

© 2019 г. Ю.С. ПОПКОВ, д-р техн. наук (popkov@isa.ru)

(Федеральный исследовательский центр

“Информатика и управление” РАН, Москва;

Институт проблем управления РАН, Москва;

Департамент программной инженерии, ОРТ Брауде Колледж, Кармиель, Израиль;

Югорский научно-исследовательский институт информационных технологий,

Ханты-Мансийск;

Московский физико-технический институт)

## ПРОЦЕДУРЫ РАНДОМИЗИРОВАННОГО МАШИННОГО ОБУЧЕНИЯ<sup>1</sup>

Предлагается новая концепция машинного обучения, основанная на компьютерной имитации энтропийно-оптимальных рандомизированных моделей. Рассмотрены процедуры рандомизированного машинного обучения (РМО) с “жесткой” и “мягкой” рандомизацией, которые сводятся либо к точному воспроизведению эмпирических балансов в первом случае, либо к приближенному в рамках принятого критерия аппроксимации. Сформулированы алгоритмы РМО в виде функциональных задач энтропийно-линейного программирования. Приведены примеры применения РМО в задачах классификации текстов и рандомизированного прогнозирования миграционного взаимодействия региональных систем.

*Ключевые слова:* рандомизация, жесткие и мягкие процедуры рандомизации, неопределенность, энтропия, матричные нормы, эмпирические балансы, классификация текстом, динамическая регрессия.

DOI: 10.1134/S0005231019090095

### 1. Введение

Многие события, явления, процессы, объекты, которые мы пытаемся изучить и использовать, происходят в некой среде, про которую мы что-то не знаем или знаем не полностью. Другими словами, приходится исследовать, моделировать или решать проблемы в условиях неопределенности, причем содержательный смысл этого термина часто оказывается весьма размытым. Тем не менее, мы оперируем уровнем неопределенности, полагая, что его можно измерить. Чтобы этот методологический путь не оказался тупиковым, необходимо как-то имитировать (моделировать) неопределенность. Наиболее продвинутым и эффективным является подход, основанный на гипотезе о стохастической природе неопределенности. Продвинутым он является потому, что существуют подходящие инструменты, превращающие гипотезу в прикладные технологии, эффективность которых базируется на математических теориях — вероятности, статистики, оптимизации и др.

Стохастическая природа неопределенности мотивирует ее моделирование с помощью искусственно генерируемой случайной среды, называемой рандомизированной средой. Разумеется эта среда не произвольная, а отражающая предположительно существующие особенности неопределенности.

---

<sup>1</sup> Работа поддержана Российским фондом фундаментальных исследований (проект 17-29-02115).

Одним из носителей этих особенностей неопределенности являются данные. Современные средства вычислительной техники позволяют накапливать и хранить огромное количество данных как в естественных, так и в оцифрованных форматах. Естественно возникает вопрос: что можно с ними делать, кроме их сохранения. Весьма привлекательным кажется предложение, попытаться извлекать из данных новые знания. По-видимому, первые декларации на эту тему, в виде концепции Data Mining появились в публикациях [1, 2]. В рамках этой концепции предлагалось интегрировать методы математической статистики и теории вероятности, погруженные в дружественную для пользователя информационно-технологическую среду. Журнал Nature посвятил проблеме извлечения новых знаний из большого объема данных спецвыпуск [3], в котором были сформулированы так называемые “5V” атрибут, характеризующие данные, пригодные для указанной цели.

Однако, методологическим ядром многих концепций, связанных с данными, является параметризованная математическая модель, а физическим ядром является компьютер, на котором обучается указанная модель с помощью соответствующего информационного и алгоритмического обеспечения. Поэтому процедуры и алгоритмы машинного обучения (МО) являются ключевыми в современных концепциях извлечения знаний. Следует напомнить, что МО имеет более чем 60-летнюю историю и опыт решения многочисленных задач. Первая публикация из этой области относится к 1957 г., когда Ф. Розенблатт создал перцептрон МАРК I [4]. Понятие эмпирического риска, ключевое для МО-процедур, было введено в монографии Я.З. Цыпкина [5]. Метод потенциальных функций в задачах классификации и распознавания был опубликован в 1970 г. в монографии М.А. Айзермана, Э.М. Бравермана, Л.И. Розоноэра [6]. Основные идеи машинной классификации были опубликованы в монографиях [7, 8], а реализованы они в методе SVM существенно позднее [9]. Современный референс-лист работ, посвященных машинному обучению, насчитывает тысячи наименований, обзору и анализу которых можно посвятить отдельную статью. Но все-таки, некоторые из них следует упомянуть (исключительно по календарному принципу) [10–18].

Существующие представления о МО базируются на параметризованных с детерминированным происхождением моделях, параметры которых неизвестны, но, используя данные, можно найти оценки их значений.

Здесь будет развиваться иной, а именно рандомизированный подход к проблемам машинного обучения (РМО), особенность которого состоит в том, что в его рамках параметризованная модель имеет случайные параметры, а данные используются для оценивания не их значений, а функций плотности распределения вероятностей параметров. Смысл такого подхода состоит в том, что реальные задачи, для решения которых привлекаются МО-процедуры, как правило, погружены в некоторую неопределенную среду. Если речь идет о данных, то они получены с ошибками, пропусками, низкой достоверностью. Формирование моделей и их параметризация — процесс неформализуемый и субъективный, зависящий от индивидуальных знаний исследователя. Поэтому в массовом применении МО-процедур уровень неопределенности достаточно высок.

В рамках РМО формируются энтропийно-оптимальные оценки функций плотности распределения вероятностей параметров и шумов, в соответствии с которыми генерируются ансамбли моделей “наилучших” в условиях максимальной неопределенности. В статье рассмотрены два класса РМО-алгоритмов, различающиеся условиями балансировки выхода модели с реальными данными. В одном из них предполагается так называемая “жесткая” балансировка, когда выход модели в точности должен совпадать с реальными данными. В другом используется “мягкая” балансировка, когда выход приближен к реальным данным оптимально в терминах принятого критерия.

В обоих классах РМО обученная модель генерирует ансамбли случайных векторов или случайных траекторий, соответствующих найденным оптимальным функциям плотности распределения вероятностей (ПРВ) параметров и шумов. Эти ансамбли используются для решения задач “мягкой” (с вероятностью) классификации и кластеризации, а также для задач рандомизированного прогнозирования с использованием моделей динамического регрессии.

## 2. РМО-процедура, структура и классификация

Рандомизация представляется альтернативным подходом к машинному обучению, направленным на повышение достоверности, надежности и гибкости МО-процедур в условиях неопределенности и функционирования как при больших, так и при ограниченных объемах данных. Для достижения этих декларируемых целей предлагается комплекс структурных, информационно-технологических и алгоритмических модификаций стандартных МО-процедур. При этом используется те же коллекции данных для реализации РМО-процедур. Они состоят из двух массивов данных, которые будем называть входными  $X$  и выходными  $Y$ .

Данные в указанных массивах измеряются в дискретные моменты времени  $j = 1, \dots, s$ . Поэтому входные данные характеризуются матрицей  $X = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(s)}]$ , а выходные — матрицей  $Y = [\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(s)}]$ . Векторы  $\mathbf{x}^{(j)} \in \mathbb{R}^n$ ,  $\mathbf{y}^{(j)} \in \mathbb{R}^m$ . Предполагается, что массив входных данных содержит точные данные, а массив выходных данных — с *ошибками интервального типа*, которые будем характеризовать матрицей случайных шумов  $\Xi = [\xi^{(1)}, \dots, \xi^{(s)}]$ . Случайные векторы  $\xi^{(j)} \in \mathbb{R}^m$ ,  $j = \overline{1, s}$ , — независимые и компоненты их также независимые:

$$(1) \quad \begin{aligned} \xi_i^{(j)} &\in E_i^{(j)} = [\xi_i^-(j), \xi_i^+(j)], \quad i = \overline{1, m}, j = \overline{1, s}; \\ \bar{\xi}^{(j)} &\in \Xi_j = \bigcup_{i=1}^m E_i^{(j)}. \end{aligned}$$

Вероятностные свойства шумов характеризуются функциями плотности распределения вероятностей (ПРВ)  $Q_j(\bar{\xi}^{(j)})$ ,  $j = \overline{1, s}$ . Предполагается, что ПРВ — непрерывно дифференцируемые функции.

