mathbf{y} - \mathbf{f})^{\intercal}\mathbf{B}(\mathbf{y} - \mathbf{f}), \tag{2}$$

получаем вместо (1) выражение

$$p(\mathbf{w}|\mathfrak{D}, \mathbf{A}, \mathbf{B}, f) = \frac{\exp\bigl(-S(\mathbf{w})\bigr)}{Z_S},$$

где Z_S — нормирующий коэффициент.

Задача восстановления регрессии имеет несколько разных постановок, каждую их которых можно условно отнести к одному из следующих типов:

1) задачи оценки параметров модели,

2) задачи выбора признаков или объектов регрессионной выборки,

3) задачи выбора регрессионных моделей,

4) задачи проверки гипотезы порождения данных.

Предполагается, что функция ошибки $S(\mathbf{w})$ задана гипотезой порождения данных. При задании функции ошибки используется байесовский вывод. Предполагается, что зависимая переменная имеет распределение из экспоненциального семейства.

Задача 1. *Задана выборка* $\mathfrak{D} = \{(\mathbf{x}_i, y_i)\}$, $i \in \mathcal{I}$, *функция ошибки модели* S *и модель — параметрическое семейство функций* $f(\mathbf{w}, \mathbf{x})$. *Требуется найти такие параметры* $\mathbf{w}$ *модели, которые бы доставляли минимум функции ошибки*

$$\mathbf{w}^* = \arg\min_{\mathbf{w}\in\mathcal{W}} S(\mathbf{w}|\mathfrak{D}, f). \tag{3}$$

В выражении (3) справа от вертикальной черты указаны фиксированные значения переменных, что читается: «при заданной выборке $\mathfrak{D}$ и модели f», аналогично обозначению, принятому для записи условной вероятности. Далее предполагается, что запись $S(\mathbf{w})$ экивалентна записи $S(\mathbf{w}|\mathfrak{D}, f)$, если специально не оговорено иное.

При использовании скользящего контроля, критерии которого описаны в предыдущем разделе, задача выбора модели ставится следующим образом.

Задача 2. *Задана выборка* $\mathfrak{D} = \{(\mathbf{x}_i, y_i)\}$, $i \in \mathcal{I}$, *где множество векторов свободных переменных* $\{\mathbf{x} = [x_1, \ldots, x_j, \ldots, x_n]\}$, *проиндексировано* $j \in \mathcal{J} = \{1, \ldots, n\}$. *Задано разбиение множества индексов элементов выборки* $\mathcal{I} = \mathcal{L} \sqcup \mathcal{C}$. *Задана функция ошибки* S *и модель — параметрическое семейство функций* $f(\mathbf{w}, \mathbf{x}) = \mu(\mathbf{w}^\intercal \mathbf{x})$, *где* μ *— функция связи. Требуется найти такое подмножество индексов* $\mathcal{A} \subseteq \mathcal{J}$, *которое бы доставляло минимум функции:*

$$\mathcal{A}^* = \arg\min_{\mathcal{A}\subseteq\mathcal{J}} S(f_\mathcal{A}|\hat{\mathbf{w}}, \mathfrak{D}_\mathcal{C}) \tag{4}$$

на подмножестве $\mathfrak{D}_\mathcal{C}$ *разбиении выборки* $\mathfrak{D}$, *определенном множеством индексов* $\mathcal{C}$. *При этом параметры* $\hat{\mathbf{w}}$ *модели должны доставлять минимум функции:*

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} S(\mathbf{w}|\mathfrak{D}_\mathcal{L}, f_\mathcal{A}) \tag{5}$$

на подмножестве $\mathfrak{D}_\mathcal{L}$ *разбиении выборки, определенном множеством индексов* $\mathcal{L}$. *Здесь* $f_\mathcal{A}$ *обозначает обобщенно-линейную модель* $f = \mu(\mathbf{w}_\mathcal{A}^\intercal \mathbf{x}_\mathcal{A})$, *включающую только столбцы матрицы* $\mathbf{X}$ *с индексами из множества* $\mathcal{A}$.

Нелинейная модель не может быть однозначно задана множеством $\mathcal{A}$ активных признаков. Поэтому для задания модели используются правила индуктивного порождения моделей детально определенные в следующем разделе. Они позволяют однозначно индексировать модели f из множества моделей $\mathcal{F}$.

Задача 3. *Задана выборка* $\mathfrak{D} = \{(\mathbf{x}_i, y_i)\}$, $i \in \mathcal{I} = \mathcal{L} \sqcup \mathcal{C}$. *Задано множество порождающих функций* $\mathcal{G} = \{g_1, \ldots, g_n\}$. *Заданы правила индуктивного порождения множества моделей* $\mathcal{F} = \{f_r\}$, *индексированных счетным множеством* $\mathcal{R} \ni r$. *Требуется найти такую модель* $f_{\hat{r}}$, *которая бы доставляла минимум функции*

$$\hat{r} = \arg\min_{r \in \mathcal{R}} S(f_r | \hat{\mathbf{w}}, \mathfrak{D}_{\mathcal{C}}) \tag{6}$$

при условии оценки оптимальных параметров $\hat{\mathbf{w}}$ *решением задачи* (5).

Оценка ковариационных матриц зависимой переменной и параметров.

Задача 4. *Задана выборка* $\mathfrak{D}$, *функция ошибки* $S(\mathbf{w})$ *и модель* $f(\mathbf{w}, \mathbf{x})$. *Требуется оценить обратные ковариационные матрицы* $\mathbf{A}, \mathbf{B}$ *максимизируя правдоподобие модели:*

$$(\hat{\mathbf{A}}, \hat{\mathbf{B}}) = \arg\max_{\mathbf{A} \in \mathbb{R}^{n^2}, \mathbf{B} \in \mathbb{R}^{m^2}} \int_{\mathbf{w} \in \mathbb{W}} p(\mathfrak{D} | \mathbf{w}, \mathbf{B}, f) p(\mathbf{w} | \mathbf{A}, f) d\mathbf{w}.$$

Обобщим предыдущую задачу на случай существенно нелинейных моделей. При этом будем считать, что $\mathcal{I} = \mathcal{B}$. Тогда задача выбора правдоподобной модели f_r с индексом r из множества моделей-претендентов $\mathcal{F}$ имеет следующий вид.

Задача 5. *Задана выборка* $\mathfrak{D} = \{(\mathbf{x}_i, y_i)\}$, $i \in \mathcal{I}$. *Задано множество моделей* $\mathcal{F} = \{f_r\}$, *индексированных счетным множеством* $\mathcal{R} \ni r$. *Требуется выбрать наиболее правдоподобную модель*

$$\hat{r} = \arg\max_{r \in \mathcal{R}} p(f_r | \mathfrak{D}) = \arg\max_{r \in \mathcal{R}} \int_{\mathbf{w} \in \mathcal{W}} p(\mathfrak{D} | \mathbf{w}, \hat{\mathbf{B}}, f_r) p(\mathbf{w} | \mathfrak{D}, \hat{\mathbf{A}}, f_r) d\mathbf{w}$$

Заметим, что оценивать параметры модели для того, чтобы выбрать наиболее правдоподобную модель, необязательно.

2. Порождение моделей

Построение моделей выполняется по итерационной схеме «порождение-выбор» в соответствии с определенными правилами порождения моделей и критерием выбора моделей. Последовательно порождаются наборы конкурирующих моделей. Каждая модель в наборе является суперпозицией элементов заданного множества гладких параметрических функций. После построения модели каждому элементу суперпозиции ставится в соответствие гиперпараметр.

Параметры и гиперпараметры модели последовательно настраиваются. Из набора выбираются наилучшие модели для последующей модификации. При модификации моделей, по значениям гиперпараметров делаются выводы о целесообразности включения того или иного элемента в модель следующего порождаемого набора.

Рассмотрим две функции $g : \mathbb{X} \to \mathbb{Y}$ и $h : \mathbb{Y}' \supseteq \mathbb{Y} \to \mathbb{Z}$ и пусть $\mathbb{Y}' \cup \mathbb{Y} \neq \emptyset$. Их композицией называется функция $f = g \circ h : \mathbb{X} \to \mathbb{Z}$, определенная равенством

$$(h \circ g)(\mathbf{x}) = h(g)(\mathbf{x}),\ \mathbf{x} \in \mathbb{X}.$$

Пусть задано множество $G = \{g_i\}$ функций. Для каждой функции g_i задана область определения $\mathbb{X}_i = \mathrm{dom}(g_i)$ и область значения $\mathbb{Y}_i = \mathrm{cod}(g_i)$. Пусть множество значений $\mathbb{Y}_i$ функции g_i содержится в области определения $\mathbb{X}_{i+1}$ функции g_{i+1}, то есть

$$g_i : \mathbb{X}_i \to \mathbb{Y}_i \subseteq \mathbb{X}_{i+1}, \quad i = 1, 2, \ldots, K-1, \tag{7}$$

то функция

$$f = g_K \circ g_{K-1} \circ \cdots \circ g_1, \quad K \geq 2, \tag{8}$$

х определяемая равенством

$$f(\mathbf{x}) = (g_K \circ g_{K-1} \circ \cdots \circ g_1)(\mathbf{x}) = g_K\big(g_{K-1}\big(\ldots(g_1)\big)\big)(\mathbf{x}), \quad \mathbf{x} \in \mathbb{X}, \tag{9}$$

называется *сложной функцией* или *суперпозицией функций* $g_1, g_2, \ldots, g_K$.

