На правах рукописи

Адуенко Александр Александрович

ВЫБОР МУЛЬТИМОДЕЛЕЙ В ЗАДАЧАХ КЛАССИФИКАЦИИ

01.01.09 — Дискретная математика и математическая кибернетика

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Москва — 2017

Работа выполнена на Кафедре интеллектуальных систем Федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (государственный университет)».

Научный руководитель:	**Стрижов Вадим Викторович** доктор физико-математических наук, Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Отдел интеллектуальных систем, научный сотрудник.
Официальные оппоненты:	**Чуличков Алексей Иванович** доктор физико-математических наук, профессор, Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М. В. Ломоносова», Кафедра математического моделирования и информатики физического факультета, профессор.
	Бурнаев Евгений Владимирович кандидат физико-математических наук, доцент, Автономная некоммерческая организация высшего профессионального образования «Сколковский институт науки и технологий», Центр по научным и инженерным вычислительным технологиям для задач с большими массивами данных, доцент.
Ведущая организация:	Федеральное государственное бюджетное образовательное учреждение высшего образования «Тульский государственный университет».

Защита состоится "18" мая 2017 года в 15:00 на заседании диссертационного совета Д 002.073.05 при Федеральном исследовательском центре «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН) по адресу: 119333, г. Москва, ул. Вавилова, д. 40.

С диссертацией можно ознакомиться в библиотеке и на сайте ФИЦ ИУ РАН http://www.frccsc.ru/

Автореферат разослан "___"_________________ 2017 года.

Ученый секретарь
диссертационного совета Д 002.073.05,
д.ф.-м.н., профессор В.В.Рязанов

Общая характеристика работы

Актуальность темы.

Исследуется проблема построения мультимоделей в задаче классификации (Verlinde: 1999, Gelman: 2006, Grün: 2007, Ge: 2006, Van: 2003, Moerbeek: 2001, Muthén: 1999, Yuksel: 2012). Задача классификации является базовой в машинном обучении, при этом задачи многоклассовой классификации могут быть эффективно сведены к решению одной или нескольких задач двухклассовой классификации (Motrenko: 2014, Joshi: 2015, Tax: 2002, Liu: 2005, Rifkin: 2004). Задачами двухклассовой классификации является задача определения наличия заболевания у пациента по набору его анализов (Tolles: 2016, Bagley: 2001), задача анализа текстов для получения настроения сообщений (Supriya: 2016) и задача кредитного скоринга (Siddiqi: 2006, Paleologo: 2010, Hosmer: 2000). Эти задачи являются актуальными в связи с распространением дистанционной диагностики, автоматических систем принятия решений.

Логистическая регрессия, являющаяся стандартом в кредитном скоринге (Paleologo: 2010, Лужбин: 2013, Siddiqi: 2006), и другие обобщенно-линейные модели не позволяют учесть неоднородности в данных, в частности зависимость важности признака от объекта, а потому неоптимальны при ее наличии. Для учета неоднородностей в данных используют композиции классификаторов (Bishop: 2006, Van: 2003, Zakrzewska: 2015). Методы построения композиции моделей позволяют учесть неоднородность в данных путем построения мультимодели, содержащей несколько одиночных моделей. Модели в мультимодели могут быть близки или совпадать, что ведет к неинтерпретируемости и снижению качества прогноза. В работе (Margineantu: 1997) предлагают эвристики для прореживания ансамбля моделей в бэггинге. В работах (Zhou: 2003, Zhou: 2002) для выбора подмножества моделей в бэггинге используют генетические алгоритмы. В работах (Bakker: 2003, Giacinto: 2001) используют кластеризацию моделей и выбор единственного представителя для каждого кластера. В работах (Martínez-Muñoz: 2006, Martinez-Muoz: 2009) предлагают жадную стратегию постепенного наращивания числа классификаторов в бэггинге. Для контроля числа моделей используют априорное поощряющее разреженность распределение весов моделей в смеси (Bishop: 2006). Структуру смеси отыскивают путем максимизации обоснованности (MacKay: 1992, MacKay: 1992, Yuksel: 2012). Однако эти методы прореживания смесей не учитывают близости между моделями, а потому мультимодель по-прежнему может содержать близкие модели. Для получения статистически различимых моделей в мультимодели используют внешнюю процедуру прореживания, основанную на статистическом сравнении моделей путем подсчета расстояний между апостериорными распределениями параметров для разных моделей, например, с помощью дивергенций Брегмана или f-дивергенций (Basseville: 2013, Veyrat-Charvillon: 2009, Frigyik: 2008, Petz: 2007). В данной работе показано, что существующие меры сходства различают неинформативную модель и совпадающую информативную, а потому не позволяют построить адекватную мультимодель. Для решения этой проблемы предложена функция сходства, позволяющая решать задачу статистического различения моделей. Предлагаемый подход позволяет учесть неоднородности в данных, получить адекватную мультимодель, содержащую меньшее число моделей и имеющую лучшее качество классификации.

Наличие избыточных или мультикоррелированных признаков влияет не только на качество классификации построенной модели, но и на ее устойчивость (Стрижов: 2013, Katrutsa: 2015). Для решения задачи отбора признаков в данной работе в рамках байесовского подхода используется принцип максимума обоснованности для определения структуры моделей (MacKay: 1992, MacKay: 1992, Yuksel: 2012, Bishop: 2006). Для решения проблемы мультиколлинеарности признаков строят набор немультиколлинеарных признаков путем оптимизации критерия качества, предложенного в (Katrutsa: 2015, Gheyas: 2010). В данной работе показано, что подход, связанный с отбором признаков, является неоптимальным. Доказано, что метод максимума обоснованности не позволяет учесть зависимости между признаками, поскольку оценка максимума обоснованности для ковариационной матрицы весов признаков является асимптотически вырожденной. Для оптимального учета информации от мультиколлинеарных признаков предлагается их комбинировать.

Цели работы.

1. Разработка статистического подхода к задаче сравнения моделей в мультимоделях.
2. Построение и теоретическое обоснование функции сходства плотностей апостериорных распределений, позволяющей решать задачу сравнения моделей.
3. Разработка методов прореживания мультимоделей для построения адекватных мультимоделей.
4. Построение метода учета мультиколлинеарности между признаками.

Задачи работы.

1. Разработать статистический подход к задаче сравнения моделей в мультимоделях с помощью функций сходства апостериорных распределений.
2. Получить оценки на максимальное число моделей в адекватной мультимодели.
3. Предложить метод совместного обучения и отбора признаков для смеси моделей.
4. Разработать алгоритмы построения адекватных оптимальных обученных мультимоделей и провести вычислительный эксперимент для проверки улучшения качества и интерпретируемости построенных мультимоделей, а также для установления границ применимости предлагаемых методов.

Основные положения, выносимые на защиту.

1. Разработаны методы выбора адекватных оптимальных обученных мультимоделей в задачах распознавания и классификации, содержащих попарно статистически различимые модели.
2. Предложена функция сходства плотностей апостериорных распределений параметров моделей, удовлетворяющая требованиям к функции сходства для решения задачи сравнения моделей.
3. Получены верхняя и нижняя оценки на максимальное число моделей в адекватной мультимодели.
4. Предложен метод комбинирования мультиколлинеарных признаков. Доказана асимптотическая вырожденность недиагональной оценки ковариационной матрицы параметров логистической модели, полученной из принципа максимума обоснованности.

Методы исследования. Для достижения поставленных целей используются методы построения мультимоделей для двухклассовой классификации (Grün: 2007, Ge: 2006,Van: 2003, Moerbeek: 2001, Yuksel: 2012). Для оценки параметров многоуровневых моделей используются методы выпуклой оптимизации (Boyd: 2004, Bishop: 2006). Для обучения смесей моделей используется вариационный EM-алгоритм (Palmer: 2005, Hoffman: 2013, Wang: 2013), а для учета многоэкстремальности используется процедура мультистарта (Morales-Enciso: 2015). Для построения оптимальных многоуровневых моделей используются методы аппроксимации обоснованности (MacKay: 1992, MacKay: 1992) с помощью аппроксимации Лапласа (Bishop: 2006) и вариационных нижних оценок (Gibbs: 2000, Blei: 2016). Построение оптимальных смесей моделей производится с помощью методов вариационного байесовского вывода (Hoffman: 2013, Palmer: 2005), а для аппроксимации обоснованности используются аппроксимация Лапласа (Bishop: 2006) и построение вариационных нижних оценок (Gibbs: 2000, Blei: 2016).

Научная новизна. Разработана теория построения адекватных мультимоделей, все модели в которых являются попарно статистически различимыми. Предложен метод статистического сравнения моделей в мультимодели на основании предложенной функции сходства апостериорных распределений параметров моделей. Показано, что предлагаемая функция сходства является корректной. Исследованы статистические свойства распределения предлагаемой функции сходства в условиях истинности гипотезы о совпадении моделей. Предложен метод совместной оптимизации параметров и отбора признаков для смесей моделей. Показана асимптотическая вырожденность недиагональной оценки максимума обоснованности для ковариационной матрицы весов признаков. Предложен метод комбинирования мультиколлинеарных признаков на основании оценки ковариационной матрицы для повышения качества классификации.

Получены верхняя и нижняя оценки на максимальное число попарно различимых моделей в мультимодели.

Теоретическая значимость. Построена функция сходства, позволяющая решить задачу статистического сравнения моделей. Исследованы асимптотические свойства распределения предложенной функции сходства в условиях истинности гипотезы о совпадении моделей. На основании этих статистических свойств построена теория выбора (s, α) – адекватных мультимоделей. Получены верхняя и нижняя оценка на максимальное число моделей в адекватной мультимодели. Предложен алгоритм совместной оптимизации параметров смеси моделей и отбора признаков. Показано, что недиагональная оценка максимума обоснованности для ковариационной матрицы весов признаков является асимптотически вырожденной, а потому для учета зависимостей между признаками предложен метод их комбинирования.