Структура РМО-процедуры представлена на рис. 1. Ее основными элементами являются рандомизированная параметризованная модель (РПМ) и алгоритм рандомизированного машинного обучения (РМО-А).

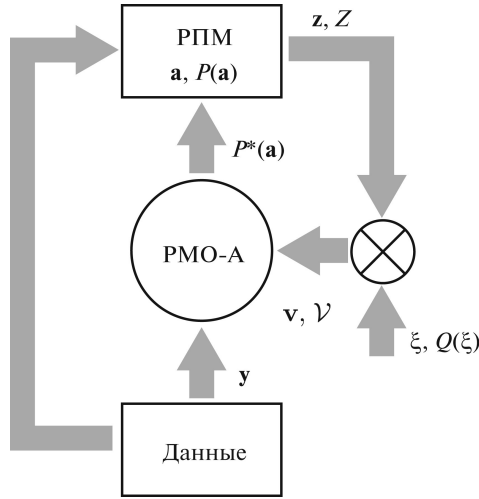


Рис. 1. Структура РМО-процедуры.

1. РПМ преобразует массив входных данных  $X$  в модельный выход, характеризуемый матрицей  $Z = [\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(s)}]$ , где  $\mathbf{z}^{(j)} \in \mathbb{R}^m$ . В общем случае связь между массивами в обучающей коллекции предполагается динамической. Это означает, что модельный выход, наблюдаемый в момент времени  $j$ , зависит от входа, наблюдаемого на некотором историческом интервале  $j - \varrho, \dots, j$ , т.е. от матрицы  $X_\varrho(j) = [\mathbf{x}^{(j-\varrho)}, \dots, \mathbf{x}^{(j)}]$ . Математическим образом этой связи является векторный функционал  $\hat{\Omega}(X_\varrho(j) | \mathbf{a}, P(\mathbf{a}))$  со случайными параметрами  $\mathbf{a}$  интервального типа

$$(2) \quad \mathbf{a} \in \mathcal{A} = [\mathbf{a}^-, \mathbf{a}^+].$$

Вероятностные свойства параметров характеризуются функцией плотности распределения вероятностей (ПРВ)  $P(\mathbf{a})$ , которая предполагается непрерывно-дифференцируемой.

Выход РПМ в  $j$ -й момент времени представляет собой ансамбль  $\hat{\mathbf{Z}}(j | P(\mathbf{a}))$  случайных векторов

$$(3) \quad \hat{\mathbf{z}}(j | \mathbf{a}) = \hat{\Omega}(X_\varrho(j) | \mathbf{a}, P(\mathbf{a})), \quad j = \overline{1, s}.$$

Для имитации влияния измерительных шумов вводятся случайные векторы

$$(4) \quad \mathbf{v}(j | \mathbf{a}, \bar{\xi}^{(j)}) = \hat{\mathbf{z}}(j | \mathbf{a}) + \bar{\xi}^{(j)}, \quad j = \overline{1, s},$$

образующие ансамбль  $\mathcal{V}(j | P(\mathbf{a}), Q_j(\bar{\xi}^{(j)}))$ . Вероятностные свойства этого ансамбля зависят от ПРВ  $P(\mathbf{a})$  и  $Q_1(\bar{\xi}^{(1)}), \dots, Q_s(\bar{\xi}^{(s)})$ . Смысл рандомизированного обучения состоит в определении таких функций ПРВ, в которых бы учитывались условия балансирования ансамбля  $\mathcal{V}(j | P(\mathbf{a}), Q_j(\bar{\xi}^{(j)}))$  с выходными обучающими данными  $Y$  (эмпирические балансы).

2. РМО-А представляет собой обучающий модуль, в котором формализован принцип оптимальности рандомизированного обучения, который опи-

сывается задачей функционального математического программирования, базовой компонентой которой является функционал информационной энтропии [19, 20], определенный на функциях ПРВ  $P(\mathbf{a})$  случайных параметров РПМ и функциях ПРВ  $Q_1(\bar{\xi}^{(1)}), \dots, Q_s(\bar{\xi}^{(s)})$  шумов:

$$(5) \quad \mathcal{H}[P(\mathbf{a}), Q_1(\bar{\xi}^{(1)}), \dots, Q_s(\bar{\xi}^{(s)})] = - \int_{\mathcal{A}} P(\mathbf{a}) \ln \frac{P(\mathbf{a})}{P^0(\mathbf{a})} d\mathbf{a} - \sum_{j=1}^s \int_{\Xi_j} Q_j(\bar{\xi}^{(j)}) \ln \frac{Q_j(\bar{\xi}^{(j)})}{Q_j^0(\bar{\xi}^{(j)})} d\bar{\xi}^{(j)},$$

где  $P^0(\mathbf{a}), Q_j^0(\bar{\xi}^{(j)})$  — априорные ПРВ параметров и шумов.

Его максимизация гарантирует получение наилучших решений при максимальной неопределенности. Эта логическая цепь была впервые декларирована в [21] (см. также [22–25]). Информационная энтропия характеризует неопределенность, связанную не только со случайными параметрами, но и с шумами наблюдений. Поэтому максимум энтропии соответствует наилучшим оценкам для максимально неопределенных (в единицах энтропии) шумов.

Алгоритмы РМО можно разделить на два класса, в одном из которых реализуется принцип “жесткой” рандомизации, а в другом — “мягкой” рандомизации. Различие между ними связано с характером выполнения условий эмпирических балансов — точно или приближенно.

2.1. Под “жесткой” рандомизацией понимаются равенства между числовыми характеристиками ансамбля  $\mathcal{V}(j | P(\mathbf{a}), Q_j(\bar{\xi}^{(j)}))$  и выходными обучающими данными  $Y$ :

$$(6) \quad \mathbf{m}^{(k)} [j | P(\mathbf{a}), Q_j(\bar{\xi}^{(j)})] = \mathbf{y}(j), \quad j = \overline{1, s}.$$

Здесь  $\mathbf{m}^{(k)} [j | P(\mathbf{a}), Q_j(\bar{\xi}^{(j)})]$  — вектор  $k$ -средних с компонентами следующего вида:

$$(7) \quad m_i^{(k)} [j | P(\mathbf{a}), Q_j(\bar{\xi}^{(j)})] = \left[ \mathcal{M}_P \{ \hat{z}_i^k((j | \mathbf{a})) \} + \mathcal{M}_{Q_j} \{ \xi_i^k(j) \} \right]^{1/k},$$

где  $\mathcal{M}_P, \mathcal{M}_{Q_j}$  — операторы математического ожидания по ПРВ  $P, Q_j$  соответственно. Здесь для характеристики эмпирических балансов при “жесткой” рандомизации будет использоваться 1-средние, т.е.:

$$(8) \quad \mathbf{m}^{(1)} [j | P(\mathbf{a}), Q_j(\bar{\xi}^{(j)})] = \mathbf{y}(j), \quad j = \overline{1, s}.$$

где

$$(9) \quad m_i^{(1)}(j | P(\mathbf{a}), Q_j(\bar{\xi}^{(j)})) = \int_{\mathcal{A}} P(\mathbf{a}) \hat{z}_i(j | \mathbf{a}) d\mathbf{a} + \int_{\Xi_j} Q_j(\bar{\xi}^{(j)}) \xi_i(j) d\bar{\xi}^{(j)},$$

$$i = \overline{1, m}, \quad j = \overline{1, s}.$$

2.2. При “мягкой” рандомизации балансы между ансамблем  $\mathcal{V}(j | P(\mathbf{a}), Q_j(\bar{\xi}^{(j)}))$ , состоящим из случайных векторов  $\hat{\mathbf{v}}(j | \mathbf{a}, \bar{\xi}^{(j)})$  (4), и выходными обучающими данными —  $(s \times m)$ -матрицей  $Y = \{\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(s)}\}$  выполняются приближенно в терминах принятого критерия.