Определение 3. *Суперпозиция f функций $\{g_1, \ldots, g_K\}$ — функция, представленная как композиция нескольких функций, определяемая выражениями (8–9) при выполнении условия* (7).

Функции $g = g(\mathbf{b}, \xi)$ с параметрами $\mathbf{b}$ и аргументом ξ, принадлежащие множеству G, далее будут называться *порождающими функциями*.

Определение 4. *Допустимой суперпозицией f называется такая суперпозиция, в которой*

$$cod(g_{i(k+1)}) \subseteq dom(g_{i(k)}) \quad \textit{для всех} \quad k = 1, \ldots, K-1.$$

Для обобщения этого определения случай функций нескольких аргументов будем считать, что функции $g_1, \ldots, g_K$, входящие в суперпозицию, являются вектор-функциями от векторных величин $\boldsymbol{\xi}$. При этом и области определения $\mathbb{X}_i$, и области значений $\mathbb{Y}_i$ этих вектор-функций являются подмножествами декартова произведения пространств соответствующих аргументов.

Пусть задано множество порождающих функций $G = \{g_1, \ldots, g_l | g = g(\mathbf{b}, \cdot)\}$, то есть, заданы

1) сама функция $g : (\mathbf{b}, \xi) \mapsto \xi'$,

2) число ее параметров $\mathbf{b}$ (возможен пустой набор),

3) число аргументов (арность) $v(g)$ функции g (возможен пустой набор) и порядок следования аргументов,

4) домен $\mathrm{dom}(g)$ и кодомен $\mathrm{cod}(g)$.

Требуется построить функцию f как суперпозицию порождающих функций из заданного множества G. Модель $f(\mathbf{w}, \mathbf{x})$ рассматривается как суперпозиция

$$f(\mathbf{w}, \mathbf{x}) = (g_{i(1)} \circ \cdots \circ g_{i(K)})(\mathbf{x}), \quad \text{где} \quad \mathbf{w} = [\mathbf{b}_{i(1)}^\intercal, \ldots, \mathbf{b}_{i(K)}^\intercal]^\intercal,$$

в которой вектор $\mathbf{w}$ состоит из присоединенных векторов-параметров $\mathbf{b}$ функций g, водящих в суперпозицию f.

Для порождения моделей требуется задать:

1) множество непорождаемых переменных $\{\xi\}$ с заданным $\mathrm{dom}(\xi)$,

2) множество порождающих функций $G = \{g_u, \mathrm{id}\}$, $g : x \mapsto x'$,

3) правило Gen порождения допустимых суперпозиций $\mathcal{G} \supset G$, где суперпозиция $g_j = g_u \circ g_v \in \mathcal{G}$, построена с учетом ограничений

   на число аргументов $v(g_u)$,

   на область определения $\mathrm{cod}(g_u)$,

   на структурную сложность суперпозиции $C(g_j) \leqslant C_{\max}$,

на число и типы входных и выходных переменных,

4) правило Rem упрощения суперпозиций: $g_k \notin \mathcal{G}$, если

$$\text{Rem} : g_k = g_u \circ g_v \mapsto g_j \in \mathcal{G}.$$

Результатом порождения допустимых суперпозиций является набор $\mathcal{F}$ моделей-претендентов f, из которого производится выбор.

Поставим в во взаимно-однозначное соответствие каждой суперпозиции f дерево Γ_f, которое строится следующим образом:

1) в вершинах V_i дерева Γ_f находятся соответствующие порождающие функции $g_s, s = s(i)$;

2) число дочерних вершин у некоторой вершины V_i равно арности соответствующей функции g_s;

3) порядок вершин, дочерних для V_i, соотвествует порядку аргументов соответствующей функции $g_{s(i)}$;

4) в листьях дерева Γ_f находятся свободные переменные x_i.

Вычисление значения выражения $f = f(\mathbf{w}, \mathbf{x})$ в некоторой точке $\mathbf{x}$ с данным вектором параметров $\mathbf{w} = \{w_1, w_2, \ldots, w_\eta\}$ эквивалентно подстановке соответствующих значений свободных переменных $\mathbf{x}$ и параметров $\mathbf{w}$ функцию f, соответствующую дереву Γ_f.

Каждое поддерево Γ_f^i дерева Γ_f, соответствующее вершине V_i, также соответствует некоторой суперпозиции, являющейся составляющей исходной суперпозиции f.

3. Сравнение элементов моделей

Решается задача последовательного добавления элементов в регрессионную модель. Вводится критерии остановки процедуры порождения и выбора моделей. Вводится понятие расстояния между последовательно порождаемыми моделями. Результатом работы алгоритма является модель удовлетворительной точности; мультикоррелирующие признаки исключены. Процедура выбора

оптимального набора признаков состоит из этапов добавления и удаления. На первом этапе последовательно добавляются признаки, доставляющие минимум правдоподобию модели на обучающей выборке, заданной множеством индексов $\mathcal{L}$. На втором этапе происходит последовательное удаление признаков, согласно модифицированному методу Белсли. Пусть на k-ом шаге алгоритма имеется активный набор признаков $\mathcal{A}_k \in \mathcal{J}$. На нулевом шаге $\mathcal{A}_0$ пуст.

Этап добавления. Находим признак доставляющий минимум S на обучающей выборке

$$j^* = \arg \min_{j \in \mathcal{J} \setminus \mathcal{A}_{k-1}} S(\mathbf{w} | \mathfrak{D}_{\mathcal{L}}, f_{\mathcal{A}_{k-1} \cup \{j\}}).$$

Затем добавляем новый признак j^* к текущему активному набору

$$\mathcal{A}_k = \mathcal{A}_{k-1} \cup \{j^*\}$$

и повторяем эту процедуру до тех пор, пока $S(f_{\mathcal{A}_k} | \mathbf{w}^*, \mathfrak{D})$ превосходит свое минимальное значение на данном этапе не более, чем на некоторое заданное значение ΔS_1.

Этап удаления. Находим индексы обусловленности и долевые коэффициенты для текущего набора признаков $\mathcal{A}_{k-1}$ согласно методу Белсли, описание которого приведено ниже. Далее находим количество достаточно больших индексов обусловленности. Достаточно большими будем считать индексы квадрат которых превосходит максимальный индекс обусловленности η_t, где $t = |A_{k-1}|$ количество признаков в текущем наборе $\mathcal{A}_{k-1}$.

$$i^* = \sum_{g=1}^{t} \left[\eta_g^2 > \eta_t \right]. \tag{10}$$

Затем ищем в матрице долевых коэффициентов $\mathbf{var}(\mathbf{w})$ столбец j^* с максимальной суммой по последним i^* долевым коэффициентам

$$j^* = \arg \max_{j \in \mathcal{A}_{k-1}} \sum_{g=t-i^*+1}^{t} q_g^j. \tag{11}$$

Удаляем j^*-ый признак из текущего набора

$$\mathcal{A}_k = \mathcal{A}_{k-1} \setminus j^*$$

и повторяем эту процедуру до тех пор, пока $S(f_{\mathcal{A}_k}|\mathbf{w}^*, \mathfrak{D})$ превосходит свое минимальное значение на данном этапе не более, чем на некоторое заданное значение ΔS_2. Повторение этапов добавления и удаления осуществляется до тех пор, пока значение $S(f_{\mathcal{A}_k}|\mathbf{w}^*, \mathfrak{D})$ не стабилизируется.

Опишем критерий удаления элементов. Для ковариационной матрицы $\mathbf{A}^{-1}$ линейной или линеаризованной модеди справедливо выражение

$$\mathbf{A} = \sigma^2\mathbf{B}\mathbf{B}^\intercal = \sigma^2(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{X}((\mathbf{X}^\intercal\mathbf{X})^{-1})^\intercal =$$

$$\sigma^2\mathbf{X}^{-1}(\mathbf{X}^\intercal)^{-1}\mathbf{X}^\intercal\mathbf{X}((\mathbf{X}^\intercal\mathbf{X})^{-1})^\intercal == \sigma^2((\mathbf{X}^\intercal\mathbf{X})^{-1})^\intercal = \sigma^2(\mathbf{X}^\intercal\mathbf{X})^{-1}.$$

Выражение $\sigma^2(\mathbf{X}^\intercal\mathbf{X})^{-1}$ является несмещенной оценкой ковариационной матрицы признаков, а в случае линейной модели оно в точности совпадает с ковариационной матрицей, то есть $\mathbf{A}^{-1} = \sigma^{-2}\mathbf{X}^\intercal\mathbf{X}$.