Практическая значимость. Предложенные в работе вычислительные методы предназначены для построения адекватных оптимальных обученных мультимоделей значимо повышают качество распознавания и классификации и снижают число моделей в мультимоделях в прикладных задачах скоринга.

Степень достоверности и апробация работы. Достоверность результатов подтверждена математическими доказательствами, экспериментальной проверкой полученных методов на реальных задачах; публикациями результатов исследования в рецензируемых научных изданиях, в том числе рекомендованных ВАК. Результаты работы докладывались и обсуждались на следующих научных конференциях.

1. Международная конференция «20th Conference of the International Federation of Operational Research Societies», 2014. *Multimodelling and Object Selection for Banking Credit Scoring.*
2. Всероссийская конференция «57я научная конференция МФТИ», 2014. *Топологический анализ пространства параметров в задаче выбора мультимоделей.*
3. Международная конференция «27th European Conference for Operational Research», 2015. *Multimodelling and Model Selection in Bank Credit Scoring.*
4. Всероссийская конференция «Математические методы распознавания образов» ММРО-17, 2015. *Анализ пространства параметров в задачах выбора мультимоделей.*
5. Международная конференция «Интеллектуализация обработки информации», 2016. *Анализ пространства параметров в задачах выбора мультимоделей.*

Публикации по теме диссертации. Основные результаты по теме диссертации изложены в 14 печатных изданиях, девять из которых изданы в журналах, рекомендованных ВАК.

Личный вклад. Все приведенные результаты, кроме отдельно оговоренных случаев, получены диссертантом лично при научном руководстве д.ф.-м.н. В. В. Стрижова.

Структура и объем работы. Диссертация состоит из оглавления, введения, пяти разделов, заключения, списка иллюстраций, списка таблиц, перечня основных обозначений и списка литературы из 110 наименований. Основной текст занимает 157 страниц.

Основное содержание работы

Во введении обоснована актуальность диссертационной работы, сформулированы цели и методы исследования, поставлены основные задачи, обоснована научная новизна, теоретическая и практическая значимость полученных результатов, приведено краткое содержание работы по главам.

В главе 1 вводятся основные понятия и определения. Рассматривается задача классификации, ее решение в общем виде, а также понятие оптимальности и обучения вероятностной модели классификации. Приводится определение модели логистической регрессии, многоуровневой модели и смеси моделей, а также априорные распределения на параметры моделей и веса моделей в мультимодели.

Определение 1. *Объектом* называется пара $(\mathbf{x},\, y)$, где $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n$ есть вектор признакового описания объекта, а $y \in \pm 1$ есть метка класса.

Определение 2. *Признаковой матрицей* для выборки $\mathfrak{D} = \{(\mathbf{x}_i, y_i)\}$, $i \in \mathcal{I} = \{1,\, \ldots,\, m\}$ размера m называется матрица $\mathbf{X} = [\mathbf{x}_1,\, \ldots,\, \mathbf{x}_m]^\intercal \in \mathbb{R}^{m \times n}$.

Определение 3. *Вектором ответов (вектором значений целевой переменной)* для выборки $\mathfrak{D} = \{(\mathbf{x}_i, y_i)\}$, $i \in \mathcal{I} = \{1,\, \ldots,\, m\}$ размера m называется вектор $\mathbf{y} = [y_1,\, \ldots,\, y_m]^\intercal \in \{-1,\, 1\}^m$.

Определение 4. *Вероятностной моделью двухклассовой классификации* называется совместное распределение вида

$$p(y,\, \mathbf{w}|\mathbf{x},\, \boldsymbol{\mu}) = p(y|\mathbf{x},\, \mathbf{w})p(\mathbf{w}|\boldsymbol{\mu}) :\ \mathcal{Y} \times \mathcal{W} \times \mathcal{X} \to \mathbb{R}^+,$$

где $\mathbf{w} \in \mathcal{W}$ есть набор параметров модели, $p(y|\mathbf{x},\, \mathbf{w})$ задает правдоподобие объекта, а $\boldsymbol{\mu} \in Q_{\boldsymbol{\mu}}$ есть набор гиперпараметров, задающих априорное распределение параметров $p(\mathbf{w}|\boldsymbol{\mu})$.

Определение 5. *Обоснованностью* вероятностной модели двухклассовой классификации для простой выборки $\mathfrak{D} = (\mathbf{X},\, \mathbf{y}) = \{\mathbf{x}_i,\, y_i\}$, $i \in \{1,\, \ldots,\, m\}$ называется величина

$$p(\mathbf{y}|\mathbf{X},\, \boldsymbol{\mu}) = \int p(\mathbf{w}|\boldsymbol{\mu}) \prod_{i=1}^{m} p(y_i|\mathbf{w},\, \mathbf{x}_i) d\mathbf{w}.$$

Определение 6. Вероятностная модель двухклассовой классификации называется *оптимальной* для простой выборки $\mathfrak{D} = (\mathbf{X},\, \mathbf{y})$, если гиперпараметры модели выбраны из условия максимума обоснованности, то есть

$$\boldsymbol{\mu}^* = \mathop{\arg\max}_{\boldsymbol{\mu}\in Q_{\boldsymbol{\mu}}} p(\mathbf{y}|\mathbf{X},\, \boldsymbol{\mu}).$$

Определение 7. *Обучением* вероятностной модели двухклассовой классификации по простой выборке $\mathfrak{D} = (\mathbf{X}, \mathbf{y})$ называется получение оценок максимума апостериорной вероятности для параметров модели, то есть

$$\mathbf{w}^* = \mathop{\arg\max}_{\mathbf{w}\in\mathcal{W}} \frac{p(\mathbf{y},\, \mathbf{w}|\mathbf{X},\, \boldsymbol{\mu})}{p(\mathbf{y}|\mathbf{X},\, \boldsymbol{\mu})} = \mathop{\arg\max}_{\mathbf{w}\in\mathcal{W}} p(\mathbf{w}|\boldsymbol{\mu}) \prod_{i=1}^{m} p(y_i|\mathbf{w},\, \mathbf{x}_i).$$

Определение 8. *Оптимальной обученной вероятностной моделью двухклассовой классификации* для простой выборки $\mathfrak{D} = (\mathbf{X},\, \mathbf{y})$ называется оптимальная вероятностная модель (см. определение 6), для которой произведено обучение по $\mathfrak{D}$ для оптимального значения гиперпараметров, то есть

$$\boldsymbol{\mu}^* = \mathop{\arg\max}_{\boldsymbol{\mu}\in Q_{\boldsymbol{\mu}}} p(\mathbf{y}|\mathbf{X},\, \boldsymbol{\mu}),\ \mathbf{w}^* = \mathop{\arg\max}_{\mathbf{w}\in\mathcal{W}} p(\mathbf{w}|\boldsymbol{\mu}^*) \prod_{i=1}^{m} p(y_i|\mathbf{w},\, \mathbf{x}_i).$$

Определение 9. *Моделью логистической регрессии (или одиночной моделью логистической регрессии)* называется вероятностная модель двухклассовой классификации, для которой

$$p(y,\, \mathbf{w}|\mathbf{x},\, \boldsymbol{\mu}) = \sigma(y\mathbf{w}^{\mathsf{T}}\mathbf{x})p(\mathbf{w}|\boldsymbol{\mu}),$$

где $\mathbf{w} \in \mathcal{W} = \mathbb{R}^n$, $p(\mathbf{w}|\boldsymbol{\mu})$ есть априорное распределение на вектор параметров $\mathbf{w}$, а $\sigma(x) = 1/(1+\exp(-x))$.

Определение 10. *Смесью моделей* называется вероятностная модель, совместное распределение которой для простой выборки $\mathfrak{D} = (\mathbf{X},\, \mathbf{y})$ размера m имеет вид

$$p(\mathbf{y},\, \boldsymbol{\pi},\, \mathbf{w}_1,\, \ldots,\, \mathbf{w}_K|\mathbf{X},\, \boldsymbol{\mu},\, \mathbf{A}_1,\, \ldots,\, \mathbf{A}_K) = \\ p(\boldsymbol{\pi}|\boldsymbol{\mu}) \prod_{k=1}^{K} p_k(\mathbf{w}_k|\mathbf{A}_k) \prod_{i=1}^{m} \left(\sum_{l=1}^{K} \pi_l f_l(\mathbf{w}_l,\, \mathbf{x}_i,\, y_i) \right),$$

где K – число моделей, входящих в мультимодель, $f_l(\mathbf{w}_l,\, \mathbf{x}_i,\, y_i)$ есть правдоподобие для объекта $(\mathbf{x}_i,\, y_i)$ в модели l; $\boldsymbol{\pi} = [\pi_1,\, \ldots,\, \pi_K]^{\mathsf{T}}$ есть веса моделей, входящих в мультимодель, $\sum_{k=1}^{K} \pi_k = 1$; $\mathbf{w}_1,\, \ldots,\, \mathbf{w}_K$ есть параметры моделей, входящих в мультимодель, а $\mu \in Q_\mu$, $\mathbf{A}_1 \in Q_{\mathbf{A}_1},\, \ldots,\, \mathbf{A}_K \in Q_{\mathbf{A}_K}$ есть гиперпараметры, определяющие априорные распределения вектора весов $\boldsymbol{\pi}$ и векторов параметров моделей $\mathbf{w}_1,\, \ldots,\, \mathbf{w}_K$ соответственно; Q_μ, $Q_{\mathbf{A}_1},\, \ldots,\, Q_{\mathbf{A}_K}$ есть множества допустимых значений гиперпараметров.