Для формирования подходящего критерия обратимся к равенству (4), определяющему случайные  $j$ -векторы, входящие в ансамбль  $\mathcal{V}$ . Близость векторов  $\hat{\mathbf{v}}(j | \mathbf{a})$  и  $\mathbf{y}^{(j)}$  будем оценивать какой-либо гельдеровской векторной нормой [26]:

$$(10) \quad \mathcal{N}^{(j)}(\mathbf{a}, \bar{\xi}^{(j)}, \mathbf{y}^{(j)}) = \|\mathbf{z}(j | \mathbf{a}) - \mathbf{y}^{(j)} + \bar{\xi}^{(j)}\|_H \leq \mathcal{N}_{z,y}^{(j)}(\mathbf{a}, \mathbf{y}^{(j)}) + \mathcal{N}_{\Xi}(\bar{\xi}^{(j)}),$$

где

$$(11) \quad \mathcal{N}_{z,y}^{(j)}(\mathbf{a}, \mathbf{y}^{(j)}) = \|\mathbf{z}(j | \mathbf{a}) - \mathbf{y}^{(j)}\|_H,$$

$$(12) \quad \mathcal{N}_{\xi}^{(j)}(\bar{\xi}^{(j)}) = \|\bar{\xi}^{(j)}\|_H.$$

Норма (11) характеризует различие между вектором выхода модели и вектором обучающих данных, а норма (12) — “мощность” шумов.

Заметим, что нормы (11) и (12) являются функциями случайных аргументов: параметров  $\mathbf{a}$  и шумов  $\bar{\xi}^{(j)}$ . Определим их математические ожидания:

$$(13) \quad \begin{aligned} \bar{\mathcal{N}}_{z,y}^{(j)}[P(\mathbf{a})] &= \int_{\mathcal{A}} P(\mathbf{a}) \mathcal{N}_{z,y}^{(j)}(\mathbf{a}, \mathbf{y}^{(j)}) d\mathbf{a}, \\ \bar{\mathcal{N}}_{\xi}^{(j)}[Q_j(\bar{\xi}^{(j)})] &= \int_{\Xi_j} Q_j(\bar{\xi}^{(j)}) \mathcal{N}_{\xi}^{(j)}(\bar{\xi}^{(j)}) d\bar{\xi}^{(j)}. \end{aligned}$$

Качество приближения эмпирических балансов и оценивания “мощности” шумов в случае мягкой рандомизации будем характеризовать функционалами (13).

### 3. РМО-алгоритмы с “жесткой” рандомизацией

“Жесткая” рандомизация предусматривает точное выполнение эмпирических балансов. Поэтому соответствующий ей РМО-А формулируется в следующем виде:

$$(14) \quad \begin{aligned} \mathcal{H}[P(\mathbf{a}), Q(\bar{\xi})] &= - \int_{\mathcal{A}} P(\mathbf{a}) \ln \frac{P(\mathbf{a})}{P^0(\mathbf{a})} d\mathbf{a} - \\ &- \sum_{j=1}^s \int_{\Xi_j} Q_j(\bar{\xi}^{(j)}) \ln \frac{Q_j(\bar{\xi}^{(j)})}{Q_j^0(\bar{\xi}^{(j)})} d\bar{\xi}^{(j)} \Rightarrow \max \end{aligned}$$

при условиях:

— нормировки функций ПРВ

$$(15) \quad \int_{\mathcal{A}} P(\mathbf{a}) d\mathbf{a} = 1, \quad \int_{\Xi_j} Q_j(\bar{\xi}^{(j)}) d\bar{\xi}^{(j)} = 1, \quad j = \overline{1, s};$$

— эмпирические балансы с использованием 1-средних

$$(16) \quad \mathbf{m}^{(1)}(j | P(\mathbf{a}), Q_j(\bar{\xi}^{(j)})) = \mathbf{y}(j), \quad j = \overline{1, s}.$$

Здесь  $P^0(\mathbf{a}), Q_j^0(\bar{\xi}^{(j)})$  — априорные ПРВ параметров и шумов; вектор  $\mathbf{m}^{(k)}$  имеет компоненты (9). Задача (14) и (16) является функциональной задачей энтропийно-линейного программирования ляпуновского типа [27–31] (все компоненты задачи описываются интегральными функционалами). Для задач этого типа условия оптимальности могут быть получены с использованием функционала и множителей Лагранжа. Поскольку функции ПРВ в РМО-А — непрерывно-дифференцируемые, то вариация функционала Лагранжа может быть определена с использованием производных Гато [32].

Введем следующие обозначения:

- переменная

$$(17) \quad w(P(\mathbf{a})) = 1 - \int_{\mathcal{A}} P(\mathbf{a}) d\mathbf{a};$$

- вектор  $\mathbf{q}(Q(\bar{\xi}))$  с компонентами

$$(18) \quad q_j(Q_j(\bar{\xi}^{(j)})) = 1 - \int_{\Xi_j} Q_j(\bar{\xi}^{(j)}) d\bar{\xi}^{(j)}, \quad j = \overline{1, s};$$

- векторы

$$(19) \quad \mathbf{e}^{(j)}(P(\mathbf{a}), Q(\bar{\xi})) = \mathbf{y}^{(j)} - \mathbf{m}^{(1)}(j | P(\mathbf{a}), Q_j(\bar{\xi}^{(j)})).$$

- множители Лагранжа  $\mu; \nu_1, \dots, \nu_s; \bar{\theta}^{(1)}, \dots, \bar{\theta}^{(s)}$ .

Здесь  $Q(\bar{\xi}) = [Q_1(\bar{\xi}^{(1)}), \dots, Q_s(\bar{\xi}^{(s)})]$ . Используя введенные обозначения, определим функционал Лагранжа

$$(20) \quad \begin{aligned} \mathcal{L}[P(\mathbb{A}), Q(\bar{\xi})] = & \mathcal{H}[P(\mathbf{a}), Q(\bar{\xi})] + \mu w(P(\mathbf{a})) + \langle \bar{\nu}, \mathbf{q}(Q(\bar{\xi})) \rangle + \\ & + \sum_{j=1}^s \langle \bar{\theta}^{(j)}, \mathbf{e}^{(j)}(P(\mathbf{a}), Q(\bar{\xi})) \rangle. \end{aligned}$$

В этом равенстве  $\langle \bullet, \bullet \rangle$  — скалярное произведение.

Условия стационарности функционала (20) в терминах производных Гато имеют вид:

$$(21) \quad \begin{aligned} \ln \frac{P^*(\mathbf{a})}{P^0(\mathbf{a})} + 1 + \mu + \sum_{j=1}^s \langle \bar{\theta}^{(j)}, \hat{\mathbf{e}}^{(j)} \rangle &= 0, \\ \ln \frac{Q_j^*(\bar{\xi}^{(j)})}{Q_j^0(\bar{\xi}^{(j)})} + 1 + \nu_j + \langle \bar{\theta}^{(j)}, \bar{\xi}^{(j)} \rangle &= 0, \\ j &= \overline{1, s}. \end{aligned}$$

Отсюда следуют выражения для энтропийно-оптимальных функций ПРВ параметров РПМ и шумов измерений:

$$(22) \quad \begin{aligned} P^*(\mathbf{a}) &= P^0(\mathbf{a}) \exp \left[ -1 - \mu - \sum_{j=1}^s \langle \bar{\theta}^{(j)}, \hat{\mathbf{z}}^{(j)} \rangle \right], \\ Q_j^*(\bar{\xi}^{(j)}) &= Q_j^0(\bar{\xi}^{(j)}) \exp \left[ -1 - \nu_j - \langle \bar{\theta}^{(j)}, \bar{\xi}^{(j)} \rangle \right], \\ j &= \overline{1, s}. \end{aligned}$$

Используя условия нормировки (17) и (18), можно исключить множители Лагранжа  $\mu$  и  $\bar{\nu}$ . Тогда выражения для энтропийно-оптимальных и нормированных функций ПРВ примут вид:

$$(23) \quad \begin{aligned} P^*(\mathbf{a}) &= \frac{P^0(\mathbf{a}) \exp \left[ - \sum_{j=1}^s \langle \bar{\theta}^{(j)}, \hat{\mathbf{z}}^{(j)} \rangle \right]}{\mathcal{P}(\bar{\theta})}, \\ Q_j^*(\bar{\xi}^{(j)}) &= \frac{Q_j^0(\bar{\xi}^{(j)}) \exp \left[ - \langle \bar{\theta}^{(j)}, \bar{\xi}^{(j)} \rangle \right]}{Q_j(\bar{\theta}^{(j)})}, \\ j &= \overline{1, s}. \end{aligned}$$

В этих равенствах

$$(24) \quad \begin{aligned} \mathcal{P}(\bar{\theta}) &= \int_{\mathcal{A}} P^0(\mathbf{a}) \exp \left[ - \sum_{j=1}^s \langle \bar{\theta}^{(j)}, \hat{\mathbf{z}}^{(j)} \rangle \right] d\mathbf{a}, \\ Q_j(\bar{\theta}^{(j)}) &= \int_{\Xi_j} Q_j^0(\bar{\xi}^{(j)}) \exp \left[ - \langle \bar{\theta}^{(j)}, \bar{\xi}^{(j)} \rangle \right] d\bar{\xi}^{(j)}, \quad j = \overline{1, s}. \end{aligned}$$

Здесь  $\bar{\theta} = \{\bar{\theta}^{(1)}, \dots, \bar{\theta}^{(s)}\}$ . Из (23) и (24) следует, что энтропийно-оптимальные функции ПРВ принадлежат экспоненциальному семейству, параметризованному множителями Лагранжа.