Используя сингулярное разложение, дисперсия параметров, найденных методом наименьших квадратов $\mathbf{w} = (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$, может быть записана как

$$\mathbf{var}(\mathbf{w}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} = \sigma^2(\mathbf{V}^\intercal)^{-1}\mathbf{\Lambda}^{-2}\mathbf{V}^{-1} = \sigma^2\mathbf{V}\mathbf{\Lambda}^{-2}\mathbf{V}^T.$$

Таким образом, дисперсия j-го регрессионного коэффициента — это j-й диагональный элемент матрицы $\mathbf{var}(\mathbf{w})$.

Каждому индексу обусловленности η_j соответствуют значения q_{ij} — долевые коэффициенты. Сумма долевых коэффициентов по индексу j равна единице.

$$\sigma^{-2}\mathbf{var}(w_i) = \sum_{j=1}^{n}\frac{\upsilon_{ij}^2}{\lambda_j^2} = (q_{i1} + q_{i2} + \ldots + q_{in})\sum_{j=1}^{n}\frac{\upsilon_{ij}^2}{\lambda_j^2},$$

где q_{ij} — отношение соответствующего слагаемого в разложении вектора $\sigma^{-2}\mathbf{var}(w_i)$ ко всей сумме, а $\mathbf{V} = (\upsilon_{ij})$.

Большие величины η_j означают, что, возможно, есть зависимость между признаками. Признак считается вовлеченным в зависимость, если его долевой коэффициент связанный с этим индексом превышает выбранный порог.

Результаты сравнения предложенного и базовых алгоритмов приведены в табл. 1. Сравнение выполнялось на задаче поиска модели волатильности опционов. В таблицу входят значения функционала качества на обучающей и контрольной выборке, информационный критерий Акаике, число переменных

модели. Исходя из значений критериев делается вывод об эффективности работы алгоритмов.

Таблица 1. Результаты работы методов выбора моделей.

Алгоритм	$S_{\mathcal{L}}$	$S_{\mathcal{C}}$	AIC	BIC	C_p	$\lg \kappa$	k
Генет.	0,073	0,107	-1152	-1072	337	13	26
МГУА	0,146	0,194	-1076	-1045	745	6	10
Шаг. рег.	0,128	0,154	-1092	-1055	644	7	12
Гребн.	0,111	0,146	-819	-330	832	33	160
Лассо	0,121	0,147	-1089	-1034	611	5	18
Ступ.	0,071	0,096	-1157	-1077	324	9	26
FOS	0,106	0,135	-1105	-1044	527	7	20
LARS	0,098	0,095	-1102	-1017	492	7	28
Предл.	0,097	0,123	-1118	-1054	469	5	21

Для каждого алгоритма вычислены значения ошибок $S_{\mathcal{L}}$ и $S_{\mathcal{C}}$ на обучении и контроле,значение информационных критериев Акаике и Байеса

$$\text{AIC} = m\left(\ln\frac{S}{m}\right) + 2k, \qquad \text{BIC} = m\left(\ln\frac{S}{m}\right) + k\ln m,$$

Маллоуза C_p, десятичный логарифм числа обусловленности κ матрицы значений отобранных признаков и сложность модели k.

4. Выбор моделей

Предложен ряд методов оптимизации структурных параметров регрессионной модели. Описан метод аппроксимации Лапласа функции ошибки для оценки правдоподобия модели. Предложен метод Монте-Карло оценки правдоподобия модели. Предложен метод оценки оптимальных параметров модели с помощью процедуры скользящего контроля. Исследованы свойства предлагаемых методов. Проведен вычислительный эксперимент на модельных и реальных данных. Проведены анализ и сравнение предлагаемых методов.

Согласно принятым в работе гипотезам, функция правдоподобия данных имеет вид

$$p(\mathfrak{D}|\mathbf{w},\beta) = \frac{\exp(-E_{\mathfrak{D}})}{Z_{\mathfrak{D}}(\mathbf{B})} = \frac{\exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{f})^{\intercal}\mathbf{B}(\mathbf{y}-\mathbf{f})\right)}{(2\pi)^{\frac{m}{2}}\det^{\frac{1}{2}}(\mathbf{B})}, \tag{12}$$

функция априорного распределения параметров, при предположении о том, что оценка матожидания вектора параметров равна $\mathbf{w}_0$ имеет вид

$$p(\mathbf{w}|\mathbf{A}) = \frac{\exp(-E_{\mathbf{w}})}{Z_{\mathbf{w}}(\mathbf{A})} = \frac{\exp\left(-\frac{1}{2}(\mathbf{w}-\mathbf{w}_0)^{\intercal}\mathbf{A}(\mathbf{w}-\mathbf{w}_0)\right)}{(2\pi)^{\frac{n}{2}}\det^{\frac{1}{2}}(\mathbf{B})}, \tag{13}$$

а функция апостериорного распределения параметров —

$$p(\mathbf{w}|\mathfrak{D},\mathbf{A},\mathbf{B}) = \frac{p(\mathfrak{D}|\mathbf{w},\mathbf{B})p(\mathbf{w}|\mathbf{A})}{p(\mathfrak{D}|\mathbf{A},\mathbf{B})} = \frac{\exp\big(-S(\mathbf{w}|\mathfrak{D},\mathbf{A},\mathbf{B})\big)}{Z_S}. \tag{14}$$

Теорема 1. *Правдоподобие линейной модели для гипотезы порождения данных имеет вид*

$$p(\mathfrak{D}|\mathbf{A},\mathbf{B}) = \frac{|\mathbf{B}|^{\frac{1}{2}}|\mathbf{A}|^{\frac{1}{2}}}{(2\pi)^{\frac{m}{2}}|\mathbf{K}|^{\frac{1}{2}}}\exp\left(\frac{1}{2}\mathbf{y}^{\intercal}(\mathbf{C}^{\intercal}\mathbf{K}\mathbf{C}-\mathbf{B})\mathbf{y}\right),$$

а его логарифм имеет вид $\ln p(\mathfrak{D}|\mathbf{A},\mathbf{B}) =$

$$= -\frac{1}{2}\big(\ln|\mathbf{K}| + m\ln 2\pi - \ln|\mathbf{B}| - \ln|\mathbf{A}| - \mathbf{y}^{\intercal}(\mathbf{C}^{\intercal}\mathbf{K}\mathbf{C}-\mathbf{B})\mathbf{y}\big).$$

Здесь

$$\mathbf{K} = \mathbf{X}^{\intercal}\mathbf{B}\mathbf{X} + \mathbf{A}, \quad \mathbf{C} = \mathbf{K}^{-1}\mathbf{X}^{\intercal}\mathbf{B}.$$

Зафиксируем значение вектора $\mathbf{w}_0$, предполагая что он доставляет локальный максимум (14). Для нахождения матриц $\mathbf{A},\mathbf{B}$ приблизим фукнцию ошибки $S(\mathbf{w}|\mathfrak{D},\mathbf{A},\mathbf{B})$ методом Лапласа. Для этого построим ряд Тейлора второго порядка логарифма числителя (14) в окрестности $\mathbf{w}_0$:

$$\ln\exp\big(-S(\mathbf{w})\big) = \ln\exp\left(S(\mathbf{w}_0) + 0 + \frac{1}{2}\Delta\mathbf{w}^{\intercal}\mathbf{H}\Delta\mathbf{w} + o(||\mathbf{w}||^3)\right),$$

где $\Delta\mathbf{w} = \mathbf{w} - \mathbf{w}_0$. При упрощении и отбрасывании малой величины получим

$$-S(\mathbf{w}) \approx -S(\mathbf{w}_0) - \frac{1}{2}\Delta\mathbf{w}^{\intercal}\mathbf{H}\Delta\mathbf{w}. \tag{15}$$

В выражении (15) нет слагаемого первого порядка, так как предполагается, что $\mathbf{w}_0$ доставляет локальный минимум функции ошибки

$$\left.\frac{\partial S(\mathbf{w})}{\partial\mathbf{w}}\right|_{\mathbf{w}=\mathbf{w}_0} = \mathbf{0}.$$

Матрица $\mathbf{H}$ — матрица Гессе функции ошибок

$$\mathbf{H} = -\nabla\nabla S(\mathbf{w})\big|_{\mathbf{w}=\mathbf{w}_0}. \tag{16}$$

Предложен ряд процедур оценивания параметров и гиперпараметров. Для оценки структурных параметров необходимо провести процедуру максимизации правдоподобия модели. Именно эта процедура является наиболее вычислительно затратной. Оптимальные структурные параметры $\mathbf{A}, \mathbf{B}$ максимизируют правдоподобие модели

$$p\left(\mathfrak{D}|\mathbf{A},\mathbf{B}\right) = \int\limits_{\mathbf{w}\in\mathbb{W}} p(\mathfrak{D}|\mathbf{w},\mathbf{B})p(\mathbf{w}|\mathbf{A})d\mathbf{w} \to \max_{\mathbf{A}\in\mathbb{M}^n,\mathbf{B}\in\mathbb{M}^m}, \tag{17}$$

где $\mathbb{M}^n$ обозначает множество положительно полуопределенных матриц размерности $n \times n$.