Определение 11. *Многоуровневой моделью* называется вероятностная модель, совместное распределение которой для простой выборки $\mathfrak{D} = (\mathbf{X},\ \mathbf{y})$ размера m имеет вид

$$p(\mathbf{y},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K|\mathbf{X},\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_K) = \prod_{k=1}^{K} p_k(\mathbf{w}_k|\mathbf{A}_k) \prod_{i=1}^{m}\prod_{l=1}^{K} f_l(\mathbf{w}_l,\ \mathbf{x}_i,\ y_i)^{[\mathbf{x}_i \in \mathcal{X}_l]},$$

где K – число моделей, входящих в многоуровневую модель, $f_l(\mathbf{w}_l,\ \mathbf{x}_i,\ y_i)$ есть правдоподобие для объекта $(\mathbf{x}_i,\ y_i)$ в модели l; $\mathbb{R}^n = \mathcal{X}_1 \sqcup \ldots \sqcup \mathcal{X}_K$ есть разбиения пространства на области действия моделей; $\mathbf{w}_1,\ \ldots,\ \mathbf{w}_K$ есть параметры моделей, входящих в мультимодель; $\mathbf{A}_1 \in Q_{\mathbf{A}_1},\ \ldots,\ \mathbf{A}_K \in Q_{\mathbf{A}_K}$ есть гиперпараметры, определяющие априорные распределения векторов параметров моделей $\mathbf{w}_1,\ \ldots,\ \mathbf{w}_K$; $Q_{\mathbf{A}_1},\ \ldots,\ Q_{\mathbf{A}_K}$ есть множества допустимых значений гиперпараметров.

В качестве моделей, составляющих мультимодель, используются одиночные модели логистической регрессии. В качестве априорных распределений на веса моделей в смеси моделей используется симметричное распределение Дирихле $p(\boldsymbol{\pi}|\mu) = \mathrm{Dir}(\boldsymbol{\pi}|\mu\mathbf{e})$, а на вектора параметров моделей, входящих в мультимодель накладываем априорное нормальное распределение с нулевым средним $p(\mathbf{w}_k|\mathbf{A}_k) = \mathcal{N}(\mathbf{w}_k|\mathbf{0},\ \mathbf{A}_k^{-1})$.

Определение 12. Смесь моделей называется *оптимальной*, если она обладает наибольшей обоснованностью (см. определение 5), то есть совместное правдоподобие мультимодели имеет вид $p(\mathbf{y},\ \boldsymbol{\pi},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K|\mathbf{X},\ \mu^*,\ \mathbf{A}_1^*,\ \ldots,\ \mathbf{A}_k^*)$, где

$$[\mu^*,\ \mathbf{A}_1^*,\ \ldots,\ \mathbf{A}_K^*] = \underset{\mu \in Q_\mu,\ \mathbf{A}_1 \in Q_{\mathbf{A}_1},\ \ldots,\ \mathbf{A}_K \in Q_{\mathbf{A}_K}}{\arg\max} p(\mathbf{y}|\mathbf{X},\ \mu,\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_K). \qquad (1)$$

Определение 13. Многоуровневая модель называется *оптимальной*, если она обладает наибольшей обоснованностью (см. определение 5), то есть совместное правдоподобие многоуровневой модели имеет вид $p(\mathbf{y},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K|\mathbf{X},\ \mathbf{A}_1^*,\ \ldots,\ \mathbf{A}_k^*)$, где

$$[\mathbf{A}_1^*,\ \ldots,\ \mathbf{A}_K^*] = \underset{\mathbf{A}_1 \in Q_{\mathbf{A}_1},\ \ldots,\ \mathbf{A}_K \in Q_{\mathbf{A}_K}}{\arg\max} p(\mathbf{y}|\mathbf{X},\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_K).$$

Определение 14. *Обучением* смеси моделей, заданной совместным правдоподобием $p(\mathbf{y},\ \boldsymbol{\pi},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K|\mathbf{X},\ \mu,\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_k)$, называется получение оценок максимума апостериорной вероятности на веса моделей, входящих в смесь, и на векторы их параметров, то есть

$$[\boldsymbol{\pi}^*,\ \mathbf{w}_1^*,\ \ldots,\ \mathbf{w}_K^*] = \underset{\boldsymbol{\pi},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K}{\arg\max}\ p(\boldsymbol{\pi},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K|\mathbf{y},\ \mathbf{X},\ \mu,\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_K).$$

Определение 15. *Обучением* многоуровневой модели, заданной совместным правдоподобием $p(\mathbf{y},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K|\mathbf{X},\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_k)$, называется получение оценок максимума апостериорной вероятности на векторы параметров моделей, входящих в многоуровневую модель, то есть

$$[\mathbf{w}_1^*,\ \ldots,\ \mathbf{w}_K^*] = \mathop{\arg\max}_{\mathbf{w}_1,\ \ldots,\ \mathbf{w}_K} p(\mathbf{w}_1,\ \ldots,\ \mathbf{w}_K|\mathbf{y},\ \mathbf{X},\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_K).$$

В главе 2 рассматривается задача построения оптимальной мультимодели, а также комбинации признаков для учета взаимосвязей между ними.

Теорема 1 (Адуенко, 2016)**.** Пусть имеется l линейно независимых факторов и целевой вектор параметров $\mathbf{v} \in \mathbb{R}^l$. Обозначим $\mathbf{f}_i$ – вектор значений факторов для i–го объекта. Пусть для каждого объекта вместо $\mathbf{f}_i$ наблюдается $\mathbf{x}_i = \mathbf{G}\mathbf{f}_i + \boldsymbol{\varepsilon}_i$, где $\mathbf{G}$ – матрица размера $n \times l$, $n \geq l$ полного ранга, а $\boldsymbol{\varepsilon}_i$ – центрированный шум с невырожденной ковариационной матрицей $\boldsymbol{\Sigma}$.

Тогда оптимальной в терминах дисперсии шума $\mathbb{E}(\mathbf{w}^\intercal\boldsymbol{\varepsilon}_i)^2$ оценкой $\mathbf{w}$, такой, что $\mathbb{E}\mathbf{w}^\intercal\mathbf{x}_i = \mathbf{v}^\intercal\mathbf{f}_i$, является оценка вида

$$\mathbf{w} = \boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{G}^\intercal\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{v}.$$

Для одиночной логистической модели оценка максимума обоснованности для матрицы ковариаций $\mathbf{A}$ имеет вид

$$\mathbf{A}^* = \mathop{\arg\max}_{\mathbf{A}\in\mathcal{M}} p_{\boldsymbol{\gamma}}(\mathbf{y}|\mathbf{X},\ \mathbf{A}) = \mathop{\arg\max}_{\mathbf{A}\in\mathcal{M}} \int p_{\boldsymbol{\gamma}}(\mathbf{y}|\mathbf{X},\ \mathbf{w})p(\mathbf{w}|\mathbf{A})d\mathbf{w}, \tag{2}$$

где $\mathcal{M}$ – некоторое подмножество симметричных неотрицательно определенных матриц. Если $\mathbf{A}^*_{jj} = \infty$, то признак с номером j является незначимым. Отметим, что интеграл (2) аналитически не считается и для вычисления оценки максимума обоснованности для матрицы ковариаций $\mathbf{A}$ используют аппроксимацию Лапласа или вариационную нижнюю оценку для сигмоидной функции.

Аппроксимация Лапласа состоит в замене задачи (2) на задачу (3).

$$\log I(\mathbf{A}) = \log p_{\boldsymbol{\gamma}}(\mathbf{y}|\mathbf{X},\, \mathbf{w}_{MP}) + \tfrac{1}{2}\log\det\mathbf{A} - \tfrac{1}{2}\mathbf{w}_{MP}^\intercal\mathbf{A}\mathbf{w}_{MP} - \tfrac{1}{2}\log\det\mathbf{H}^{-1} \to \max_{\mathbf{A}\in\mathcal{M}}, \tag{3}$$

где $\mathbf{w}_{MP}$ есть оценка максимума апостериорной вероятности для вектора параметров модели. Для решения этой задачи используется итерационный алгоритм, состоящий в поочередном пересчете оценок $\mathbf{A}$ и $\mathbf{w}_{MP}$.

Определение 16. *Вариационной оценкой функции* $f(x) : \mathbb{R} \to \mathbb{R}$ называется функция $g(x, \xi) : \mathbb{R}^2 \to \mathbb{R}$, такая что:

$$f(x) \geq g(x, \xi) \quad \forall\, x, \xi;\ f(\xi) = g(\xi, \xi) \quad \forall\, \xi.$$

Для сигмоиды существует вариационная нижняя оценка вида

$$\sigma(x) \geq \sigma(\xi)\exp\{(x-\xi)/2 - \lambda(\xi)(x^2 - \xi^2)\},\ \text{ где } \lambda(\xi) = \frac{1}{2\xi}\left[\sigma(\xi) - \frac{1}{2}\right].$$

Задача максимизации обоснованности (2) заменяется задачей максимизации нижней оценки на ее логарифм

$$[\mathbf{A},\ \boldsymbol{\xi}] = \mathop{\arg\max}_{\boldsymbol{\xi},\ \mathbf{A}\in\mathcal{M}} \Bigg[\tfrac{1}{2}\log\det\mathbf{A} - \tfrac{1}{2}\log\det\mathbf{A}' + \sum_{i=1}^{m}\gamma_i\log\sigma(\xi_i) - \tfrac{1}{2}\sum_{i=1}^{m}\gamma_i\xi_i + \sum_{i=1}^{m}\gamma_i\lambda(\xi_i)\xi_i^2 + \tfrac{1}{2}\mathbf{v}^{\mathsf{T}}\mathbf{A}'^{-1}\mathbf{v}\Bigg],$$

для решения которой используется итерационный алгоритм с поочередным пересчетом $\boldsymbol{\xi}$ при фиксированной $\mathbf{A}$ и $\mathbf{A}$ при фиксированном $\boldsymbol{\xi}$.