Значения указанных множителей Лагранжа определяются следующими уравнениями, возникающими из эмпирических балансов (16) при подстановке в них функций ПРВ из (23) и (24):

$$(25) \quad \begin{aligned} & \mathcal{P}^{-1}(\bar{\theta}) \int_{\mathcal{A}} P^0(\mathbf{a}) \exp \left[ - \sum_{j=1}^s \langle \bar{\theta}^{(j)}, \hat{\mathbf{z}}^{(j)} \rangle \right] d\mathbf{a} + \\ & + Q_j^{-1}(\bar{\theta}) \int_{\Xi_j} Q_j^0(\bar{\xi}^{(j)}) \exp \left[ - \langle \bar{\theta}^{(j)}, \bar{\xi}^{(j)} \rangle \right] d\bar{\xi}^{(j)} = \mathbf{y}^{(j)}, \\ & j = \overline{1, s}. \end{aligned}$$



Уравнения (25) образуют особый класс нелинейных уравнений, в которые входят интегральные компоненты, определяющие функциональные зависимости левых частей этих уравнений от множителей Лагранжа  $\bar{\theta}$ .

#### 4. РМО-алгоритмы с “мягкой” рандомизацией

При “мягкой” рандомизации эмпирические балансы выполняются приближенно в рамках условной минимизации синтетического функционала (5) и (13):

$$(26) \quad J[P(\mathbf{a}), Q(\bar{\xi})] = \mathcal{H}[P(\mathbf{a}), Q(\bar{\xi})] - \sum_{j=1}^s \bar{\mathcal{N}}_{z,y}^{(j)}(P(\mathbf{a})) - \sum_{j=1}^s \bar{\mathcal{N}}_{\xi}^{(j)}(Q_j(\bar{\xi}^{(j)})) \Rightarrow \max,$$

при условиях нормировки ПРВ

$$(27) \quad \int_{\mathcal{A}} P(\mathbf{a}) d\mathbf{a} = 1, \quad \int_{\Xi_j} Q_j(\bar{\xi}^{(j)}) d\bar{\xi}^{(j)} = 1, \quad j = \overline{1, s}.$$

Эта задача ляпуновского типа, и ее решение может быть найдено с помощью условий оптимальности в терминах производных Гато [32].

Будем иметь:

$$(28) \quad P^*(\mathbf{a}) = \frac{P^0(\mathbf{a}) \exp \left[ - \sum_{j=1}^s \left( \mathcal{N}_{z,y}^{(j)}(\mathbf{a}, \mathbf{y}^{(j)}) + \mathcal{N}_{\xi}^{(j)}(\bar{\xi}^{(j)}) \right) \right]}{\mathcal{P}},$$

$$Q_j^*(\bar{\xi}^{(j)}) = \frac{\exp \left[ - \mathcal{N}_{\xi}^{(j)}(\bar{\xi}^{(j)}) \right]}{Q_j}, \quad j = \overline{1, s},$$

где

$$(29) \quad \mathcal{P} = \int_{\mathcal{A}} P^*(\mathbf{a}) d\mathbf{a},$$

$$Q_j = \int_{\Xi_j} Q_j^*(\bar{\xi}^{(j)}) d\bar{\xi}^{(j)}, \quad j = \overline{1, s}.$$

Отсюда видно, что при “мягкой” рандомизации требуются вычислительные ресурсы только на вычисление многомерных интегралов (29).

## 5. Некоторые приложения рандомизированного машинного обучения

### 5.1. РМО с “жесткой” рандомизацией для бинарной вероятностной классификации текстов

Бинарная вероятностная классификация предполагает определение принадлежности объекта соответствующему классу с вероятностью, вычисляемой с учетом имеющейся обучающей коллекции.

Рассмотрим в качестве примера рандомизированный линейный бинарный классификатор. Пусть имеются две коллекции текстовых документов: обучающая  $\mathbb{E} = \{e_1, \dots, e_m\}$  и тестовая  $\mathbb{T} = \{t_1, \dots, t_s\}$ . В обучающей коллекции документы помечены принадлежностью либо к классу 1, либо к классу 2. В тестовой коллекции документы никак не помечены.

Документы в обеих коллекциях представлены векторами, компонентами которых являются веса ключевых слов (термов), встречающихся в документе. Тем самым формируются наборы векторов, размерность которых равна количеству  $n$  ключевых слов:

$$(30) \quad \begin{aligned} \mathcal{E} &= \{\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(m)}\}, & \mathbf{e}^{(j)} &\in \mathbb{R}^n, & j &= \overline{1, m}, \\ \mathcal{T} &= \{\mathbf{t}^{(1)}, \dots, \mathbf{t}^{(s)}\}, & \mathbf{t}^{(k)} &\in \mathbb{R}^n, & k &= \overline{1, s}. \end{aligned}$$

Здесь  $\mathbb{R}^n$  — пространство признаков, в данном случае — весов ключевых слов.

*1. Этап обучения.* Документы в обучающей коллекции помечены, т.е. она состоит из пар “номер документа, класс”:  $\mathcal{O} = \{(1, 1), (2, 1), \dots, (m, 1)\}$  длины  $m$ . Трансформируем ее в последовательность чисел из интервала  $[0, 1]$ , где номеру элемента в коллекции могут соответствовать числа из интервала  $[1/2, 1]$ , если документ принадлежит первому классу, и числа из интервала  $[0, 1/2)$ , если документ принадлежит второму классу. Таким образом, получаем “обучающий” вектор  $\mathbf{y}$  с компонентами, значения которых лежат в интервале  $[0, 1]$ , и размерностью, равной количеству  $m$  документов в коллекции, т.е.

$$(31) \quad \mathbf{y} = \{y^{(1)}, \dots, y^{(m)}\}.$$

Рандомизированная модель (решающее правило) представляется случайным вектором  $\hat{\mathbf{z}}(\mathbf{e}^{(j)}, \mathbf{a})$ , зависящим от случайных параметров  $\mathbf{a}$  однослойной нейронной сети

$$(32) \quad \hat{\mathbf{z}}^{(j)}(\mathbf{a}) = \text{sigm} \left( \langle \mathbf{e}^{(j)}, \mathbf{a} \rangle \right), \quad j = \overline{1, m}.$$

Здесь

$$(33) \quad \text{sigm}(x) = \frac{1}{1 + \exp[-\alpha(x - \Delta)]},$$

где параметры  $\alpha$  и  $\Delta$  фиксированы. Аргумент этой функции случайный, так как параметры  $\mathbf{a}$  рандомизированной модели случайные. Значения функции

$\text{sigm}(x)$  на интервале  $[1/2, 1]$  соответствуют первому классу, и значения в открытом интервале  $[0, 1/2)$  — второму классу.