Теорема 2. *Для гипотезы нормального распределения зависимой переменной вариант: биномиального при фиксированных ковариационных матриц $\mathbf{A}^{-1}, \mathbf{B}^{-1}$ итерационный алгоритм оценки параметров обобщенно-линейной модели*

$$\Delta\mathbf{w}_{k+1} = \left(\mathbf{X}^\intercal\mathbf{B}\mathbf{X} + \mathbf{A}\right)^{-1}\mathbf{X}^\intercal\mathbf{B}^\intercal\mathbf{y} - \mathbf{w}_k, \quad \textit{вариант:}$$

$$\Delta\mathbf{w}_{k+1} = (\mathbf{X}^\intercal\mathbf{B}\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{B}\big(\mathbf{X}\mathbf{w}_k - \mathbf{B}^{-1}(\mathbf{f}-\mathbf{y})\big) + \frac{1}{2}\mathbf{w}_k^\intercal\mathbf{A}\mathbf{w}_k$$

доставляет локальный минимум функции ошибки общего вида $S(\mathbf{w}|\mathfrak{D},\mathbf{A},\mathbf{B},\mathbf{f})$ при сходимости последовательности векторов $\mathbf{w}_k$.

Оптимальные значения гиперпараметров α и β — элементов матриц $\mathbf{A}$ и $\mathbf{B}$ вычисляются итеративно следующим образом. При фиксированных параметрах $\mathbf{w}_0$ находятся оптимальные значения α. С использованием α находятся оптимальные значения β. Далее новые β определяют новые значения вспомогательной переменной λ. Цикл повторяется до тех пор, пока изменение значений α, β на соседних шагах не станет менее заранее заданной границы.

Оптимальные значения параметров $\mathbf{w}$ переоцениваются с использованием функции ошибки $S(\mathbf{w}|\mathfrak{D},\mathbf{A},\mathbf{B})$, определенной числителем (14) при фиксированных на данном шаге значениях матриц $\mathbf{A}, \mathbf{B}$. Таким образом, параметры $\mathbf{w}$ и гиперпараметры $\mathbf{A}, \mathbf{B}$ регрессионной модели f оцениваются по отдельности.

На каждой итерации сначала при фиксированных значениях гиперпараметров отыскиваются параметры путем оптимизации функции $S(\mathbf{w})$. При этом используется алгоритм Левенберга-Марквардта или его модификации, описанные в первом разделе.

Для того, чтобы оценить структурные параметры $\mathbf{A}, \mathbf{B}$ совместно с параметрами $\mathbf{w}$ регрессионной модели, воспользуемся методом аппроксимации Лапласа функции правдоподобия модели.

Для нахождения оптимальных структурных параметров $\hat{\mathbf{A}}, \hat{\mathbf{B}}$ выражение (17) преобразуется следующим образом:

$$\frac{|\mathbf{B}|^{\frac{1}{2}}}{(2\pi^{\frac{m}{2}})}\frac{|\mathbf{A}|^{\frac{1}{2}}}{(2\pi^{\frac{n}{2}})}\int\limits_{\mathbf{w}\in\mathbb{W}} \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{f})^{\intercal}\mathbf{B}(\mathbf{y}-\mathbf{f})\right)\exp\left(-\frac{1}{2}\mathbf{w}^{\intercal}\mathbf{A}\mathbf{w}\right)d\mathbf{w}. \qquad (18)$$

Примем за функцию ошибки $S(\mathbf{w}, \mathbf{A}, \mathbf{B})$ показатель экспоненты выражения (18) с отрицательным знаком:

$$S(\mathbf{w}, \mathbf{A}, \mathbf{B}) = \frac{1}{2}(\mathbf{y}-\mathbf{f})^{\intercal}\mathbf{B}(\mathbf{y}-\mathbf{f}) + \frac{1}{2}\mathbf{w}^{\intercal}\mathbf{A}\mathbf{w}, \qquad (19)$$

тогда оптимизационная задача (18) перепишется в более удобном виде:

$$\frac{|\mathbf{B}|^{\frac{1}{2}}}{(2\pi^{\frac{m}{2}})}\frac{|\mathbf{A}|^{\frac{1}{2}}}{(2\pi^{\frac{n}{2}})}\int\limits_{\mathbf{w}\in\mathbb{W}} \exp\big(-S\left(\mathbf{w}, \mathbf{A}, \mathbf{B}\right)\big)d\mathbf{w} \to \max_{\mathbf{A},\mathbf{B}}.$$

Отметим, что оптимальными параметрами $\hat{\mathbf{w}}$ модели $\mathbf{f} = \mathbf{f}(\mathbf{w}, \mathbf{X})$ являются те, которые максимизируют апостериорное распределение параметров или, в нашем случае, минимизируют функцию ошибки

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathbb{W}} S(\mathbf{w}|\hat{\mathbf{A}}, \hat{\mathbf{B}}),$$

где $\hat{\mathbf{A}}, \hat{\mathbf{B}}$ — оптимальные структурные параметры, максимизирующие выражение (18).

Теорема 3. *Для гипотезы порождения данных* (**??**) *при фиксированных ковариационных матриц* $\mathbf{A}^{-1}, \mathbf{B}^{-1}$ *итерационный алгоритм оценки параметров*

$$\Delta\mathbf{w}_{k+1} = (\mathbf{J}^{\intercal}\mathbf{J})^{-1}\left(\mathbf{J}^{\intercal}\big(\mathbf{y}-\mathbf{f}(\mathbf{w}, \mathbf{X})\big) - \frac{1}{\beta}\mathbf{A}^{-1}\mathbf{w}_k\right)$$

доставляет локальный минимум функции ошибки общего вида $S(\mathbf{w}|\mathfrak{D}, \mathbf{A}, \mathbf{B}, \mathbf{f})$ *при сходимости последовательности векторов* $\mathbf{w}_k$.

Замечание. Итерационный алгоритм $\mathbf{w}_{k+1} = \Delta\mathbf{w}_{k+1} + \mathbf{w}_k$ предполагает известное начальное приближение $\mathbf{w}_0$. Последовательность $\|\mathbf{w}_{k+1} - \mathbf{w}_k\|^2$ монотонно убывает с увеличением номера шага k.

Вместо оптимизации выражения (18), будем оптимизировать аппроксимированное выражение

$$\frac{|\mathbf{B}|^{\frac{1}{2}}}{(2\pi^{\frac{m}{2}})}\frac{|\mathbf{A}|^{\frac{1}{2}}}{(2\pi^{\frac{n}{2}})}\exp\bigl(S(\hat{\mathbf{w}})\bigr)\int\limits_{\mathbf{w}\in\mathbb{W}}\exp\left(-\frac{1}{2}\Delta\mathbf{w}^{\intercal}\mathbf{H}\Delta\mathbf{w}\right)d\mathbf{w} \to \max_{\mathbf{A},\mathbf{B}}. \tag{20}$$

Отметим, что подынтегральное выражение в (20) является частью нормального распределения, поэтому весь интеграл в (20) можно заменить на нормировочную константу и получить оптимизационную задачу вида:

$$g(\mathbf{A},\mathbf{B}) = \frac{|\mathbf{B}|^{\frac{1}{2}}}{(2\pi^{\frac{m}{2}})}\frac{|\mathbf{A}|^{\frac{1}{2}}}{(2\pi^{\frac{n}{2}})}\exp\bigl(S(\hat{\mathbf{w}})\bigr)\frac{(2\pi^{\frac{n}{2}})}{|\mathbf{H}|^{\frac{1}{2}}} \to \max_{\mathbf{A},\mathbf{B}}. \tag{21}$$

Прологарифмируем выражение (21) и будем искать оптимум в виде:

$$-\ln g(\mathbf{A},\mathbf{B}) = -\frac{m}{2}\ln(2\pi) + \frac{1}{2}\ln|\mathbf{A}| + \frac{1}{2}\ln|\mathbf{B}| - S(\mathbf{w}_0) - \frac{1}{2}\ln|\mathbf{H}|. \tag{22}$$

Для дальнейших рассуждений примем некоторые ограничения на вид матриц $\mathbf{A},\mathbf{B}$, позволяющие упростить вид функции $\ln g(\mathbf{A},\mathbf{B})$. В частности, везде далее будем рассматривать случай скалярной матрицы $\mathbf{B} = \beta\mathbf{I}$. В случае скалярной матрицы $\mathbf{B} = \beta\mathbf{I}$, функция ошибки (19) записывается следующим образом:

$$S(\mathbf{w},\mathbf{A},\beta) = \frac{\beta}{2}(\mathbf{y}-\mathbf{f})^{\intercal}(\mathbf{y}-\mathbf{f}) + \frac{1}{2}\mathbf{w}^{\intercal}\mathbf{A}\mathbf{w} = \beta S_{\mathfrak{D}}(\mathbf{w}) + \frac{1}{2}\mathbf{w}^{\intercal}\mathbf{A}\mathbf{w}, \tag{23}$$

где

$$S_{\mathfrak{D}}(\mathbf{w}) = \frac{1}{2}(\mathbf{y}-\mathbf{f})^{\intercal}(\mathbf{y}-\mathbf{f}), \tag{24}$$

и гессиан $\mathbf{H}$ записывается в виде

$$\mathbf{H} = \beta\mathbf{H}_{\mathfrak{D}} + \mathbf{A},$$

где $\mathbf{H}_{\mathfrak{D}}$ — гессиан функции $S_{\mathfrak{D}}(\mathbf{w})$ в точке $\mathbf{w} = \hat{\mathbf{w}}$.

