ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА
на диссертацию Перминова Андрея Игоревича
«Доверенный байесовский классификатор для данных малой размерности на основе многослойного персептрона», представленную на соискание учёной степени кандидата физико-математических наук по специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»

Актуальность темы. Стандартные современные методы предсказательного моделирования (в частности, нейросетевые или на основе градиентного бустинга) позволяют добиться высокой средней точности предсказаний, однако не содержат внутренних механизмов оценки относительной надежности этих предсказаний для конкретных входных объектов. В частности, в случае применения классификатора к объекту, сильно отличающемуся от элементов обучающей выборки, классификатор может уверенно предсказывать неверный класс. Родственные трудности могут возникать и в случае несбалансированной классификации. Преодоление этих недостатков является важной темой современных исследований и ключевой темой данной диссертации. Развиваемый автором подход позволяет модели оценивать пределы своей компетенции и в случае необходимости отказываться от принятия решения. В этом же контексте автор анализирует защищенность модели от состязательных атак и предлагает способы статистической интерпретации предсказаний модели и коррекции дисбаланса классов.

Структура диссертации. Основная часть диссертации состоит из 5 глав.

Глава 1 представляет собой введение в проблематику доверенной классификации. В главе дается обзор существующих методов классификации и подходов к оценке неопределенности, решению проблемы обнаружения выхода за границы распределения и обработки дисбаланса классов.

В **Главе 2** предлагается и развивается метод придания модели представления об области своей применимости путем включения в обучающую выборку “фоновых” данных, выходная компонента которых позволяет детектировать выход за границы этой области. В качестве базового классификатора рассматривается полносвязная многослойная нейронная сеть с кусочно-линейной функцией активации. Пространство входных векторов при этом разбивается на многогранники, на которых предсказания сети линейны. Каждый такой многогранник задается указанием одной из линейных компонент кусочно-линейной функции активации в каждом нейроне. Обсуждается, что с многогранниками разбиения можно связывать гистограммные статистики обучающих данных; предлагается иерархическая организация многогранников в виде объясняющего дерева eXBTree. Демонстрируется более высокая защищенность модифицированных моделей по отношению к состязательным атакам.

В **Главе 3** идея “фоновых” данных развивается далее с целью применения к несбалансированным задачам, для которых предлагается отдельно обучать “унарные” классификаторы, отличающие отдельные классы от фона. Унарные классификаторы затем агрегируются в двух- или многоклассовые. Для полноценной оцен-

ки работы этой схемы предлагается несколько новых метрик классификации (мощность, эффективность, неразделимость классов).

В **Главе 4** предложенные унарные классификаторы применяются для генерации синтетических табличных данных с распределением, заданным некоторой обучающей выборкой. Для этого по заданной выборке и добавленной фоновой выборке сначала обучается унарный классификатор. Желаемые синтетические данные затем получаются из случайных равномерно распределенных точек путем пороговой фильтрации по значениям этого классификатора.

В **Главе 5** описывается программная реализация предлагаемых методов. Разработанная система представляет собой автономное клиентское web-приложение на JavaScript. Приложение позволяет настраивать параметры архитектуры нейронной сети (размеры и число слоев, функции активации), параметры обучения (функции потерь, оптимизаторы и др.), пошагово визуализировать процесс обучения и результаты работы моделей. Приложение имеет модульную архитектуру (модули вычислительного ядра, управления данными, системы визуализации и экспериментов). Для ускорения вычислений применяется разворачивание циклов и другие средства оптимизации. Сравнительное тестирование реализованных алгоритмов с референсными методами в PyTorch показало согласие в пределах машинной точности.

Теоретическая и практическая значимость результатов, их новизна. Работа содержит ряд интересных идей, связанных с внедрением в модель механизма оценки своей уверенности на основе включения в обучающие данные "фоновых" элементов. Эти идеи последовательно развиваются, иллюстрируются и проверяются в главах 2 – 4: сначала в контексте отказа от предсказания на объектах, далеких от обучающей выборки, затем для коррекции несбалансированной классификации с помощью отдельных унарных классификаторов, и затем для генерации синтетических данных с заданным распределением. Предлагаемый метод в основном ориентирован на задачи небольшой размерности ($\lesssim 10$ входных переменных). Хотя работа включает сравнительно небольшой объем эмпирических тестов, из результатов можно сделать вывод о сопоставимой эффективности предлагаемых методов и существующих альтернатив. Работа также включает несколько теоретических результатов, уточняющих структуру модифицированных классификаторов и дающих асимптотические оценки их сложности и точности.

Отдельно отметим, что сильной стороной работы является открытая и удобная, автономная и кроссплатформенная программная реализация разработанных методов в виде web-приложения на JavaScript. Она включает наглядную демонстрацию обучения и работы моделей, обеспечивает широкую доступность и воспроизводимость экспериментов. Представляется, что потенциально она может иметь существенную педагогическую ценность.

Замечания к содержанию и оформлению диссертации. В целом текст диссертации логичный и понятный. Имеется несколько замечаний.

1. Изложение материала, посвященного объясняющим деревьям eXBTree в разделе 2.5, недостаточно ясное. Нет четкого определения этих деревьев; не указано, в каком порядке обходятся отвечающие нейронам узлы дерева. Не поясняется,

что в формулировке теоремы 2 понимается под “временной сложностью алгоритма построения полного объясняющего двочного дерева”. Из доказательства можно понять, что речь идет просто о размере дерева. Однако сложность построения дерева не определяется, вообще говоря, лишь его размером – скажем, ветвления в узлах отвечают возможным пересечениям многогранников гиперплоскостями; естественно было бы ожидать, что сложность определения этих пересечений также включается в общую сложность построения дерева.

2. Формула атакующего вектора x^* на с. 55 содержит опечатку: вместо b^T должно быть $y_t - b$.

3. Пример нахождения значения метрик качества унарных классификаторов в разделе 3.6.4 недостаточно ясный. Непонятно, например, почему вложенность классов влияет на мощность классификаторов – согласно определению из раздела 3.6.1 она определяется для разных классов независимо друг от друга. Также, по своему определению, метрики напрямую зависят от пороговых значений β, однако в примере они вообще не упоминаются.

Указанные замечания не снижают начную и практическую ценность проведенных исследований.

Заключение. Диссертация Перминова Андрея Игоревича «Доверенный байесовский классификатор для данных малой размерности на основе многослойного персептрона» является самостоятельным и завершенным исследованием, обладающим научной и практической значимостью. Работа отвечает требованиям ВАК о порядке присуждения ученых степеней к кандидатским диссертациям, а соискатель заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Официальный оппонент
Яроцкий Дмитрий Александрович,
доктор физико-математических наук, профессор
Автономной некоммерческой образовательной организации высшего пефессионального образования «Сколковский институт науки и технологий»

30 марта 2026 г.