В рандомизированной модели (32) параметры  $\mathbf{a} = \{a_1, \dots, a_n\}$  — независимые, интервального типа:

$$(34) \quad \begin{aligned} a_k &\in \mathcal{A}_k = [a^-_k, a^+_k], & k = \overline{1, n}, \\ \mathcal{A} &= \bigotimes_{k=1}^n \mathcal{A}_k. \end{aligned}$$

На этих интервальных множествах существует функция ПРВ  $P(\mathbf{a})$ . Поэтому для каждого документа с номером  $j$  имеет место ансамбль  $\hat{\mathbf{z}}^{(j)}$  случайных чисел из интервала  $(0, 1)$  (32) и (53). Средние значения компонент будут иметь следующий вид

$$(35) \quad \mathcal{M}\{\hat{\mathbf{z}}^j(\mathbf{a})\} = \int_{\mathcal{A}} P(\mathbf{a}) \text{sigm}(\langle \mathbf{e}^{(j)}, \mathbf{a} \rangle) d\mathbf{a}.$$

Итак, РМО-А классификации с “жесткой” рандомизацией представляются в следующем виде

$$(36) \quad \mathcal{H}[P(\mathbf{a})] = - \int_{\mathcal{A}} P(\mathbf{a}) \ln P(\mathbf{a}) d\mathbf{a} \Rightarrow \max$$

при условиях:

– нормировки

$$(37) \quad \int_{\mathcal{A}} P(\mathbf{a}) d\mathbf{a} = 1,$$

– эмпирических балансов

$$(38) \quad \int_{\mathcal{A}} P(\mathbf{a}) \text{sigm}(\langle \mathbf{e}^{(j)}, \mathbf{a} \rangle) d\mathbf{a} = y^{(j)}, \quad j = \overline{1, m}.$$

Введем множители Лагранжа  $\bar{\theta} = \{\theta_1, \dots, \theta_m\}$  для ограничений (38). Тогда решение этой задачи, согласно (22) и адаптированное к модели (53), с точностью до множителей Лагранжа  $\theta_1, \dots, \theta_m$  имеет вид

$$(39) \quad P^*(\mathbf{a}) = \frac{W^*(\mathbf{a})}{\mathcal{P}(\bar{\theta})},$$

где

$$(40) \quad W^*(\mathbf{a}) = \exp(-\langle \bar{\theta}, \hat{\mathbf{z}}(\mathbf{a}) \rangle),$$

$$(41) \quad \mathcal{P}(\bar{\theta}) = \int_{\mathcal{A}} \exp[-\langle \bar{\theta}, \hat{\mathbf{z}}(\mathbf{a}) \rangle] d\mathbf{a}.$$

Таблица 1

$j$	$e_1^{(j)}$	$e_2^{(j)}$	$e_3^{(j)}$	$e_4^{(j)}$
1	0,11	0,75	0,08	0,21
2	0,91	0,65	0,11	0,81
3	0,57	0,17	0,31	0,91

Множители Лагранжа  $\bar{\theta}$  определяются системой балансовых уравнений (38) следующего вида

$$(42) \quad \mathcal{P}^{-1}(\bar{\theta}) \int_{\mathcal{A}} \exp[-\langle \bar{\theta}, \hat{\mathbf{z}}(\mathbf{a}) \rangle] \hat{z}^{(j)}(\mathbf{a}) d\mathbf{a} = y^{(j)}, \quad j = \overline{1, m}.$$

Заметим, что размерность этой системы равна количеству документов в обучающей коллекции. Поэтому с точки зрения вычислительной трудоемкости рандомизированная процедура решения задачи классификации более эффективна для ограниченных объемов обучающей коллекции. Рассмотрим два примера, которые демонстрируют основную идею РМО-А с “жесткой” рандомизацией для бинарной классификации текстов.

*Пример 1 – Обучение.* Размерность алгоритма равна 4, обучающая коллекция состоит из трех документов, каждый из которых описывается четырьмя весами, значения которых показаны в табл. 1.

Рандомизированная модель (32) и (53) имеет параметры:  $\alpha = 1,0$  и  $\Delta = 0$ . Ответы  $\mathbf{y} = \{0,18; 0,81; 0,43\}$  ( $y_i < 0,5$  соответствует классу 2,  $y_i \geq 0,5$  соответствует классу 1). Множители Лагранжа для энтропийно-оптимальной ПРВ (39) и (40) имеют следующие значения:  $\bar{\theta} = \{0,2524; 1,7678; 1,6563\}$ . Параметры  $a_i \in [-10, 10]$ ,  $i = \overline{1, 4}$ . Энтропийно-оптимальная для данной обучающей коллекции функция  $W^*(\mathbf{a})$  (39) имеет вид

$$(43) \quad W^*(\mathbf{a}) = \exp\left(-\sum_{i=1}^3 \theta_i z_i(\mathbf{a})\right),$$

$$z_i(\mathbf{a}) = \left(1 + \exp\left(-\sum_{k=1}^4 e^{(j)}_{k, a_k}\right)\right)^{(-1)}.$$

На рис. 2 показано двумерное сечение функции (39) при  $a_3 = 0,5$ ;  $a_4 = 0,5$ .

*2. Этап тестирования.* На этом этапе используется коллекция  $\mathbb{T} = \{t_1, \dots, t_s\}$ , где каждый элемент коллекции характеризуется вектором  $\mathbf{t}^{(j)} \in \mathbb{R}^n$ . Рассмотрим процедуру классификации для произвольного документа  $\mathbf{t}^{(j)}$ .

*Шаг 1.* Генерируется ансамбль  $\hat{\mathbf{z}}^{(i)}$  выхода рандомизированной модели (решающих правил) (32) и (53) с функцией  $P^*(\mathbf{a})$  (39). Ансамбль содержит  $N$  случайных чисел из интервала  $[0, 1]$ .

*Шаг 2.* Если случайное число из этого ансамбля больше  $1/2$ , то документ  $\mathbf{t}^{(i)}$  относится к классу 1. Если — меньше  $1/2$ , то — к классу 2.

*Шаг 3.* Пусть  $N_1$  чисел оказались в классе 1 и  $N_2$  — в классе 2. Поскольку число испытаний  $N$  достаточно велико, то можно считать, что величины

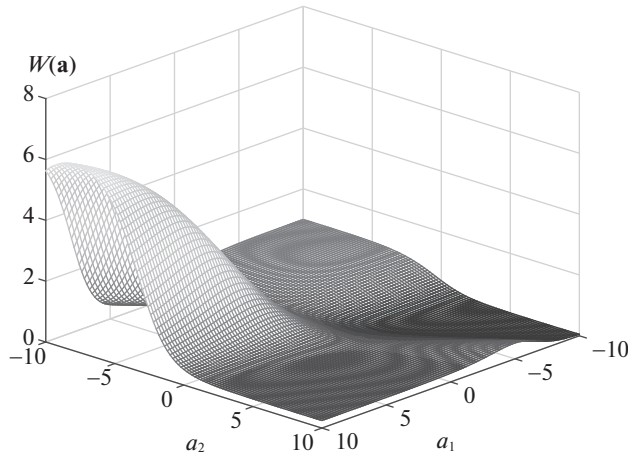


Рис. 2. Двумерное сечение функции ПРВ для  $a_3 = 0,5$ ,  $a_4 = 0,5$ ,  $\theta = \{3,2807, -3,5127, 1,6373\}$ .

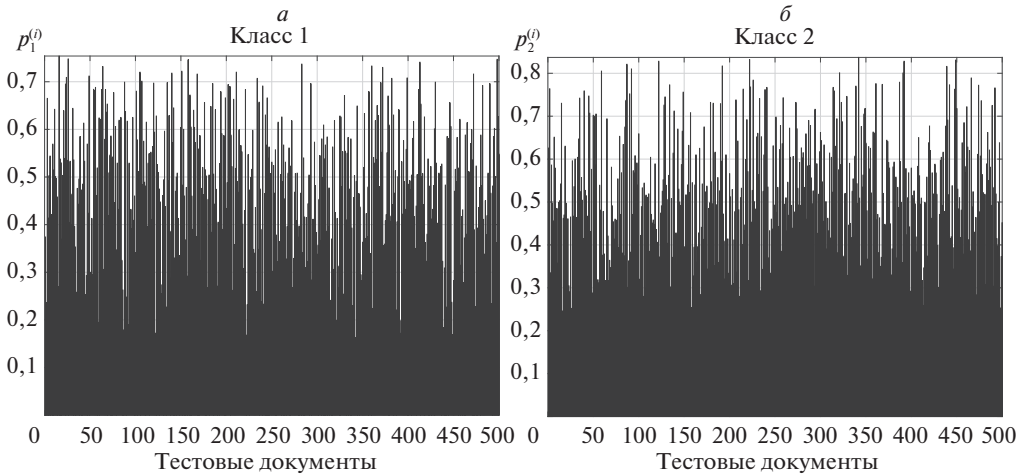


Рис. 3. Эмпирические вероятности принадлежности классам.

$p_1^{(i)} = N_1/N$  и  $p_2^{(i)} = N_2/N$  есть эмпирические вероятности попадания документа  $\mathbf{t}^{(i)}$  в соответствующий класс. Повторяя операции шагов 2, 3 для всей коллекции  $\mathbb{T}$ , получим распределения вероятностей попадания документов в класс 1 или 2.