Пусть вектор параметров $\mathbf{w}_0 = [w_{1(0)},\ldots,w_{n(0)}]^{\intercal}$ фиксирован.

Теорема 4. *В окрестности вектора параметров* $\mathbf{w}_0$ *оценка ковариационных матриц* $\mathbf{A}^{-1}, \mathbf{B}^{-1}$ *для гипотезы нормального распределения зависимой переменной имеет вид*

$$\alpha_i = \frac{1}{2}\lambda_i \left(\sqrt{1 + \frac{4}{(w_i - w_{i(0)})^2 \lambda_i}} - 1 \right), \text{ где } \lambda_i = \beta \mathbf{diag}(h_i),$$

$$\beta = \frac{m - \gamma}{2(\mathbf{f} - \mathbf{y})^\intercal \mathbf{B}'(\mathbf{f} - \mathbf{y})}, \text{ где } \gamma = \sum_{j=1}^{W} \frac{\lambda_j}{\lambda_j + \alpha_j}.$$

Последовательности $\|\mathbf{A}_{k+1} - \mathbf{A}_k\|^2$ *и* $\|\beta_{k+1} - \beta_i\|^2$ *монотонно убывают с увеличением номера шага* k.

В работе рассматриваются частные случаи скалярной и диагональной матрицы $\mathbf{A}$, что позволяет дифференцировать слагаемое

$$\frac{1}{2} \ln |\beta \mathbf{H}_{\mathfrak{D}} + \mathbf{A}|,$$

а также стохастический алгоритм, позволяющий получать матрицу общего вида.

5. Выбор моделей для данных в разнородных шкалах и экспертных оценок

В этом разделе описаны способы выбора моделей при построении интегральных индикаторов качества сложных объектов с использованием экспертных оценок.

Задано множество $\Upsilon = \{\upsilon_1, ..., \upsilon_m\}$ объектов и множество показателей $\Psi = \{\psi_1, ..., \psi_n\}$. Произвольный объект υ_i описывается с помощью вектора-строки $\mathbf{x}_i = \langle x_{i1}, x_{i2}, ..., x_{in} \rangle : \mathbf{x}_i \in \mathbb{R}^n$. Множество измерений представляется в виде матрицы исходных данных, обозначаемой $\mathbf{X} = \{x_{ij}\}_{i,j=1}^{m,n}$ в пространстве действительных чисел: $\mathbf{X} \in \mathbb{R}^{m \times n}$. Элемент x_{ij} — значение j-го показателя ψ_j для i-го объекта υ_i.

Интегральным индикатором объекта $\upsilon_i \in \Upsilon$ с номером i называется скаляр y_i, поставленный в соответствие набору $\mathbf{x}_i$ описаний объекта. При рассмотрении множества объектов Υ вектор $\mathbf{y} = \langle y_1, ..., y_m \rangle^\intercal : \mathbf{y} \in \mathbb{R}^m$ считается интегральным индикатором множества объектов, описанных матрицей $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^m : \mathbf{X} \in \mathbb{R}^{m \times n}$.

Согласованными значениями интегрального индикатора и весов показателей называются такие векторы $\mathbf{y}$ и $\mathbf{w}$, при которых выполняются условия

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\mathbf{w}, \\ \mathbf{w} &= \mathbf{X}^{+}\mathbf{y}, \end{aligned} \tag{25}$$

где $\mathbf{X}^{+}$ — линейное отображение, псевдообратное отображению $\mathbf{X}$, такое, что $\mathbf{X}\mathbf{X}^{+}\mathbf{X} = \mathbf{X}$, $\mathbf{X}^{+}\mathbf{X}\mathbf{X}^{+} = \mathbf{X}^{+}$ и $(\mathbf{X}\mathbf{X}^{+})^{\intercal} = \mathbf{X}\mathbf{X}^{+}$, $(\mathbf{X}^{+}\mathbf{X})^{\intercal} = \mathbf{X}^{+}\mathbf{X}$. Задачей предлагаемого метода является такое уточнение экспертных оценок, которое соответствовало бы условию (25).

Заданы экспертные оценки $\mathbf{y}_0, \mathbf{w}_0$, допускающие произвольные монотонные преобразования. Задана матрица описаний объектов $\mathbf{X} \in \mathbb{R}^{m\times n}$.Без ограничения общности будем считать, что на наборах экспертных оценок введено отношение порядка такое, что

$$y_1 \geqslant y_2 \geqslant ... \geqslant y_m \geqslant 0 \quad \text{и} \quad w_1 \geqslant w_2 \geqslant ... \geqslant w_n \geqslant 0. \tag{26}$$

Для выполнения этого условия достаточно переставить элементы векторов $\mathbf{y}_0$, $\mathbf{w}_0$ и соответствующие им строки и столбцы матрицы $\mathbf{X}$ местами.

Обозначим двухдиагональную матрицу $\mathbf{J}$ и перепишем (26) в виде

$$\mathbf{J}_m\mathbf{y} \geqslant \mathbf{0} \quad \text{и} \quad \mathbf{J}_n\mathbf{w} \geqslant \mathbf{0}.$$

Число строк квадратной матрицы $\mathbf{J}$ равно числу неравенств в системе, а число элементов каждой строки равно числу элементов вектора ($\mathbf{y}$ или $\mathbf{w}$).

Обозначим конусы, заданные экспертными оценками в пространстве интегральных индикаторов и в пространстве весов показателей, соотвественно $\mathcal{Q}$ и $\mathcal{W}$:

$$\begin{aligned} \mathcal{Q} &= \{\mathbf{y} | \mathbf{J}_m\mathbf{y} \geqslant \mathbf{0}\}, \\ \mathcal{W} &= \{\mathbf{w} | \mathbf{J}_n\mathbf{w} \geqslant \mathbf{0}\}. \end{aligned} \tag{27}$$

Нижний индекс 0, указывающий на то, что оценка поставлена экспертом, опущен, так как векторы $\mathbf{y}, \mathbf{w}$ рассматриваются как произвольные элементы множеств.

Линейное отображение $\mathbf{X}$ переводит конус $\mathcal{W} \ni \mathbf{w}_0$ экспертных оценок показателей (27) в вычисленный конус $\mathbf{X}\mathcal{W} = \mathcal{P} \ni \mathbf{w}_1$:

$$\mathbf{X} : \mathcal{W} \to \mathcal{P},$$
$$\mathbf{X} : \mathbf{w}_0 \mapsto \mathbf{y}_1.$$

Рассмотрим следующие варианты:

1) конусы $\mathcal{P}$ и $\mathcal{Q}$ пересекаются, в этом случае экспертные оценки считаются согласованными и найдется такая пара $\mathbf{y}_p \in \mathcal{P} \cup \mathcal{Q}, \mathbf{w}_p = \mathbf{X}^+\mathbf{y}_p \in \mathcal{W}$, которая удовлетворяет условию согласованности (25);

2) пересечение конусов $\mathcal{P}$ и $\mathcal{Q}$ пусто, в этом случае требуется уточнение экспертных оценок.