Теорема 2 (Адуенко, 2016)**.** Пусть имеется одиночная логистическая модель, заданная совместным правдоподобием

$$p(\mathbf{y},\ \mathbf{w}|\mathbf{X},\ \mathbf{A}) = \prod_{i=1}^{m}\sigma(y_i\mathbf{w}^{\mathsf{T}}\mathbf{x}_i)\mathcal{N}(\mathbf{w}|\mathbf{0},\ \mathbf{A}^{-1}),$$

где $y_i \in \{-1,\ 1\}$, $\mathbf{x}_i \in \mathbb{R}^2$, то есть признаковое пространство имеет размерность $n = 2$. Пусть также $\mathbf{w} = [w_1,\ w_2]$, $w_1,\ w_2 \neq 0$.
Обозначим

$$\boldsymbol{\Sigma} = \mathbf{H}^{-1} = \mathbf{X}^{\mathsf{T}}\mathbf{R}\mathbf{X} = \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix},\ \mathbf{A} = \begin{pmatrix}s_1^2 & \kappa s_1 s_2\\ \kappa s_1 s_2 & s_2^2\end{pmatrix},$$

где $\mathbf{H}$ есть гессиан $-\log p(\mathbf{y}|\mathbf{X},\ \mathbf{w})$ в точке максимума апостериорной вероятности $\mathbf{w}_{MP} = \arg\max_{\mathbf{w}} p(\mathbf{y},\ \mathbf{w}|\mathbf{X})$, а $\mathbf{R} = \mathrm{diag}(\sigma(y_i\mathbf{w}_{MP}^{\mathsf{T}}\mathbf{x}_i)\sigma(-y_i\mathbf{w}_{MP}^{\mathsf{T}}\mathbf{x}_i))$,

Тогда если при $m \to \infty$ выполнено

$$\sigma_1^2,\ \sigma_2^2 \overset{\text{п.н.}}{\to} \infty,\ \mathbb{P}(\omega:\ \exists\, c_\omega > 0,\ \exists\, m_\omega:\ \forall\, m \geq m_\omega\ 1 - \rho_m^2 \geq c_\omega) = 1,$$

то $s_1^*,\ s_2^* \overset{\text{п.н.}}{\to} \infty$, $\kappa^* \overset{\text{п.н.}}{\to} -\mathrm{sign}(w_1 w_2)$.

Определение 17. Набор признаков с индексами $j \in \mathcal{J}$ называется δ – *мультиколлинеарным* для $\delta > 0$ с вектором весов $\boldsymbol{\theta}$, $\theta_j \neq 0 \iff j \in \mathcal{J}$, если $\|\hat{\boldsymbol{\Sigma}}\boldsymbol{\theta}\|_1 < \delta$.

Для детектирования и учета мультиколлинеарности предлагается следующий алгоритм, имеющий два параметра τ и δ, которые определяются с помощью кросс-валидации на обучающей выборке.

- Находим наиболее мультиколлинеарный набор признаков

  $$\boldsymbol{\theta}^* = \mathop{\arg\max}_{\boldsymbol{\theta}}\left[\|\hat{\boldsymbol{\Sigma}}\boldsymbol{\theta}\|_1 + \tau\|\boldsymbol{\theta}\|_1\right],$$ где $\tau > 0$ – коэффициент регуляризации.

- Если $\|\hat{\boldsymbol{\Sigma}}\boldsymbol{\theta}^*\|_1 \geq \delta$, то останавливаемся, поскольку найденный набор не является δ-мультиколлинеарным. Иначе переходим на шаг 3.

- Вычисляем невязку $\boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\theta}^*$ и активные признаки $\mathcal{J} = \{j: \theta_j^* \neq 0\}$.
- Пусть $j_0 \in \mathcal{J}$. Поправим признаки с номерами из $\mathcal{J}$ так, что $\mathbf{X}^{\text{new}}\boldsymbol{\theta}^* = \mathbf{0}$,

$$\mathbf{f}_j - \frac{1}{\mathbf{e}^{\mathsf{T}}\boldsymbol{\theta}^*}\boldsymbol{\varepsilon} \to \mathbf{f}_j,\ j \in \mathcal{J}.$$

- Удаляем признак с номером j_0 и шкалируем признаки с номерами $j \in \mathcal{J} \setminus \{j_0\}$ до дисперсии 1.
- Пересчитываем матрицу ковариации $\hat{\boldsymbol{\Sigma}}$. Возвращаемся на шаг 1.

В главе 3 рассматривается задача обучения мультимоделей. Для обучения смеси моделей используется вариационный EM-алгоритм, который позволяет обучить смесь моделей при известных гиперпараметрах смеси. Предложен также алгоритм совместного обучения и оптимизации смеси моделей, основанный на аппроксимации Лапласа и вариационном EM-алгоритме.

Задача обучения одиночной модели логистической регрессии имеет вид

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} -\log p(\mathbf{y},\ \mathbf{w}|\mathbf{X},\ \mathbf{A}) = \arg\min_{\mathbf{w}} l(\mathbf{w}), \tag{4}$$

которая имеет единственное решение в силу выпуклости $l(\mathbf{w})$.

Обучение многоуровневой модели состоит в решении K независимых задач обучения одиночных логистических моделей вида

$$\mathbf{w}_k^* = \arg\min_{\mathbf{w}_k} -\log p(\mathbf{y}_{\mathcal{I}_k},\ \mathbf{w}_k|\mathbf{X}_{\mathcal{I}_k},\ \mathbf{A}_k),\ k = \overline{1,\ K}.$$

Для смеси моделей совместное правдоподобие имеет вид

$$p(\mathbf{y},\ \boldsymbol{\pi},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K|\mathbf{X},\ \mu,\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_k) = \\ p(\boldsymbol{\pi}|\mu)\prod_{k=1}^{K} p_k(\mathbf{w}_k|\mathbf{A}_k)\prod_{i=1}^{m}\left(\sum_{l=1}^{K}\pi_l\sigma(y_i\mathbf{w}_l^{\mathsf{T}}\mathbf{x}_i)\right),\ \text{где}$$

$$p(\boldsymbol{\pi}|\mu) = \frac{\Gamma(K\mu)}{\Gamma^K(\mu)}\prod_{k=1}^{K}\pi_k^{\mu-1},\ p(\mathbf{w}_k|\mathbf{A}_k) = \frac{\sqrt{\det \mathbf{A}_l}}{(2\pi)^{n/2}}\exp\left(-\tfrac{1}{2}\mathbf{w}_l^{\mathsf{T}}\mathbf{A}_l\mathbf{w}_l\right),\ k = 1,\ \ldots,\ K.$$

Обучение смеси моделей состоит в решении задачи

$$[\boldsymbol{\pi}^*,\ \mathbf{w}_1^*,\ \ldots,\ \mathbf{w}_K^*] = \arg\max_{\boldsymbol{\pi},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K} \log p(\mathbf{y},\ \boldsymbol{\pi},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K|\mathbf{X},\ \mu,\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_K). \tag{5}$$

Оценим $\log p(\mathbf{y},\ \boldsymbol{\pi},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K|\mathbf{X},\ \mu,\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_K)$ снизу, вводя распределение $q(\mathbf{Z})$, получим

$$\log p(\mathbf{y},\ \boldsymbol{\pi},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K|\mathbf{X},\ \mu,\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_k) \geq L(q,\ \boldsymbol{\pi},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K) = \\ \mathbb{E}_q \log p(\mathbf{y},\ \boldsymbol{\pi},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K,\ \mathbf{Z}|\mathbf{X},\ \mu,\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_k) - \mathbb{E}_q \log q(\mathbf{Z}). \tag{6}$$

Вместо исходной задачи (5) решаем задачу

$$L(q,\ \boldsymbol{\pi},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K) \to \max_{q,\,\boldsymbol{\pi},\,\mathbf{w}_1,\,\ldots,\,\mathbf{w}_K}. \tag{7}$$

с помощью вариационного EM-алгоритма, используя для q факторизацию вида

$$Q = \left\{ q:\ q(\mathbf{Z}) = \prod_{i=1}^{m} q_{\mathbf{z}_i}(\mathbf{z}_i) \right\}.$$

Это приводит к следующему итеративному алгоритму.

1. Задаем начальное приближение для $\boldsymbol{\pi},\ \mathbf{w}_k,\ k=\overline{1,\ K}$.
2. Пересчитываем распределение скрытых переменных
$$\mathbb{P}(z_{il}=1) = \frac{\pi_l \sigma(y_i \mathbf{w}_l^\mathsf{T}\mathbf{x}_i)}{\sum_{k=1}^{K} \pi_k \sigma(y_i \mathbf{w}_k^\mathsf{T}\mathbf{x}_i)}.$$
3. Пересчитываем веса моделей в мультимодели
$$\pi_k = \begin{cases} 0, \text{ если } \mu - 1 + \sum_{i=1}^{m} \mathbb{E}z_{ik} \leq 0, \\ \dfrac{\mu - 1 + \sum_{i=1}^{m} \mathbb{E}z_{ik}}{\sum_{l=1}^{K} \max\left(0,\ \mu - 1 + \sum_{i=1}^{m} \mathbb{E}z_{il}\right)}, \text{ если } \mu - 1 + \sum_{i=1}^{m} \mathbb{E}z_{ik} > 0. \end{cases}$$
4. Пересчитываем векторы параметров моделей, решая задачу
$$l_k(\mathbf{w}_k) = \tfrac{1}{2} \sum_{k=1}^{K} \mathbf{w}_k^\mathsf{T} \mathbf{A}_k \mathbf{w}_k - \sum_{i=1}^{m} \mathbb{E}z_{ik} \log \sigma(y_i \mathbf{w}_k^\mathsf{T}\mathbf{x}_i) \to \min_{\mathbf{w}_k},\ k=\overline{1,\ K},$$
совпадающую с задачей обучения одиночной логистической модели (4) для взвешенных объектов.
5. При значительном отличии оценок $\boldsymbol{\pi},\ \mathbf{w}_k,\ k=\overline{1,\ K}$ переходим на шаг 2.