*Пример 1 – Тестирование.* Генерируется массив  $(500 \times 4)$  четырехмерных случайных векторов  $\mathbf{t}^{(i)}$ ,  $i = \overline{1, 500}$ , с независимыми компонентами, равномерно распределенными в интервалах  $[0, 1]$ . Для каждого элемента этой выборки генерируются случайные параметры модели (32) и (53) в соответствии с ПРВ (39) и (40) ( $N = 1000$ ) методом исключения [33] и вычисляется ее выход. Выполняются шаги 2, 3.

На рис. 3а–3б показаны эмпирические вероятности  $p_1^{(i)}, p_2^{(i)}$  принадлежности  $t_i$ -документа классу 1 и 2.

5.2. РМО с “мягкой” рандомизацией для прогнозирования миграционного взаимодействия региональных систем

Взаимовлияние миграционных процессов в региональных системах — проблема, актуальность которой возрастает в современном мире. Неоднородность социально-экономического статуса регионов возрастает вместе с ростом политической и военной напряженности. Все это приводит к резкому перераспределению миграционных потоков и, как следствие, к изменению численности регионального населения и росту затрат на его обеспечение. Поэтому важными оказываются инструментальные средства (математические модели, алгоритмы, программное обеспечение), позволяющие в режиме адаптации к динамике миграции прогнозировать ее распределение с учетом имеющихся обеспечивающих ресурсов.

Рассмотрим динамическую модель миграционного взаимодействия с ограничением на общий ресурс и дискретным временем, изложенную в [34]. Она состоит из двух блоков. Первый блок моделирует миграционные потоки внутри системы  $\mathcal{S}_1$  и описывается уравнением динамической регрессии:

$$(44) \quad \mathbf{K}[(s+1)h] = (A - E)\mathbf{K}[sh] + \mathbf{F}(z[sh]), \quad (\mathbf{K}, \mathbf{F}) \in R^N, \quad s = \overline{0, K-1},$$

где  $A$  — матрица пространственных передвижек,  $E$  — диагональная матрица мобильности,  $\mathbf{K}[sh]$  — распределение населения в региональной системе  $\mathcal{S}_1$  в момент времени  $sh$ .

Локально-стационарное распределение в момент времени  $sh$  иммиграционных потоков из региональной системы  $\mathcal{S}_2$  в систему  $\mathcal{S}_1$ , характеризуемое энтропийным оператором, моделирует второй блок, допускающий описание вектор-функцией  $\mathbf{F}(z[sh])$  с компонентами

$$(45) \quad f_n[sh] = h \sum_{j=1}^M b_{jn}(z[sh])^{c_{jn}}, \quad n = \overline{1, N}, \quad s = \overline{0, K-1}.$$

Переменная  $z$  — экспоненциальный множитель Лагранжа в задаче энтропийно-оптимального распределения иммиграционных потоков определяется решением уравнения

$$(46) \quad \sum_{k=1}^M \sum_{n=1}^N c_{kn} b_{kn}(z[sh])^{c_{kn}} = T[sh],$$

где  $T[sh]$  — объем общего для регионов системы  $\mathcal{S}_1$  ресурса, который используется для поддержки иммигрантов.

Входными данными для модели являются объемы  $T[0], T[h], \dots, T[(K-1)h]$  и выходными — распределения регионального населения  $\mathbf{K}[0], \mathbf{K}[h], \dots, \mathbf{K}[(K-1)h]$  в системе  $\mathcal{S}_1$ .

Рассмотрим применение “мягкой” рандомизации для оценивания и прогнозирования иммиграционных потоков из Сирии (1) и Ливии (2) (система  $\mathcal{S}_2$ ) в Германию (1), Францию (2) и Италию (3) (система  $\mathcal{S}_1$ ).

Таблица 2

$m_n$	0,43	0,50	0,40
$h_{1n}$	0	0,3	0,3
$h_{2n}$	0,3	0	0,3
$h_{3n}$	0,5	0,4	0
$b_{1n}$	0,4	0,3	0,3
$b_{2n}$	0,3	0,1	0,4
$c_{1n}$	0,4	0,4	0,3
$c_{2n}$	0,4	0,4	0,3

1. Рандомизированная модель, параметры, ошибки измерений, временные интервалы и коллекции реальных данных. Воспользуемся рандомизированной математической моделью (44)–(46), применив в ней нормализованные переменные

$$(47) \quad p_n[sh] = \frac{K_n[sh]}{K_{\max}}, \quad n = \overline{1, 3}.$$

Получим:

$$(48) \quad p_n[(s+1)h] = (1 - b_1 m_n) p_n[sh] + h b_1 b_2 \sum_{i=1, i \neq n}^3 m_i h_{in} p_i[sh] + h f_n[sh],$$

$$f_n[sh] = \sum_{i=1}^M b_{in} b_3^{c_{in}}, \quad n = \overline{1, 3},$$

$$T[sh] = \sum_{n=1}^3 \sum_{i=1}^2 c_{in} b_{in} b_3^{c_{in}}.$$

В этих уравнениях переменные состояния системы  $\mathcal{S}_1$  и иммиграционные потоки из системы  $\mathcal{S}_2$  нормализованы, т.е.

$$(49) \quad 0 \leq p_n[sh] \leq 1, \quad 0 \leq f_n[sh] \leq 1, \quad n = \overline{1, N}.$$

Переменная  $z^*$  есть характеристика энтропийного оператора иммиграционного процесса, которая определяется решением последнего уравнения в (48). Значения параметров  $m_i$ ,  $h_{in}$ ,  $b_{in}$  и  $c_{in}$  указаны в табл. 2.

Согласно этой таблице  $m_{\max} = 0,5$ ;  $h_{\max} = 0,5$ ;  $b_{\min} = 0,3$ ;  $b_{\max} = 0,4$ ;  $c_{\max} = c_{\min} = c = 0,5$ . Ошибки измерений численности (в нормализованных единицах)  $\bar{\xi}[sh] \in \mathbb{R}^3$  интервального типа

$$(50) \quad \bar{\xi}[sh] \in \Xi = [\bar{\xi}_-, \bar{\xi}_+], \quad \xi_n^\pm = 0,01,$$

причем границы интервалов предполагаются одинаковыми для моментов времени  $sh$ .

Нормализованный наблюдаемый выход модели имеет вид

$$(51) \quad \mathbf{v}[sh] = \mathbf{p}[sh] + \bar{\xi}[sh].$$

**Таблица 3.** Входные и выходные коллекции данных

год	2009	2010	2011	2012	2013
$s$	0	1	2	3	4
$Y_1[s]$	81,90	81,77	80,27	80,42	80,64
$y_1[s]$	1,00	0,998	0,980	0,982	0,985
$Y_2[s]$	62,47	62,80	63,11	63,41	63,70
$y_2[s]$	0,762	0,767	0,771	0,774	0,778
$Y_3[s]$	59,39	59,53	59,63	59,71	59,75
$y_3[s]$	0,725	0,727	0,728	0,729	0,726
$T[s]$ (млрд)	0,093	0,094	0,095	0,096	0,097

Модель (48) со случайными параметрами будет использоваться для оценивания их характеристик и тестирования на соответствующих временных интервалах с шагом  $h = 1$  год:

- интервал оценивания  $T_{est} = 2009 - 2013$  годы;
- интервал тестирования  $T_{tst} = 2014 - 2018$  годы.

2. Энтропийное оценивание ПРВ параметров и шумов (интервал  $T_{est}$ ). Для решения этой задачи используются данные о распределении населения в Германии  $n = 1$ , Франции  $n = 2$  и Италии  $n = 3$  и об общих затратах на обустройство иммигрантов в интервале оценивания, которые приведены в табл. 3, (UNDATA a world information, data.un.org).