В случае пустого пересечения конусов $\mathcal{Q}_p = \mathcal{Q} \cap \mathbf{X}\mathcal{W}$ и $\mathcal{W}_p = \mathcal{W} \cap \mathbf{X}^+\mathcal{Q}$ предлагается использовать модифицированный метод уточнения экспертных в линейных шкалах. В пространстве интегральных индикаторов рассмотрим лучи, заданные векторами $\mathbf{y} \in \mathcal{Q}$ и $\mathbf{p} \in \mathcal{P} = \mathbf{X}\mathcal{W}$. Найдем ближайшие друг к другу лучи на ребрах или гранях конусов $\mathcal{Q}, \mathcal{P}$,

$$\cos(\mathbf{y}, \mathbf{p}) = \frac{\mathbf{y}^\intercal \mathbf{p}}{\|\mathbf{y}\|\|\mathbf{p}\|} \to \max.$$

и выполним процедуру уточнения на точках, задающих эти лучи. Отыскиваемая пара $\mathbf{y}, \mathbf{p}$ должна выполнять следующие условия:

$$\begin{array}{ll} \text{maximize} & \mathbf{y}^\intercal \mathbf{p} \\ \text{subject to} & \mathbf{y}^\intercal \mathbf{y} = 1, \quad \mathbf{p}^\intercal \mathbf{p} = 1, \\ & \mathbf{J}_n \mathbf{y} \geqslant \mathbf{0} \quad \mathbf{X}\mathbf{J}_m \mathbf{p} \geqslant \mathbf{0}. \end{array}$$

Построим итерационный алгоритм, последовательно находящий приближения векторов $\mathbf{y}^{(2k)}, \mathbf{p}^{(2k+1)}$ на четном и нечетном шаге. Векторы $\mathbf{x} = \mathbf{y}^{(2k)}$ и $\mathbf{y} = \mathbf{p}^{(2k+1)}$ будем считать решениями двух последовательно решаемых оптимизационных задач, полагая произвольным вектор $\mathbf{p}^{(0)} \in \mathcal{P}$ на шаге $k = 0$.

Задача $2k$:		Задача $2k+1$:	
maximize	$\mathbf{x}^\intercal \mathbf{p}^{(2k)}$	maximize	$\mathbf{y}^{T(2k+1)}\mathbf{y}$
subject to	$\mathbf{x}^\intercal \mathbf{x} = 1,$	subject to	$\mathbf{y}^\intercal \mathbf{y} = 1,$
	$\mathbf{J}_n \mathbf{x} \geqslant 0.$		$\mathbf{X}\mathbf{J}_m \mathbf{y} \geqslant 0.$

При решении задач, на каждом шаге значение констант $\mathbf{p}^{(2k)}$ и $\mathbf{y}^{(2k+1)}$ принимается равным значениям соответствующих решений $\mathbf{x}$ и $\mathbf{y}$ предыдущего шага. Так как максимизируемые функции и ограничения обеих задач являются выпуклыми, то решение будет найдено за счетное число шагов. Методы выпуклой оптимизации, используемые для получения численных решений, хорошо исследованы и описаны, например, в [Boyd: 2004, Minoux: 1990].

Получив решения задачи — векторы $\hat{\mathbf{p}}$ и $\hat{\mathbf{y}}$, выполняем процедуру линейного уточнения оценок интегрального индикатора

$$\mathbf{y}_\alpha = (1-\alpha)\hat{\mathbf{p}} + \alpha\hat{\mathbf{y}},$$

при условии существования нетривиального решения $\mathbf{y}_\alpha$, то есть, $\hat{\mathbf{p}}^\intercal\hat{\mathbf{y}} \neq -1$. Как было показано ранее, вектор $\mathbf{y}_\alpha$ и соответствующий ему вектор $\mathbf{w}_\alpha = \mathbf{X}^+\mathbf{y}_\alpha$ удовлетворяют условию согласованности (25). Эти векторы задают в соответствующих пространствах конусы $\mathcal{W}$ и $\mathcal{Q}$, причем пересечение $\mathcal{W}_p = \mathbf{X}\mathcal{W} \cap \mathcal{Q}$ не пусто. Так же, как и в случае уточнения оценок у линейных шкалах, при значении параметра $\alpha \to 0$, предпочтение отдается экспертным оценкам качества объектов. При $\alpha \to 1$ предпочтение отдается экспертным оценкам важности показателей.

6. Анализ прикладных задач

Поставленные в первой главе задачи регрессионного анализа и их решения играют важную роль в ряде прикладных областей. Принятый способ постановки задач, в терминах $\arg\max$, позволяет разбить работы по решению прикладных задач на несколько независимых частей. Постановка прикладной задачи как задачи регрессионного анализа включает следующие шаги.

1. Строится регрессионная выборка, определяются общие цели моделирования.

2. Назначается функция ошибки и ограничения на регрессионную модель. Функция ошибки может быть назначена исходя из гипотезы порождения данных, либо исходя из прикладных соображений, например, из требова-

ний к минимизации риска, максимизации прибыли, из стандартов физико-химических измерений и прочих.

3. Назначается класс регрессионных моделей, из которых будет выбрана модель оптимальной структурной или статистической сложности.

4. Задача выбора модели ставится как оптимизационная задача с ограничениями. Выбираются алгоритмы оптимизации для ее решения.

5. Исходя из гипотезы порождения данных или исходя из прикладных соображений выполняется ряд тестов, которые оценивают качество и свойства выбранной модели.

Рекурсивная форма векторной авторегрессионной модели имеет вид

$$\mathbf{y}_t = \sum_{\tau=0}^{r}(\mathbf{A}_\tau \mathbf{y}_{t-\tau} + \mathbf{B}_\tau \mathbf{u}_{t-\tau} + \mathbf{C}_\tau \mathbf{z}_{t-\tau}) + \mathbf{m} + \boldsymbol{\varepsilon}_t. \tag{28}$$

Здесь вектор управляющих воздействий $\mathbf{u}^\intercal$ и присоединенный к нему справа вектор внешних воздействий $\mathbf{z}^\intercal$ образуют транспонированный вектор экзогенных переменных, а матрица коэффициентов $\mathbf{B}$ и присоединенная к ней справа матрица $\mathbf{C}$ образуют матрицу коэффициентов, на которую вектор экзогенных переменных умножается слева.

В выражении (28) переменная t — дискретное время $t = 1, ..., t_0$, t_0 — последний наблюдаемый такт времени. Переменная τ обозначает глубину лагирования, причем $\tau = 1, ..., r < t_0$. Также переменная $\mathbf{m}$ есть регрессионное среднее и $\boldsymbol{\varepsilon}_t$ — регрессионный остаток, в общем различный в каждый момент времени. Так как состояние $\mathbf{y}$ объекта управления описано m переменными, а управляющие $\mathbf{u}$ и неуправляемые $\mathbf{z}$ внешние воздействия описаны соответственно q и k переменными, то матрицы $\mathbf{A} \in \mathbb{R}^{m\times m}, \mathbf{B} \in \mathbb{R}^{m\times p}, \mathbf{C} \in \mathbb{R}^{m\times k}$ и векторы $\mathbf{y}, \mathbf{m}, \boldsymbol{\varepsilon} \in \mathbb{R}^m, \mathbf{u} \in \mathbb{R}^p, \mathbf{z} \in \mathbb{R}^k$.

В первом столбце таблицы показаны значения лаговой переменный τ, которые соответствуют матрицам коэффициентов напротив. Левая часть уравнения (28) — вектор $\mathbf{y}$ показан в верхней строке таблицы. Он получается путем суммирования всех матриц, транспонированных и умноженных слева на

соответствующие векторы, которые показаны в правом столбце таблицы, а также транспонированного вектора регрессионного среднего $\mathbf{m}$ (нижняя строка таблицы) и вектора авторегрессионного остатка $\boldsymbol{\varepsilon}_t$ (в таблице не показан). Из таблицы видно, что заполняемость ненулевыми коэффициентами невысока. В частности, все элементы матриц $\mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4, \mathbf{C}_3$ равны нулю, а матрицы $\mathbf{C}_2, \mathbf{C}_4$ имеют только по одному ненулевому элементу. В данной работе обсуждается только этот способ идентификации моделей (28), поэтому анализ заполняемости матриц $\mathbf{A}, \mathbf{B}, \mathbf{C}$ и нахождение оптимальной глубины лагирования τ останется за рамками исследования.

Представим выражение (28) в приведенной форме. Для этого перенесем вектор $\mathbf{y}_t$ состояния объекта управления в левую часть и получим выражение

$$\mathbf{I}\mathbf{y}_t - \mathbf{A}_0\mathbf{y}_t = \mathbf{B}_0\mathbf{u}_t + \mathbf{C}_0\mathbf{z}_t + \sum_{\tau=1}^{r}(\mathbf{A}_\tau\mathbf{y}_{t-\tau} + \mathbf{B}_\tau\mathbf{u}_{t-\tau} + \mathbf{C}_\tau\mathbf{z}_{t-\tau}) + \mathbf{m} + \boldsymbol{\varepsilon}_t,$$

здесь $\mathbf{I}$ — единичная матрица. Матрица линейного оператора $\mathbf{A} : \mathbf{y} \longrightarrow \mathbf{y}$ предлагаемой эконометрической модели имеет полный ранг. Следовательно, матрица $(\mathbf{I} - \mathbf{A}_0)$ не вырождена. Найдем обратную матрицу $(\mathbf{I} - \mathbf{A}_0)^{-1}$ и получим выражение

$$\mathbf{y}_t = (\mathbf{I} - \mathbf{A}_0)^{-1}\left(\mathbf{B}_0\mathbf{u}_t + \mathbf{C}_0\mathbf{z}_t + \sum_{\tau=1}^{r}(\mathbf{A}_\tau\mathbf{y}_{t-\tau} + \mathbf{B}_\tau\mathbf{u}_{t-\tau} + \mathbf{C}_\tau\mathbf{z}_{t-\tau}) + \mathbf{m} + \boldsymbol{\varepsilon}_t\right). \qquad (29)$$

Пусть известно состояние $\mathbf{y}_t$ объекта управления и внешние воздействия $\mathbf{u}_t, \mathbf{z}_t$ в течение времени $t = 1, ..., t_0$. Чтобы спрогнозировать состояние объекта управления для момента времени $t = t_1$ необходимо подставить в выражение (29) значения векторов измерений экзогенных переменных $\mathbf{u}_t, \mathbf{z}_t$ в моменты времени $t = t_0, t_0 - 1..., t_0 - r$, вектора $\mathbf{y}_t$ измерений эндогенных переменных в моменты времени $t = t_0 - 1..., t_0 - r$, а также значения матриц коэффициентов $\mathbf{A}_\tau, \mathbf{B}_\tau, \mathbf{C}_\tau$, где τ=0, ..., r.