Предложим далее алгоритм, который позволяет одновременно с обучением смеси моделей производить оценку гиперпараметров априорных распределений. Воспользуемся оценкой $\log p(\mathbf{y},\ \boldsymbol{\pi},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K | \mathbf{X},\ \mu,\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_K)$ снизу (6) как аппроксимацией и будем решать задачу (7) с помощью вариационного EM-алгоритма, учитывая, что там содержатся неизвестные матрицы $\mathbf{A}_1,\ \ldots,\ \mathbf{A}_K$, которые определяются в соответствии с принципом максимума обоснованности (1), и от их значений зависят полученные оценки максимума апостериорной вероятности. Для получения оценок матриц $\mathbf{A}_1, \ldots, \mathbf{A}_K$ используем порожденную нижней оценкой аппроксимацию обоснованности

$$p(\mathbf{y}|\mathbf{X},\ \mu,\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_K) \approx e^{-\mathbb{E}_q \log q(\mathbf{Z})} \int p(\boldsymbol{\pi}|\mu) \prod_{k=1}^{K} \pi_k^{\sum_{i=1}^{m} \mathbb{E}z_{ik}} d\boldsymbol{\pi}$$
$$\prod_{k=1}^{K} \int \mathcal{N}(\mathbf{w}_k|\mathbf{0},\ \mathbf{A}_k^{-1}) \prod_{i=1}^{m} \sigma(y_i \mathbf{w}_k^\mathsf{T}\mathbf{x}_i)^{\mathbb{E}z_{ik}} d\mathbf{w}_k,$$

откуда оценки матриц $\mathbf{A}_k,\ k = \overline{1,\ K}$ являются решением задачи

$$\mathbf{A}_k^* = \arg\max_{\mathbf{A}_k} \int \mathcal{N}(\mathbf{w}_k|\mathbf{0},\ \mathbf{A}_k^{-1}) \prod_{i=1}^{m} \sigma(y_i \mathbf{w}_k^{\mathsf{T}} \mathbf{x}_i)^{\mathbb{E}z_{ik}} d\mathbf{w}_k,\ k = \overline{1,\ K},$$

совпадающей с задачей получения оценки максимума обоснованности для ковариационной матрицы одиночной логистической модели с экспоненциально взвешенными объектами с вектором весов $\boldsymbol{\gamma}_k = [\mathbb{E}z_{ik},\ i = \overline{1,\ m}]^{\mathsf{T}}$, которая рассмотрена ранее (2).

В главе 4 вводится понятие адекватной мультимодели и ставится задача статистического сравнения моделей с помощью расчета функции сходства между апостериорными распределениями параметров моделей. Рассматриваются требования к корректным функциям сходства между распределениями. Предложена корректная функция сходства, получены ее свойства, а также предложены методы построения адекватных мультимоделей. Показано, что существующие функции сходства не являются корректными.

Несмотря на прореживание мультимодели, она может содержать похожие модели. По этой причине поставим задачу сравнения моделей, решение которой используется при построении адекватной мультимодели.

- Даны две модели f_1 и f_2, векторы параметров моделей $\mathbf{w}_1$, $\mathbf{w}_2$.
- Имеем выборки $(\mathbf{X}_1,\ \mathbf{y}_1)$ и $(\mathbf{X}_2,\ \mathbf{y}_2)$, $y_{1,i} = f_1(\mathbf{x}_{1,i},\ \mathbf{w}_1),\ \ y_{2,i} = f_2(\mathbf{x}_{2,i},\ \mathbf{w}_2)$.
- Априорные распределения параметров моделей $\mathbf{w}_1 \sim p_1(\mathbf{w})$, $\mathbf{w}_2 \sim p_2(\mathbf{w})$.
- Апостериорные распределения $p(\mathbf{w}_1|\mathbf{X}_1,\ \mathbf{y}_1)$ и $p(\mathbf{w}_2|\mathbf{X}_2,\ \mathbf{y}_2)$, обозначаемые далее $g_1(\mathbf{w})$ и $g_2(\mathbf{w})$.

Требуется построить функцию сходства, определенную на паре распределений $g_1(\mathbf{w})$ и $g_2(\mathbf{w})$. Она должна удовлетворять следующим требованиям.

1. определена в случае несовпадения носителей,
2. $s(g_1,\ g_2) \leq s(g_1,\ g_1)$,
3. $s \in [0,\ 1]$,
4. $s(g_1,\ g_1) = 1$,
5. близка к 1, если $g_2(\mathbf{w})$ — малоинформативное распределение,
6. симметрична, $s(g_1,\ g_2) = s(g_2,\ g_1)$.

Свойство 5 является основным в решаемой задаче, поскольку обеспечивает неотличимость модели, про параметры которой ничего неизвестно, от любой другой модели.

Определение 18. Назовем распределение $g_2(\cdot) :\ \Omega \to \mathbb{R}^+$ *неинформативным* относительно распределения $g_1(\cdot)\ :\ \ \Omega \to \mathbb{R}^+$ с конечным носителем $\mathrm{supp}(g_1) = A$, если $\exists\, B :\ A \subseteq B$, что $\forall\, \mathbf{v} \in B :\ g_2(\mathbf{w}) = 1/|B|$.

Определение 19. Назовем распределение $g_2(\cdot) : \Omega \times \Omega_2 \to \mathbb{R}^+$ *неинформативным* относительно распределения $g_1(\cdot) :\ \Omega \times \Omega_1 \to \mathbb{R}^+$ с конечным носителем $\mathrm{supp}(g_1) = A$, если $\Omega_1 = \varnothing$, то есть g_1 определено на подпространстве области определения g_2 и $\exists \tau > 0,\ B : A \times [-\tau, \tau]^{\dim(\Omega_2)} \subseteq B$, что $\forall \mathbf{v} \in B : g_2(\mathbf{w}) = 1/|B|$.

Определение 20. Назовем последовательность распределений $g_2^1(\cdot),\ \ldots,\ g_2^k(\cdot),\ \ldots\ \Omega\ \to\ \mathbb{R}^+$ *малоинформативной* на Ω, если выполнены следующие условия

$$\forall\, a > 0,\ g_2^k(\cdot)|_A \to U(A),\ \text{где}\ A = \{\mathbf{w}:\ \|\mathbf{w}\| \leq a\},$$
$$\exists\, 0 \leq B < \infty,\ \exists\, k_0:\ \forall\, k \geq k_0 \sup_{\{\mathbf{w}:\, \|\mathbf{w}\| \geq B\}} g_2^k(\mathbf{w}) \leq \sup_{\{\mathbf{w}:\, \|\mathbf{w}\| \leq B\}} g_2^k(\mathbf{w}),$$

где в условии (8) $g_2^k(\cdot)|_A$ есть сужение распределения $g_2^k(\cdot)$ на множество A

$$g_2^k|_A(\mathbf{w}) = \begin{cases} 0,\ \text{если}\ \|\mathbf{w}\| > a, \\ \dfrac{g_2^k(\mathbf{w})}{\int_A g_2^k(\mathbf{v})d\mathbf{v}},\ \mathbf{w} \in A. \end{cases}$$

Сходимость в свойстве (8) понимается равномерная, то есть

$$g_2^k(\cdot)|_A \to U(A) \Longleftrightarrow \sup_{\mathbf{w}\in A} |g_2^k(\mathbf{w}) - 1/|A|| \to 0\ \text{при}\ k \to \infty.$$

Определение 21. Функция сходства $s(g_1,\ g_2)$, определенная на паре распределений $g_1(\mathbf{w}):\ \Omega \times \Omega_1 \to \mathbb{R}^+$ и $g_2(\mathbf{w}):\ \Omega \times \Omega_2 \to \mathbb{R}^+$, где $\Omega = \mathbb{R}^n$, $\Omega_1 = \mathbb{R}^{n_1}$, $\Omega_2 = \mathbb{R}^{n_2}$, $n,\ n_1,\ n_2 \geq 0$ называется *корректной*, если она удовлетворяет следующим требованиям.

1. $s(g_1,\ g_2)$ определена $\forall\, n,\ n_1,\ n_2 \geq 0$, если $g_1(\mathbf{w},\ \mathbf{w}_1),\ g_2(\mathbf{w},\ \mathbf{w}_2) < \infty\, \forall\, \mathbf{w} \in \Omega,\ \mathbf{w}_1 \in \Omega_1,\ \mathbf{w}_2 \in \Omega_2$,
2. $s(g_1,\ g_2) \leq s(g_1,\ g_1)$,
3. $s(g_1,\ g_2) \in [0,\ 1]$,
4. $s(g_1,\ g_1) = 1$,
5.
   - Если g_2 является неинформативным относительно g_1, то $s(g_1, g_2) = 1$;
   - $\forall\, g_1:\ n_1 = 0$, то есть $g_1(\mathbf{w},\ \mathbf{w}_1) = g_1(\mathbf{w})$, для малоинформативной последовательности распределений $g_2^1,\ \ldots,\ g_2^k,\ \ldots$ выполнено

$$s(g_1,\ g_2^k) \to 1\ \text{при}\ k \to \infty,$$

6. $s(g_1,\ g_2) = s(g_2,\ g_1)$.

Для перехода от расстояния / дивергенции к функции сходства будет использовать следующее преобразование

$$s_\rho(g_1,\ g_2) = \exp(-\rho(g_1,\ g_2)). \tag{10}$$

Теорема 3 (Адуенко, 2014)**.** Перечисленным требованиям к функции сходства для пары распределений не удовлетворяют функции сходства, порожденные а) дивергенцией Кульбака-Лейблера; б) расстоянием Дженсона-Шеннона; в) расстоянием Хеллингера; г) расстоянием Бхаттачарайа.