В этой модели параметры  $b_1, b_2, b_3$  — случайные со значениями из интервалов:

$$(52) \quad b_1 \in \mathcal{B}_1 = [1,0; 2,5]; \quad b_2 \in \mathcal{B}_2 = [0,5; 1,8], \quad b_3 \in \mathcal{B}_3 = [0,3; 1,5].$$

Согласно (4)

$$(53) \quad U_1 = 0,5; \quad U_2 = 00,5; \quad U_3 = 1,2, \quad U_4 = 0,986.$$

Тогда метод “мягкого” РМО дает следующие оптимальные ПРВ параметров и шумов:

$$(54) \quad W^*(\mathbf{b}) = \frac{\exp\left(-0,5b_1 - 0,5b_1b_2 - 1,2b_3^{0,5} - 0,986\right)}{\mathcal{W}},$$

$$Q^*(\bar{\xi}) = \frac{\exp\left(-\sum_{n=1}^3 \xi_n^2\right)}{\mathcal{Q}},$$

где

$$(55) \quad \mathcal{W} = \int_{\mathcal{B}_1} \int_{\mathcal{B}_2} \int_{\mathcal{B}_3} \exp\left(-0,5b_1 - 0,5b_1b_2 - 1,2b_3^{0,5} - 0,986\right) db_1 db_2 db_3,$$

$$\mathcal{Q} = \prod_{n=1}^3 \int_{-0,01}^{0,01} \exp(-\xi^2) d\xi.$$



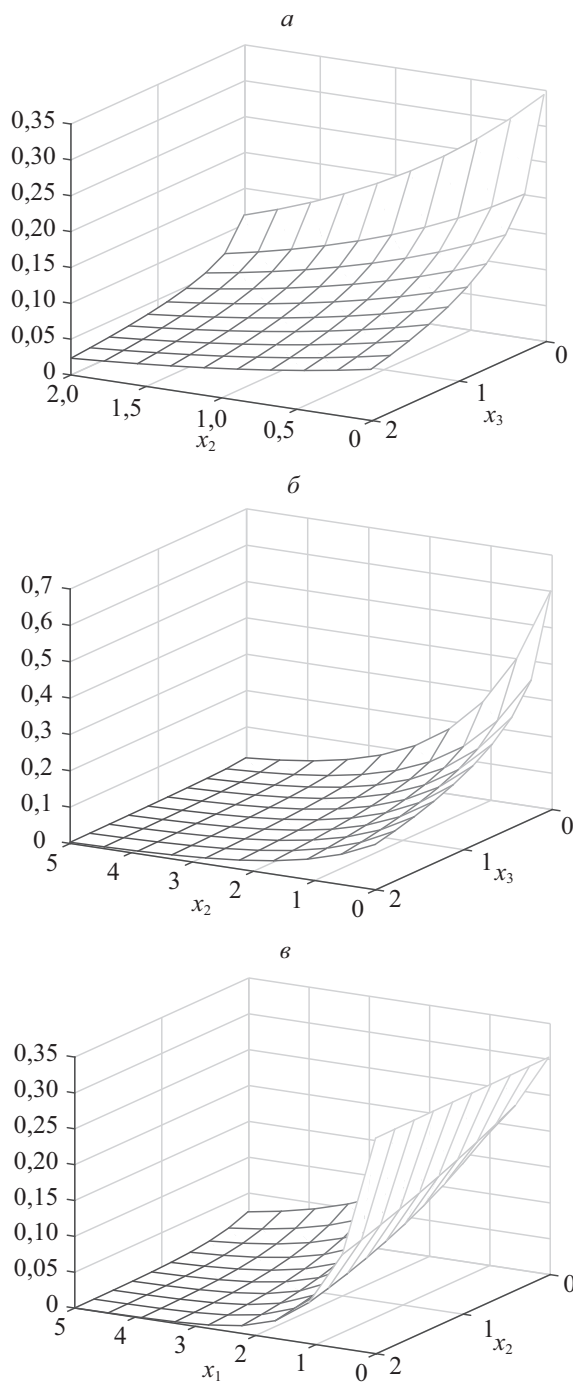


Рис. 4. *a* — Двумерное сечение функции ПРВ параметров  $W^*(x_1, x_2, x_3)$ ,  $x_1 = 1,1189$ ,  $I = 0,54562$ . *б* — Двумерное сечение функции ПРВ параметров  $W^*(x_1, x_2, x_3)$ ,  $x_2 = 0,45222$ ,  $I = 0,54562$ . *в* — Двумерное сечение функции ПРВ параметров  $W^*(x_1, x_2, x_3)$ ,  $x_3 = 0,45222$ ,  $I = 0,54562$ .

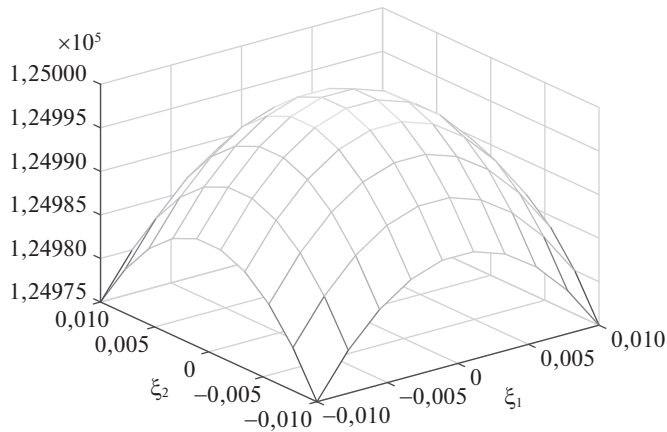


Рис. 5. Двумерное сечение функции ПРВ шума  $Q^*(\xi_1, \xi_2, \xi_3)$ ,  $\xi_3 = 0,01$ ,  $I = 7,9992E - 06$ .

На рис. 4,а–4,в приведены графики двумерных сечений 3-мерных функции ПРВ параметров, и на рис. 5 — функции ПРВ шума.

3. Тестирование модели. Для тестирования рандомизированной модели (50) с оптимальными ПРВ (54) и (55) используются тестовые данные о численности населения в исследуемых странах из UNDATA a world information (data.un.org), приведенные в табл. 4. Там же приведены результаты тестирования в терминах средних по ансамблю траекторий  $\bar{p}_1[sh]$ ,  $\bar{p}_2[sh]$ ,  $\bar{p}_3[sh]$ .

Тестирование производится путем сэмлирования значений рандомизированных интервальных параметров, имеющих функции ПРВ (54) и (55), и построения соответствующих траекторий согласно уравнениям (50). На рис. 6,а–6,в показаны ансамбли таких траекторий  $v_1[sh]$ ,  $v_2[sh]$ ,  $v[sh]$ . На этих же рисунках нанесены средние  $\bar{v}_1[sh]$ ,  $\bar{v}_2[sh]$ ,  $\bar{v}_3[sh]$  по ансамблям траектории (штриховая линия) и траектории  $y_1[sh]$ ,  $y_2[sh]$ ,  $y_3[sh]$  реальных изменений региональных численностей (сплошная линия), а также указаны границы дисперсионных трубок  $\bar{p}_1^*[sh] \pm \sigma_1$ ,  $\bar{p}_2^*[sh] \pm \sigma_2$ ,  $\bar{p}_3^*[sh] \pm \sigma_3$  (пунктирная линия).

**Таблица 4.** Входные и выходные коллекции данных

год	2014	2015	2016	2017	2018
$s$	0	1	2	3	4
$Y_1[s]$	81,489	81,707	82,063	82,386	82,674
$y_1[s]$	0,985	0,988	0,993	0,996	1,000
$\bar{p}_1[sh]$	0,986	0,615	0,743	0,639	0,999
$Y_2[s]$	64,190	64,457	64,791	65,134	65,484
$y_2[s]$	0,721	0,472	0,564	0,529	0,708
$\bar{p}_2[sh]$	0,722	0,695	0,707	0,691	0,715
$Y_3[s]$	59,585	59,504	59,504	59,509	59,516
$y_3[s]$	0,775	0,609	0,562	0,699	0,650
$\bar{p}_3[sh]$	0,776	0,617	0,607	0,705	0,628
$T[s]$ (млрд)	0,097	0,097	0,097	0,098	0,098