Модель субъекта определяет связь между списком альтернатив принимаемых решений и переменных, которые управляют субъектом. Для заданных элементов a из множества альтернатив $\mathcal{A} = \{a\}$ определяются значения управляемых переменных, $\mathbf{u} = \mathbf{u}(\mathcal{A})$. Принятие той или иной управляющей альтернативы определяет состояние субъекта управления и индикатора состояния

объекта. И наоборот, задавая индикатор состояния или переменные состояния объекта мы определяем значения управляемых переменных и находим ближайшую соответствующую этим значениям альтернативу.

При моделировании систем управления различают две задачи: *прямую и обратную*. Прямая задача заключается в нахождении состояния объекта управления при заданных управляющих воздействиях, см. (29). Обратная задача заключается в нахождении управляющих воздействий, которые требуются для достижения заданного состояния объекта при некоторых условиях, которые будут описаны ниже.

Прямая задача нахождения состояния $\mathbf{y}_t$ объекта управления по экзогенным переменным $\mathbf{u}_t, \mathbf{z}_t$, согласно эконометрической модели, решается посредством выражения (29). Для решения задачи управления, то есть, нахождения таких управляющих воздействий $\mathbf{u}$, которые бы привели объект управления в заданное состояние $\bar{\mathbf{y}}$, рассмотрим зависимость состояния $\mathbf{y}_t$ от управляющих воздействий $\mathbf{u}_t, ..., \mathbf{u}_{t-r}$. Для этого выберем из множества элементов $\{u^{(1)}_{t,\tau}, ..., u^{(k)}_{t,\tau}, t = t_0, \tau = 0, ..., r\}$ векторов $\mathbf{u}_{t-\tau}$, такие элементы $u^{*(j)}$, составляющие вектор управления $\mathbf{u}_t = [u^{*(1)}, ..., u^{*(k)}]^\intercal$ что для $i = 1, ..., p$ и $j = 1, ..., k$ выполняется условие

$$b_{ij,\tau} \neq 0, \tau = \min(0, ..., r),$$

где $\mathbf{B}_\tau = \{b_{ij,\tau}\}$. Другими словами выберем такие элементы вектора управляющих воздействий, которые для данного прогнозируемого состояния в момент времени t являются существенными, имеют ненулевые коэффициенты. При этом необходимо учитывать, что управляющее воздействие было последним по времени относительно состояния объекта управления. Также рассмотрим в качестве примера влияние управляемых переменных `gt` и `tr` на вектор $\mathbf{y}$ состояния объекта управления.

Подставляя в выражение (29) значения векторов фазовых траекторий $(\mathbf{y}_{t_0-1}, ..., \mathbf{y}_{t_0-r})$, $(\mathbf{z}_{t_0-1}, ..., \mathbf{z}_{t_0-r})$ и $(\mathbf{u}_{t_0-1}, ..., \mathbf{u}_{t_0-r})$ за исключением элементов вектора $\mathbf{u}_t$ и упрощая это выражение, получаем

$$\mathbf{y}_t = \mathbf{G}_r \mathbf{u}_t + \mathbf{h}_{t,r}, \tag{30}$$

где $\mathbf{G}_r \in \mathbb{R}^{p\times k}$ — новая матрица коэффициентов для управляемых переменных $\mathbf{u}_t$, значение которой вычисляется для заданного r и $\mathbf{h}_{t,r} \in \mathbb{R}^m$ — вектор, вычисляемый для заданного момента времени по известным значениям фазовых траекторий.

Уравнение обратной задачи

$$\mathbf{u}_t = \mathbf{G}_r^+(\mathbf{y}_t - \mathbf{h}_{t,r}) \tag{31}$$

получается путем псевдообращения оператора $\mathbf{G}$. Так как $\mathbf{G} \in \mathbb{R}^{m\times p}$, то псевдообратная матрица $\mathbf{G}^+ \in \mathbb{R}^{p\times m}$ при выполнении условия $\mathbf{G}^+\mathbf{G} = \mathbf{I}_p$.

Для псевдообращения используется сингулярное разложение матрицы $\mathbf{G} = \mathbf{W}\mathbf{\Lambda}\mathbf{V}^\intercal$. Так как $\mathbf{W}$ и $\mathbf{V}$ являются ортогональными матрицами, а $\mathbf{\Lambda}$ — диагональная матрица, то справедливо равенство $\mathbf{G}^+ = \mathbf{V}^\intercal\mathbf{\Lambda}^{-1}\mathbf{W}$, причем $\mathbf{G}^+\mathbf{G} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{W}^\intercal\mathbf{W}\mathbf{\Lambda}\mathbf{V}^\intercal = \mathbf{I}_k$.

Требуется подобрать такую последовательность управляющих воздействий $(\mathbf{u}_{t_1}, ..., \mathbf{u}_{t_n})$, при ограничениях на управление $\mathbf{u}_t \in \Delta\mathbf{u}_t$, которая бы при некотором заданном сценарии внешних воздействий обеспечивала бы через n шагов состояние $\bar{\mathbf{y}}_{t_n} \in \Delta\mathbf{u}_{t_n}$. В рамках данной задачи определим две: задачу *наискорейшего приближения* к целевому состоянию и задачу *оптимальноого управления.*

Задача наискорейшего приближения не является оптимальной в том смысле, что для ее решения не назначается функция общей стоимости управления; требуется подобрать такие векторы управления $(\mathbf{u}_{t_0}, ..., \mathbf{u}_{t_n})$ при ограничениях $\mathbf{u}_t \in \Delta\mathbf{u}_t$, которые бы минимизировали расстояние между целевым вектором $\bar{\mathbf{y}}_{t_n}$ и вектором текущего состояния $\mathbf{y}_t$ на каждом шаге.

Для этого на каждом шаге, начиная с t_0, отыскивается такое новое состояние $\mathbf{y}_{t+1} = \alpha\bar{\mathbf{y}}_{t_n} + (1-\alpha)\mathbf{y}_t$ объекта управления, что

$$\alpha = \arg\min_{\mathbf{u}_{t+1}\in\Delta\mathbf{u}_{t+1}} ||\bar{\mathbf{y}}_{t_n} - \mathbf{y}_{t+1}||^2,$$

где параметр $\alpha \in [0, 1]$. Данный алгоритм стремится достичь заданное состояние независимо от характера внешних воздействий $(\mathbf{z}_{t_1}, ..., \mathbf{z}_{t_n})$.

В задаче оптимального управления, как и в предыдущей, заданы сценарий внешних воздействий $(\mathbf{z}_{t_1}, ..., \mathbf{z}_{t_n})$, ограничения $\Delta\mathbf{u}_t$ на управляющие воздей-

ствия $\mathbf{u}_t$ и целевой вектор $\bar{\mathbf{y}}_n$. Требуется найти такую последовательность векторов $(\mathbf{u}_{t_1}, ..., \mathbf{u}_{t_n})$, при ограничениях $\mathbf{u}_t \in \Delta\mathbf{u}_t$, которая приводила бы объект управления из начального состояния $\mathbf{y}_{t_0}$ в целевое состояние $\bar{\mathbf{y}}_{t_n}$ за n шагов при минимальной стоимости управления $F(\mathbf{u}_{t_1}, ..., \mathbf{u}_{t_n}) \to \min$.

Заключение

Решены следующие задачи.

1. Исследована связь между пространством данных и пространством параметров моделей. Результаты опубликованы в работах, посвященных связанному байесовскому выводу.

2. Исследована связь между целевой функцией и распределением параметров модели. При этом использованы методы тестирования статистических гипотез.

3. Исследованы методы выбора из счетного или континуального множества линейных, полиномиальных, криволинейных и обобщенно-линейных моделей. Методы выбора моделей определяются классом рассматриваемых моделей. При этом выбор производится из конечного числа моделей.