Определение 22. *Дивергенцией Брегмана* $D_F(p,\, q) : \Omega \times \Omega \to \mathbb{R}$, где Ω – выпуклое множество, называется заданная строго выпуклой непрерывно дифференцируемой функцией $F:\, \Omega \to \mathbb{R}$ функция двух аргументов вида

$$D_F(p,\, q) = F(p) - F(q) - \langle \nabla F(q),\, p - q \rangle.$$

Определение 23. *Симметризованной дивергенцией Брегмана* $\tilde{D}_F(p,\, q) : \Omega \times \Omega \to \mathbb{R}$ назовем

$$\tilde{D}_F(p,\, q) = \sqrt{\tfrac{1}{2} D_F\left(p, \tfrac{1}{2}(p+q)\right) + \tfrac{1}{2} D_F\left(q, \tfrac{1}{2}(p+q)\right)}.$$

Теорема 4 (Адуенко, 2016)**.** Функции сходства, порожденные дивергенциями Брегмана и симметризованными дивергенциями Брегмана в соответствии с (10), не являются корректными.

Определение 24. *f-дивергенцией* $d_f(p,\, q) : \Omega \times \Omega \to \mathbb{R}$, где Ω – множество распределений с конечными плотностями над $\mathbb{R}^n$, называется заданная выпуклой функцией $f:\, \mathbb{R} \to \mathbb{R}$, $f(1) = 0$ функция двух аргументов вида

$$d_f(p,\, q) = \int_{\mathbb{R}^n} f\left(\frac{p(\mathbf{w})}{q(\mathbf{w})}\right) q(\mathbf{w}) d\mathbf{w}.$$

Теорема 5 (Адуенко, 2016)**.** Функции сходства, порожденные f – дивергенциями в соответствии с (10), не являются корректными.

Предложим далее корректную функцию сходства s.

Определение 25. Назовем функцией сходства *s-score* пары распределений $g_1 :\, \mathbb{R}^n \to \mathbb{R}^+$, $g_2 :\, \mathbb{R}^n \to \mathbb{R}^+$, определенных на одном пространстве, функцию вида

$$s_0(g_1,\, g_2) = \frac{\int g_1(\mathbf{w}) g_2(\mathbf{w}) d\mathbf{w}}{\max_{\mathbf{b} \in \mathbb{R}^n} \int g_1(\mathbf{v}) g_2(\mathbf{v} - \mathbf{b}) d\mathbf{v}}.$$

Определение 26. Назовем пару распределений $(g_1^\tau(\mathbf{w},\, \mathbf{w}_1,\, \mathbf{w}_2) :\, \mathbb{R}^k \times \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}^+,\, g_2^\tau(\mathbf{w},\, \mathbf{w}_1,\, \mathbf{w}_2) :\, \mathbb{R}^n \times \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}^+)$ τ – расширением пары распределений $(g_1(\mathbf{w},\, \mathbf{w}_1) : \mathbb{R}^n \times \mathbb{R}^{n_1} \to \mathbb{R}^+,\, g_2(\mathbf{w},\, \mathbf{w}_2) :\, \mathbb{R}^n \times \mathbb{R}^{n_2} \to \mathbb{R}^+)$, если

$$g_1^\tau(\mathbf{w},\, \mathbf{w}_1,\, \mathbf{w}_2) = g_1(\mathbf{w},\, \mathbf{w}_1) U(Q_2^\tau),\; g_2^\tau(\mathbf{w},\, \mathbf{w}_1,\, \mathbf{w}_2) = g_2(\mathbf{w},\, \mathbf{w}_2) U(Q_1^\tau), \text{ где}$$

$$Q_1^\tau = \{\mathbf{w}_1 :\, \|\mathbf{w}_1\|_\infty \le \tau\},\; Q_2^\tau = \{\mathbf{w}_2 :\, \|\mathbf{w}_2\|_\infty \le \tau\}.$$

Определение 27. Назовем функцией сходства *s-score* пары распределений $g_1(\mathbf{w},\, \mathbf{w}_1)$, $g_2(\mathbf{w},\, \mathbf{w}_2)$, $g_1 :\, \mathbb{R}^n \times \mathbb{R}^{n_1} \to \mathbb{R}^+$, $g_2 :\, \mathbb{R}^n \times \mathbb{R}^{n_2} \to \mathbb{R}^+$ функцию вида

$$s(g_1,\, g_2) = \lim_{\tau \to 0} s_0(g_1^\tau,\, g_2^\tau).$$

Утверждение 1. Функция сходства $s(g_1,\, g_2)$ определена корректно.

Теорема 6 (Адуенко, 2014)**.** Предлагаемая функция сходства s-score является корректной.

Теорема 7 (Адуенко, 2014)**.** Пусть модели, задаваемые математическим ожиданием и ковариационной матрицей апостериорного распределения параметров $(\mathbf{m}_1,\, \mathbf{\Sigma}_1)$ и $(\mathbf{m}_2,\, \mathbf{\Sigma}_2)$ считаются различимыми, если

$$\mathrm{s-score}\,(\mathcal{N}(\mathbf{m}_1,\, \mathbf{\Sigma}_1),\, \mathcal{N}(\mathbf{m}_2,\, \mathbf{\Sigma}_2)) \le C \in (0,\, 1).$$

Тогда, если указанные модели различимы по приведенному критерию, то и модели, задаваемые $(\mathbf{m}_1,\, \mathbf{\Sigma}_1)$ и $(\mathbf{m}_2,\, \lambda\mathbf{\Sigma}_2)$, $\lambda \in [0,\, 1]$ будут различимы согласно приведенному критерию.

Теорема 8 (Адуенко, 2014)**.** Пусть рассматриваются K моделей с $\|\mathbf{m}_1\| = \ldots = \|\mathbf{m}_K\| = \lambda_1 > 0$ и $\mathbf{\Sigma}_1 = \ldots = \mathbf{\Sigma}_K = \lambda_2\mathbf{I}$. В качестве критерия отличимости моделей рассматривается следующий: модели с номерами $i \ne j$ разные, если

$$\mathrm{s-score}\,(\mathcal{N}(\mathbf{m}_i,\, \mathbf{\Sigma}_i),\, \mathcal{N}(\mathbf{m}_j,\, \mathbf{\Sigma}_j)) \le C \in (0,\, 1).$$

Тогда максимальное число попарно различимых моделей, которое может быть в наборе, есть

$$K_{max} = \left\lfloor \sqrt{\pi}\frac{n\Gamma(\frac{n+1}{2})}{(n-1)\Gamma(\frac{n}{2}+1)}\frac{1}{\int_0^{\theta/2}\sin^{n-2}\varphi d\varphi} \right\rfloor.$$

Здесь $\theta \in [0,\, \pi]$, $\cos\theta = \rho = \max(-1,\, 1 + 2\lambda_2/\lambda_1^2 \ln C)$. При этом можно построить K_{min} попарно различимых моделей, где

$$K_{min} = \left\lfloor \sqrt{\pi}\frac{n\Gamma(\frac{n+1}{2})}{(n-1)\Gamma(\frac{n}{2}+1)}\frac{1}{\int_0^{\theta}\sin^{n-2}\varphi d\varphi} \right\rfloor.$$

Определение 28. Назовем *обобщенно-линейной моделью с натуральной функцией связи* и априорным распределением на вектор параметров $p(\mathbf{w}|\mathbf{A})$ вероятностную модель, совместное правдоподобие которой имеет вид

$$p(\mathbf{y},\, \mathbf{w}|\mathbf{X},\, \mathbf{A}) = p(\mathbf{y}|\mathbf{X},\, \mathbf{w})p(\mathbf{w}|\mathbf{A}),\ \text{где}\ p(\mathbf{y}|\mathbf{X},\, \mathbf{w}) = \prod_{i=1}^{m} p(y_i|\mathbf{x}_i,\, \mathbf{w}),$$

$$p(y_i|\mathbf{x}_i,\, \mathbf{w}) = c(y_i)\exp(\theta_i y_i - b(\theta_i)),\ \text{где}\ \theta_i = \mathbf{w}^{\mathsf{T}}\mathbf{x}_i.$$

Введем следующие обозначения

$$l_m(\mathbf{v}) = \log p(\mathbf{y}|\mathbf{X},\, \mathbf{v}),\ \mathbf{s}_m(\mathbf{v}) = \nabla l_m(\mathbf{v}),\ \mathbf{H}_m(\mathbf{v}) = -\frac{\partial^2 l_m(\mathbf{v})}{\partial \mathbf{v}^2}.$$

Обозначим также $\tilde{l}_m(\mathbf{v}) = \log p(\mathbf{y}|\mathbf{X},\, \mathbf{v}) + \log p(\mathbf{w}|\mathbf{A})$. Обозначим $\mathbf{H}_m(\mathbf{w}) = \mathbf{H}_m$ и введем $O_m^{\delta}(\mathbf{w}) = \{\mathbf{v} :\ \|\mathbf{H}_m^{T/2}(\mathbf{v} - \mathbf{w})\| \le \delta\}$, $m = 1,\, \ldots$ последовательность окрестностей истинного параметра $\mathbf{w}$, где $\mathbf{H}_m^{1/2}$ есть квадратный корень Холецкого.