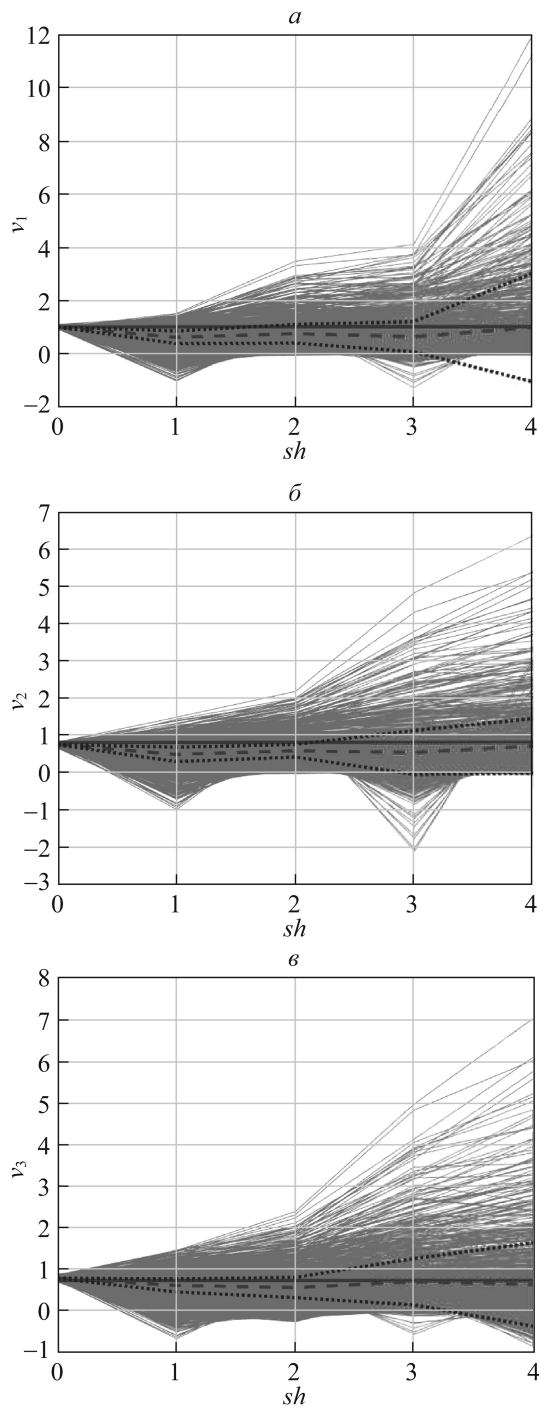


Рис. 6. *a* — Ансамбль траекторий  $\bar{v}_1[4] = 0,96678$ . *б* — Ансамбль траекторий  $\bar{v}_2[4] = 0,69204$ . *в* — Ансамбль траекторий  $\bar{v}_3[4] = 0,62205$ .

Погрешность при тестировании оценивалась относительной среднеквадратичной ошибкой

$$(56) \quad \delta_n = \frac{\sqrt{\sum_{s=0}^4 (\bar{p}_n[sh] - y_n[sh])^2}}{\sqrt{\sum_{s=0}^4 (\bar{p}_n[sh])^2 + \sum_{s=0}^4 (y_n[sh])^2}}.$$

В данном примере она составила: по региону 1 — 4,6 %, по региону 2 — 3,5 %, по региону 3 — 2,6 %.

## 6. Заключение

Рандомизированное машинное обучение является эффективным инструментом решения задач оценивания, классификации, распознавания и прогнозирования в условиях неопределенности. Математическое описание алгоритмов РМО формализуется в терминах функциональных задач оптимизации. Рассмотрены задачи с “жесткой” и “мягкой” рандомизацией и получены соответствующие алгоритмы их решения. Приведены примеры задач классификации текстов и оценивания (тестирования) моделей динамической регрессии.

## СПИСОК ЛИТЕРАТУРЫ

1. *Piatetsky-Shapiro G., Frawley W.* Knowledge Discovery in Databases. AAAI/MIT Press, 1991.
2. *Witten I.H., Frank E.* Data Mining: Practical Learning Tools and Techniques (2nd Ed.). Morgan Kaufmann, 2005.
3. Editorial Community cleverness required // Nature. 2008. V. 455. No. 1.
4. *Rosenblatt M.* The Perceptron — perceiving and recognizing automaton. Report 85-460-1, 1957.
5. *Цыпкин Я.З.* Основы теории обучающихся систем. М.: Наука, 1970.
6. *Айзерман М.А., Браверман Э.М., Розоноэр Л.И.* Метод потенциальных функций в теории обучения машин. М.: Наука, 1970.
7. *Вапник В.Н., Червоненкис А.Я.* Теория распознавания образов. М.: Наука, 1974.
8. *Вапник В.Н., Червоненкис А.Я.* Восстановление зависимостей по эмпирическим данным. М.: Наука, 1979.
9. *Bishop C.M.* Pattern Recognition and Machine Learning. Serise: Information Theory and Statistics. Springer. 2006.
10. *Dempster A.P., Laird N.M., Rubin D.B.* Maximum Likelihood from incomplete data via the EM algorithm // J. Royal Statistical Society. Ser. B. 1977. No. 34. P. 1–38.
11. *Загоруйко Н.Г.* Прикладные методы анализа данных и знаний. Новосибирск: Наука, 1998.
12. *Jain A., Murty M., Flunn P.* Data clustering: A review. ASM Computing Surveys, 1999. V. 31. No. 3. P. 264–323.
13. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning. Springer, 2001. <http://www-stat.stanford.edu/tibs/ElemStatLearn>
14. *Воронцов К.В.* Математические методы обучения по прецедентам. Курс лекций МФТИ, 2006.

15. Мерков А.Б. Распознавание образов. Введение в методы статистического обучения. М.: Едиториал УРСС, 2010.
16. Золотых Н.Ю. Машинное обучение и анализ данных. 2013.  
<http://www.uic.unn.ru/zny/ml>
17. Флаш П. Машинное обучение. М.: ДМК Пресс, 2015.
18. Abellan J., Castellano J.G. Improving the Naive Bayes Classifier via a Quick Variable Selection Method Using Maximum of Entropy // Entropy. 2017. V. 19. No. 6. P. 246–254.
19. Kullback S., Leibler R.A. On Information and Sufficiency // Ann. Math. Stat. 1951. V. 22 (1). P. 79–86.
20. Kapur J.N. Maximum entropy models in science and engineering. John Wiley & Sons, Inc., 1989.
21. Jaynes E.T. Information Theory and Statistical Mechanics // Physics Review Notes. 1957. V. 106. P. 620–630.
22. The maximum entropy formalism / Eds. R.D. Levin, M. Tribus. MIT Press, 1979.
23. Jaynes E.T. Papers on Probability, Statistics and Statistical Physics. Dordrecht: Kluwer Acad. Publisher, 1989.
24. Jaynes E.T. Probability Theory. The logic and science. Cambridge Univ. Press, 2003.
25. Racine J., Maasoumi E. A Versatile and Robust Metric Entropy Test of Time-Reversibility, and Other Hypotheses // J. Econometrics. 2007. V. 138. P. 547–567.
26. Воеводин В.В., Кузнецов Ю.А. Матрицы и вычисления. М.: Наука, 1984.
27. Kaashoek M.A., Seatzu S., van der Mee C. Recent Advances in Operator Theory and its Applications. Springer, 2006.
28. Иоффе А.Д., Тихомиров В.М. Теория экстремальных задач. М.: Наука, 1974.
29. Алексеев В.М., Тихомиров В.М., Фомин С.В. Оптимальное управление. М.: Наука, 1979.
30. Darkhovskii B.S., Popkov Y.S., Popkov A.Y. Monte Carlo Method of Batch Iterations: Probabilistic Characteristics // Autom. Remote Control. 2015. V. 76. No. 5. P. 776–785.  
*Дарховский Б.С., Попков Ю.С., Попков А.Ю. Метод пакетных итераций Монте-Карло: вероятностные характеристики // АИТ. 2015. № 5. С. 60–71.*
31. Попков Ю.С., Попков А.Ю., Дарховский Б.С. Параллельный Монте-Карло для построения энтропийно-робастных оценок // Математическое моделирование. 2015. Т. 27. № 6. С. 14–32.
32. Kolmogorov F.N., Fomin S.V. Elements of the Theory of Functions and Functional Analysis. Duver Books of Mathematics, 1999.
33. Rubinstein R.Y., Kroese D.P. Simulation and the Monte Carlo Method, John Wiley & Sons, 2008.
34. Попков Ю.С. Динамическая модель миграционного взаимодействия региональных систем с энтропийным оператором // Тр. ИСА РАН. 2018. Т. 68. № 3. С. 3–11. <https://doi.org/10.14357/20790279180301>

*Статья представлена к публикации членом редколлегии А.В. Назинным.*

Поступила в редакцию 06.06.2018

После доработки 13.09.2018

Принята к публикации 08.11.2018