4. Развит метод порождения моделей, который заключается в последовательной модификации элементов суперпозиций. Исследованы и описаны топологические свойства пространства параметров моделей. На основании этого исследования предложен метод последовательного порождения некоторых классов нелинейных моделей.

5. Исследована проблема порождения топологически изоморфных моделей с различной структурой суперпозиции. Сформулирован набор ограничений, которые требуется наложить на порождаемое множество моделей для исключения топологически изоморфных моделей.

6. Развиты и предложены новые методы оценки гиперпараметров моделей. В частности, исследована связь между методом оптимального прореживания,

методом наименьших углов и аппроксимацией Лапласа для пространства параметров. Указаны классы моделей, к которым применимы предложенные методы оценки гиперпараметров.

Получены следующие научные результаты.

1. Метод оценки гиперпараметров в связанном байесовском выводе обобщен на случай многомерного распределения параметров моделей, описываемого ковариационной матрицей. Новизна заключается в том, что обобщенный метод может быть использован для оценки информативности элементов широкого класса нелинейных моделей.

2. Выполнена работа по сравнению различных алгоритмов порождения и выбора линейных регрессионных моделей. Цель данной работы – изучение способов нахождения глобального минимума суммы квадратов невязок при последовательном добавлении элементов модели. Основной результат работы – получен новый эвристический алгоритм выбора линейной модели на основе ранее предложенного метода наименьших углов. Этот алгоритм имеет большую устойчивость к данным с множественной корреляцией признаков.

3. Исследованы ранговые регрессионные модели. В частности, разработан новый метод согласования экспертных оценок при построении интегральных индикаторов в ранговых шкалах. Исследованы свойства регрессионных моделей, параметры и зависимые переменные которых принадлежат конусам. Исследованы способы сравнения и выбора таких моделей. Предложенный метод использован для построения интегрального индикатора воздействия тепловых электростанций на окружающую среду.

4. Разработаны способы оценки гиперпараметров для основных классов моделей. Результатом этого исследования является методология создания оптимизационных алгоритмов, итеративно вычисляющих параметры и гиперпараметры моделей. В частности, рассмотрены линейные модели, обобщенные линейные модели, криволинейные модели с нелинейными параметрами базовых функций, полилинейные модели, одно и двухслойные нейронные сети, функции радиального базиса и существенно нелинейные модели.

5. Созданы базовые алгоритмы порождения оптимальных нелинейных регрессионных моделей: алгоритм обобщенного индуктивного порождения произвольных суперпозиций нелинейных параметрических функций и алгоритм порождения линейных и полиномиальных суперпозиций нелинейных параметрических функций.

6. Созданы алгоритмы выбора регрессионных моделей из индуктивно заданного множества: комбинаторный алгоритм перебора моделей ограниченной сложности, многорядный алгоритм выбора наиболее информативных мономов полинома Колмогорова-Габора, алгоритм генетического выбора наиболее информативных мономов и генетический оптимизационный алгоритм выбора моделей, представленных произвольными суперпозициями нелинейных функций.

7. Разработана система алгоритмов поиска оптимальных моделей для решения задач нелинейной регрессии и получения адекватных устойчивых регрессионных моделей. В качестве элементов, порождающих множество моделей, был использован набор аналитических функций. Модели были идентифицированы по ряду тестовых и реальных обучающих выборок, выполнен анализ адекватности этих моделей. Параметры моделей оцениваются с помощью квазиньютоновских методов оптимизации. Для поиска моделей используются алгоритмы стохастической оптимизации.

8. Показано, что поиск моделей в неявно заданном множестве возможно выполнять путем анализа значений гиперпараметров, поставленных в соответствие элементам моделей.

Публикации по теме диссертации

Нижеприведенные работы опубликованы в журналах из списка ВАК.

[1] Kuznetsov M.P., Strijov V.V. Methods of expert estimations concordance for integral quality estimation // Expert Systems with Applications, 2014, 41(4): 1988-1996.

[2] Motrenko A.P., Strijov V.V., Weber G.-W. Bayesian sample size estimation for logistic regression // Journal of Computational and Applied Mathematics, 2014, 255: 743-752.

[3] Strijov V., Krymova E.A., Weber G.W. Evidence optimization for consequently generated models // Mathematical and Computer Modelling, 2013, 57(1-2): 50-56.

[4] Strijov V., Granic G. et al. Integral Indicator of Ecological Footprint for Croatian Power Plants // Energy, 2011, 36(7): 4144-4149.

[5] Strijov V., Weber G.-W. Nonlinear regression model generation using hyperparameter optimization // Computers and Mathematics with Applications, 2010, 60(4): 981-988.

[6] Цыганова С.В., Стрижов В.В. Построение иерархических тематических моделей коллекции документов // Прикладная информатика, 2013, 1: 109-115.

[7] Стрижов В.В. Функция ошибки в задачах восстановления регрессии // Заводская лаборатория, 2013, 79(5): 65-73.

[8] Медведникова М.М., Стрижов В.В. Построение интегрального индикатора качества научных публикаций методами ко-кластеризации // Известия Тульского государственного университета, Естественные науки, 2013, 1: 154-165.

[9] Адуенко А.А., Стрижов В.В. Алгоритм оптимального расположения названий коллекции документов // Программная инженерия, 2013, 3: 21-25.

[10] Будников Е.А., Стрижов В.В. Оценивание вероятностей появления строк в коллекции документов // Информационные технологии, 2013, 4: 40-45.

[11] Зайцев А.А., Стрижов В.В., Токмакова А.А. Оценка гиперпараметров регрессионных моделей методом максимального правдоподобия // Информационные технологии, 2013, 2: 11-15.

[12] Иванова А.В., Адуенко А.А., Стрижов В.В. Алгоритм построения логических правил при разметке текстов // Программная инженерия, 2013, 6: 41-48.

[13] Кузьмин А.А., Стрижов В.В. Проверка адекватности тематических моделей коллекции документов // Программная инженерия, 2013, 4: 16-20.

[14] Медведникова М.М., Стрижов В.В. Построение интегрального индикатора качества научных публикаций методами ко-кластеризации // Известия Тульского государственного университета, Естественные науки, 2013, 1: 154-165.

[15] Рудой Г.И., Стрижов В.В. Алгоритмы индуктивного порождения суперпозиций для аппроксимации измеряемых данных // Информатика и её применения, 2013, 7(1): 17-26.

[16] Адуенко А.А., Кузьмин А.А., Стрижов В.В. Выбор признаков и оптимизация метрики при кластеризации коллекции документов // Известия Тульского государственного университета, Естественные науки, 2012, 3: 119-131.

[17] Мотренко А.П., Стрижов В.В. Многоклассовая логистическая регрессия для прогноза вероятности наступления инфаркта // Известия ТулГУ, 2012, 1: 153-162.

[18] Стрижов В.В., Кузнецов М.П., Рудаков К.В. Метрическая кластеризация последовательностей аминокислотных остатков в ранговых шкалах // Математическая биология и биоинформатика, 2012, 7(1): 345-359.

[19] Сандуляну Л.Н., Стрижов В.В. Выбор признаков в авторегрессионных задачах прогнозирования // Информационные технологии, 2012, 7: 11-15.

[20] Токмакова А.А., Стрижов В.В. Оценивание гиперпараметров линейных и регрессионных моделей при отборе шумовых и коррелирующих признаков // Информатика и её применения, 2012, 6(4): 66-75.

[21] Кузнецов М.П., Стрижов В.В., Медведникова М.М. Алгоритм многоклассовой классификации объектов, описанных в ранговых шкалах // Научно-технический вестник С.-Пб.ПГУ. Информатика. Телекоммуникации. Управление 2012, 5: 92-95.

[22] Крымова Е.А., Стрижов В.В. Алгоритмы выбора признаков линейных регрессионных моделей из конечного и счетного множеств, Заводская лаборатория, 2011, 77(5): 63-68.

[23] Стрижов В.В., Крымова Е.А. Выбор моделей в линейном регрессионном анализе // Информационные технологии, 2011, 10: 21-26

[24] Стрижов В.В. Уточнение экспертных оценок, выставленных в ранговых шкалах, с помощью измеряемых данных // Заводская лаборатория, 2011, 77(7): 72-78.

[25] Стрижов В.В., Сологуб Р.А. Индуктивное порождение регрессионных моделей предполагаемой волатильности для опционных торгов // Вычислительные технологии, 2009, 14(5): 102-113.

[26] Стрижов В.В., Казакова Т.В. Устойчивые интегральные индикаторы с выбором опорного множества описаний // Заводская лаборатория, 2007, 73(7): 72-76.

[27] Стрижов В.В. Поиск параметрической регрессионной модели в индуктивно заданном множестве // Вычислительные технологии, 2007, 1: 93-102.

[28] Стрижов В.В. Уточнение экспертных оценок с помощью измеряемых данных // Заводская лаборатория, 2006, 72(7): 59-64.