Теорема 9 (Адуенко (2016))**.** Пусть имеется имеется обобщенно-линейная модель с натуральной функцией связи с истинным вектором параметров $\mathbf{w}$ и нормальным априорным распределением на вектор параметров $\mathbf{w} \sim \mathcal{N}(\mathbf{w}|\mathbf{0},\, \mathbf{A}_m^{-1})$, причем $\exists\, c_0 < \infty :\ \|\mathbf{A}_m\| < c_0\ \forall\, m$. Пусть также $\sum_{i=1}^m \mathbf{x}_i\mathbf{x}_i^{\intercal}$ имеет полный ранг для $m \geq m_0$, $\lambda_{\min}(\mathbf{H}(\mathbf{w})) \to \infty$ при $m \to \infty$ и

$$\exists\, \delta > 0,\ c > 0,\ \theta > 0,\ m_1 :\ \forall\, m \geq m_1\, \forall\, \mathbf{v} \in O_m^{\delta}(\mathbf{w})\ \lambda_{\min}(\mathbf{H}(\mathbf{v})) \geq c\lambda_{\max}^{1/2+\theta}(\mathbf{H}(\mathbf{w})).$$

Тогда оценка максимума апостериорной вероятности $\hat{\mathbf{w}} = \arg\max_{\mathbf{v}} p(\mathbf{y}, \mathbf{v}|\mathbf{X}, \mathbf{A})$ существует с вероятностью, стремящейся к 1 при $m \to \infty$, и $\hat{\mathbf{w}} \xrightarrow{\text{п.н.}} \mathbf{w}$.

Теорема 10 (Адуенко (2016))**.** Пусть имеется обобщенно-линейная модель с натуральной функцией связи с истинным вектором параметров $\mathbf{w}$ и нормальным априорным распределением на вектор параметров $\mathbf{w} \sim \mathcal{N}(\mathbf{w}|\mathbf{0},\, \mathbf{A}_m^{-1})$, причем $\exists\, c_0 < \infty :\ \|\mathbf{A}_m\| < c_0\ \forall\, m$. Пусть также $\sum_{i=1}^m \mathbf{x}_i\mathbf{x}_i^{\intercal}$ имеет полный ранг для $m \geq m_0$, $\lambda_{\min}(\mathbf{H}(\mathbf{w})) \to \infty$ при $m \to \infty$ и

$$\forall\, \delta > 0\ \max_{\mathbf{v} \in O_m^{\delta}(\mathbf{w})} \|\mathbf{H}(\mathbf{w})^{-1/2}\mathbf{H}(\mathbf{v})\mathbf{H}(\mathbf{w})^{-T/2} - \mathbf{I}\| \to 0 \text{ при } m \to \infty.$$

Тогда $\tilde{\mathbf{H}}(\hat{\mathbf{w}})^{1/2}(\hat{\mathbf{w}} - \mathbf{w}) \xrightarrow{d} \mathcal{N}(\mathbf{0},\, \mathbf{I}_n)$, где $\hat{\mathbf{w}} = \arg\max_{\mathbf{v}} p(\mathbf{y},\, \mathbf{v}|\mathbf{X},\, \mathbf{A})$ есть оценка максимума апостериорной вероятности для вектора параметров модели.

Теорема 11 (Адуенко (2016))**.** Пусть имеется две обобщенно-линейные модели с натуральной функцией связи с одинаковым истинным вектором параметров $\mathbf{w}$ и нормальным априорным распределением на вектор параметров $\mathbf{w} \sim \mathcal{N}(\mathbf{w}|\mathbf{0},\, \mathbf{A}_{m^k}^{-1})$, $k = 1,\, 2$, причем $\exists\, c_0^k < \infty :\ \|\mathbf{A}_{m^k}^k\| < c_0^k\ \forall\, m^k,\ k = 1,\, 2$. Пусть для обеих моделей выполнено, что $\sum_{i=1}^{m^k} \mathbf{x}_i\mathbf{x}_i^{\intercal}$ имеет полный ранг для $m^k \geq m_0^k$, $k = 1,\, 2$, $\lambda_{\min}(\mathbf{H}_{m^k}(\mathbf{w})) \to \infty$ при $m^k \to \infty$, $k = 1,\, 2$ и

$$\forall \delta > 0\ \max_{\mathbf{v} \in O_{m^k}^{\delta}(\mathbf{w})} \|\mathbf{H}_{m^k}(\mathbf{w})^{-1/2}\mathbf{H}_{m^k}(\mathbf{v})\mathbf{H}_{m^k}(\mathbf{w})^{-T/2} - \mathbf{I}\| \to 0 \text{ при } m^k \to \infty, k = 1, 2.$$

$$\text{Пусть также } \|\tilde{\mathbf{H}}_{m^1}(\hat{\mathbf{w}}_1)\|\|\tilde{\mathbf{H}}_{m^2}^{-1}(\hat{\mathbf{w}}_2)\| \xrightarrow{\mathbb{P}} 0 \text{ при } m = \min(m^1,\, m^2) \to \infty.$$

Тогда $-2\log \text{s-score} = (\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}_1)^{\intercal}(\tilde{\mathbf{H}}_{m^1}^{-1}(\hat{\mathbf{w}}_1) + \tilde{\mathbf{H}}_{m^2}^{-1}(\hat{\mathbf{w}}_2))^{-1}(\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}_1) \xrightarrow{d} \chi^2(n)$, где $\hat{\mathbf{w}}_k$, $k = 1,\, 2$ есть оценки максимума апостериорной вероятности для вектора параметров модели, а m^k – число объектов в выборке для модели $k = 1,\, 2$.

Определение 29. Мультимодель (смесь моделей или многоуровневая модель) называется $(s,\, \alpha)$ – адекватной, если все модели $f_1,\, \ldots,\, f_l$, входящие в нее, являются попарно статистически различимыми с помощью функции сходства s на уровне значимости α. Под статистической различимостью моделей предполагается статистическая различимость апостериорных распределений параметров моделей $p(\mathbf{w}_k|\mathbf{y},\, \mathbf{X},\, \mu,\, \mathbf{A}_1,\, \ldots,\, \mathbf{A}_K)$, $k = \overline{1,\, K}$ для мультимодели и $p(\mathbf{w}_k|\mathbf{y},\, \mathbf{X},\, \mathbf{A}_1,\, \ldots,\, \mathbf{A}_K)$, $k = \overline{1,\, K}$ для многоуровневой модели.

Отметим, что так как модели, входящие в многоуровневую модель, оптимизируются независимо и никаких ограничений на их похожесть не накладывается, а разбиение признакового пространства на области действия моделей может не отражать реальной неоднородности в данных, построенная многоуровневая модель может быть не $(s,\,\alpha)$ – адекватной. При построении смеси моделей модели оптимизируются совместно, но несмотря на прореживание смесь может содержать похожие модели. Опишем теперь методы построения $(s,\,\alpha)$ – адекватной мультимодели по имеющейся оптимальной обученной мультимодели. В качестве апостериорных распределений на параметры $\mathbf{w}_1,\,\ldots,\,\mathbf{w}_K$ используем их нормальные аппроксимации. Пусть задана некоторая функция сходства s. Обозначим $\mathbf{S} = \|s_{kl}(g_k(\mathbf{w}_k),\,g_l(\mathbf{w}_l))\|,\ k,\,l\ =\ \overline{1,\,K}$ матрицу значений попарных сходств моделей, входящих в мультимодель, а $\mathbf{T} = \|t_{kl}\|,\ k,\,l = \overline{1,\,K}$ матрицу соответствующих достигаемых уровней значимости в условиях истинности гипотезы о совпадении моделей, то есть $t_{kl} = \mathbb{P}(s(g_k(\mathbf{w}_k),\,g_l(\mathbf{w}_l)) < s_{kl}|\mathbf{w}_k = \mathbf{w}_l)$.

Метод последовательного объединения максимальных клик по наибольшему сходству.

1. Находим клики попарно неразличимых моделей максимального размера

$$\tilde{\mathcal{K}} = \mathop{\mathrm{Arg\,max}}_{\mathcal{K}\in 2^{\{1,\,\ldots,\,K\}}} |\mathcal{K}| \min_{k,\,l\in\mathcal{K}} [t_{kl} \geq \alpha].$$

2. Если для $\mathcal{K} \in \tilde{\mathcal{K}}$ $\min_{k,\,l\in\mathcal{K}}[t_{kl} \geq \alpha] = 0$, то есть в клике есть различимые модели, останавливаемся, поскольку построена $(s,\,\alpha)$ – адекватная многоуровневая модель. Иначе переходим на шаг 3.
3. Среди найденных клик находим клику с максимальной суммой достигаемых уровней значимости: $\mathcal{K}^* = \arg\max_{\mathcal{K}\in\tilde{\mathcal{K}}} \sum_{k,\,l\in\mathcal{K}} t_{kl}$.
4.
    - *Для многоуровневых моделей:*
    Объединяем модели с индексами из $\mathcal{K}^*$ в одну с номером $k^* \in \mathcal{K}^*$ и пересчитываем $g_{k^*}(\mathbf{w}_{k^*})$.

$$\sqcup_{k\in\mathcal{K}^*}\mathcal{I}_k \to \mathcal{I}_{k^*},\ \mathbf{A}^*_{k^*} = \mathop{\arg\max}_{\mathbf{A}_{k^*}} p(\mathbf{y}_{\mathcal{I}_{k^*}}|\mathbf{X}_{\mathcal{I}_{k^*}},\ \mathbf{A}_{k^*}),$$

$$g_{k^*}(\mathbf{w}_{k^*}) = \mathcal{N}(\mathbf{w}_{k^*}|\mathbf{w}^*_{k^*},\ \mathbf{\Sigma}^*_{k^*}),\ \mathbf{w}^*_{k^*} = \mathop{\arg\max}_{\mathbf{w}_{k^*}} p(\mathbf{y}_{\mathcal{I}_{k^*}},\ \mathbf{w}_{k^*}|\mathbf{X}_{\mathcal{I}_{k^*}},\ \mathbf{A}^*_{k^*}),$$

$$\mathbf{\Sigma}^{*-1}_{k^*} = \mathbf{X}^{\mathsf{T}}_{\mathcal{I}_{k^*}}\mathbf{R}_{k^*}\mathbf{X}_{\mathcal{I}_{k^*}} + \mathbf{A}^*_{k^*},\ \mathbf{R}_{k^*} = \mathrm{diag}(\sigma(\mathbf{w}^{*\mathsf{T}}_{k^*}\mathbf{x}_i)\sigma(-\mathbf{w}^{*\mathsf{T}}_{k^*}\mathbf{x}_i),\ i\in\mathcal{I}_{k^*}).$$

    - *Для смесей моделей:*
    Объединяем модели с индексами из $\mathcal{K}^*$ и перенастраиваем смесь моделей в соответствии. Начальное приближение:

$$\sum_{l\in\mathcal{K}^*}\pi_l \to \pi_{k^*}, 0 \to \pi_l, l\in\mathcal{K}\backslash\{k^*\}, \frac{1}{|\mathcal{K}^*|}\sum_{l\in\mathcal{K}^*}\mathbf{w}_l \to \mathbf{w}_{k^*}, \mathbf{w}_k \to \mathbf{w}_k, k \notin \mathcal{K}^*.$$

5. Удаляем столбцы матриц $\mathbf{S}$ и $\mathbf{T}$ с номерами из $\mathcal{K}^* \setminus \{k^*\}$ и пересчитываем s_{k^*l} и t_{k^*l} для $l \neq k^*$.

$$s_{k^*l} = s(g_{k^*}(\mathbf{w}_{k^*}),\ g_l(\mathbf{w}_l)),\ t_{k^*l} = \mathbb{P}(s(g_{k^*}(\mathbf{w}_{k^*}),\ g_l(\mathbf{w}_l)) < s_{k^*l}|\mathbf{w}_{k^*} = \mathbf{w}_l).$$

   Для смесей моделей пересчитываем все попарные сходства, так как оптимизация могла изменить значения весов и параметров других моделей.

6. Переходим на шаг 1.

В главе 5 приводится сравнение результатов на синтетических и реальных данных с существующими методами. Рассматриваются задачи классификации данных из репозитория UCI по потребительским кредитам, по качеству белого вина, по локализации белков в клетках, по ценам домов, по заболеваниям сердца в Южной Африке. Данные по ценам домов относятся к задаче регрессии, а данные по качеству белого вина – к задаче многоклассовой классификации, а потому была произведена дискретизация целевой переменной.

Для оценки качества использовалась кросс-валидация по 50 независимым разбиениям выборки на обучение и контроль. Сравнение построения (s, α) – адекватных многоуровневых моделей для $K = 20$ с исходными и (s, α) – адекватных смесей моделей с исходными при автоматическом выборе числа моделей в смеси путем прореживания приведены в табл. 1 и 2 соответственно. В таблицах использованы следующие обозначения.

- $\mathrm{AUC}_{\mathrm{m/sp}}$ – среднее качество на кросс-валидации для многоуровневой модели/смеси моделей;
- $\mathrm{AUC}^{\mathrm{adeq}}_{\mathrm{m/sp}}$ – среднее качество на кросс-валидации для $(s,\ \alpha)$ – адекватной многоуровневой модели/смеси моделей;
- $t_{\mathrm{m/sp}}$ – значение выборочной t-статистики при сравнении $(s,\ \alpha)$ – адекватной многоуровневой модели/смеси моделей с исходной многоуровневой моделью/смесью моделей;
- $K_{\mathrm{m/sp}}$ – среднее количество моделей в обученной оптимальной многоуровневой модели/смеси моделей;
- $K^{\mathrm{adeq}}_{\mathrm{m/sp}}$ – среднее количество моделей в $(s,\ \alpha)$ – адекватной обученной оптимальной многоуровневой модели/смеси моделей.

Предлагаемые методы построения $(s,\ \alpha)$ – адекватных мультимоделей позволяют не только значительно сократить число моделей в мультимодели и получить адекватную мультимодель, состоящую из попарно различимых моделей, но и приводят в большинстве случаев к значимому повышению качества классификации по сравнению с исходными мультимоделями. При этом качество классификации построенных мультимоделей устойчиво к избыточному числу моделей в исходных мультимоделях, а в случае отсутствия неоднородности в данных различие в качестве построенных $(s,\ \alpha)$ – адекватных мультимоделей по отношению к одиночной модели значительно меньше, чем у исходных моделей. Эти

Таблица 1. Результаты построения многоуровневых моделей и (s, α) – адекватных многоуровневых моделей для $K = 20$.

Выборки	$\mathrm{AUC_m}$	$\mathrm{AUC_m^{adeq}}$	t_m	K_m	$K_\mathrm{m}^\mathrm{adeq}$
Кредиты	0.6789	**0.7604**	16.7	20	2.48
Локализация белков	0.6721	**0.6959**	8.83	20	2.56
Качество вина	**0.8185**	**0.8189**	0.57	20	9.08
Заболевания сердца	0.6757	**0.7392**	11.56	20	2.62
Цены домов	0.897	**0.9427**	11.34	20	2.56

Таблица 2. Результаты построения смесей моделей и (s, α) – адекватных смесей моделей с автоматическим выбором числа моделей в смеси.

Данные	$\mathrm{AUC_{sp}}$	$\mathrm{AUC_{sp}^{adeq}}$	t_sp	K_sp	$K_\mathrm{sp}^\mathrm{adeq}$
Кредиты	0.7693	**0.7784**	4.53	15.5	2.66
Локализация белков	**0.6988**	**0.7001**	1.67	17.4	2.11
Качество вина	0.8038	**0.8125**	14.01	31.5	6.48
Заболевания сердца	0.7644	**0.7705**	3.03	9.66	2.60
Цены домов	0.9411	**0.9461**	2.46	7.66	2.10

свойства позволяют рекомендовать предлагаемый подход с построением (s, α) – адекватных мультимоделей для решения задач классификации, где в данных могут содержатся неоднородности.

В заключении представлены основные результаты диссертационной работы.

1. Предложена функция сходства плотностей апостериорных распределений параметров моделей, удовлетворяющая требованиям к функции сходства для решения задачи сравнения моделей. Показано, что известные функции сходства требованиям не удовлетворяют.
2. Получены верхняя и нижняя оценки на максимальное число моделей в адекватной мультимодели.
3. Разработаны методы выбора адекватных оптимальных обученных мультимоделей, содержащих попарно статистически различимые модели.
4. Предложен метод комбинирования мультиколлинеарных признаков. Доказана асимптотическая вырожденность недиагональной оценки максимума обоснованности ковариационной матрицы параметров логистической модели.
5. Разработан программный комплекс для построения адекватных мультимоделей при решении задач классификации. Разработанные методы показывают значимое повышение качества и снижение числа моделей в мультимодели по сравнению с известными методами.

Публикации соискателя по теме диссертации

Публикации в журналах из списка ВАК.

1. А. А. Адуенко, А. А. Кузьмин, В. В. Стрижов Выбор признаков и оптимизация метрики при кластеризации коллекции документов // Известия ТулГУ, 2012. № 3. С. 119-131.
2. А. А. Адуенко, В. В. Стрижов Алгоритм оптимального расположения названий коллекции документов // Программная инженерия, 2013. № 3. С. 21–25.
3. А. В. Иванова, А. А. Адуенко, В. В. Стрижов Алгоритм построения логических правил при разметке текстов // Программная инженерия, 2013. № 6. С. 41–47.
4. А. А. Адуенко, Н. И. Амелькин О предельных движениях волчка с внутренней диссипацией в однородном поле тяжести // Труды МФТИ, 2013. № 18(2). С. 126-133.
5. А. А. Кузьмин, А. А. Адуенко, В. В. Стрижов Тематическая классификация тезисов крупной конференции с использованием экспертной модели // Информационные технологии, 2014. № 6. С. 22-26.
6. А. А. Адуенко, Н. И. Амелькин Асимптотические свойства движений тяжелого волчка с внутренней диссипацией // ПММ, 2014. Т. 78. Вып. 1. С. 13-28.
7. А. А. Адуенко, В. В. Стрижов Совместный выбор объектов и признаков в задачах многоклассовой классификации коллекции документов // Инфокоммуникационные технологии, 2014. № 1. С. 47–53.
8. А. А. Адуенко, А. С. Василейский, А. И. Карелов, И. А. Рейер, К. В. Рудаков, В. В. Стрижов Алгоритмы выделения и совмещения устойчивых отражателей на спутниковых снимках // Компьютерная оптика, 2015. Т. 39. Вып. 4. С. 622–630.
9. А. А. Адуенко, Н. И. Амелькин О резонансных вращениях маятника с вибрирующим подвесом // ПММ. 2015. Т. 79. Вып. 6. С. 756–767.

Остальные публикации.

10. Адуенко А.А. Выбор признаков и шаговая логистическая регрессия для задачи кредитного скоринга // Машинное обучение и анализ данных, 2012. № 3. С. 279-291.
11. A. A. Aduenko , V. V. Strijov Multimodelling and Object Selection for Banking Credit Scoring // 20th Conference of the International Federation of Operational Research Societies. — Barcelona: 2014.— P. 136.
12. A. A. Aduenko , V. V. Strijov Multimodelling and Model Selection in Bank Credit Scoring // 27th European Conference for Operational Research. — Glasgow: 2015.— P. 273.

13. А. А. Адуенко, В. В. Стрижов Анализ пространства параметров в задачах выбора мультимоделей // Математические методы распознавания образов ММРО-17. Тезисы докладов 17-й Всероссийской конференции с международным участием. — г. Светлогорск, Калининградская область: Торус пресс, 2015. С. 10–11.
14. А. А. Адуенко, В. В. Стрижов Анализ пространства параметров в задачах выбора мультимоделей // Интеллектуализация обработки информации ИОИ-2016. Тезисы докладов 11-й Международной конференции. — Москва, Россия-Барселона, Испания: Торус пресс, 2016. С. 10–11.