

## ОТЗЫВ

члена диссертационного совета на диссертацию Боярова Андрея Александровича на тему «Рандомизированный подход к обучению в условиях отсутствия разметки и малого количества данных», представленную на соискание ученой степени кандидата физико-математических наук по специальности 01.01.09 – Дискретная математика и математическая кибернетика.

### **Актуальность темы диссертации**

Диссертационная работа Боярова Андрея Александровича «Рандомизированный подход к обучению в условиях отсутствия разметки и малого количества данных» посвящена тематике адаптивного оценивания параметров модели при обучении в условиях отсутствия разметки у данных или наличия лишь слабой разметки. Ее автором исследуется подход, основанный на использовании стохастической аппроксимации в методах машинного обучения и анализа данных. Как известно, для успешной работы основных классических алгоритмов машинного обучения с учителем необходимы возможность вычисления градиента для оптимизируемого функционала среднего риска и данные для обучения как можно большего объёма, при этом тренировочные данные должны быть размечены и максимально очищены от ошибок. Однако, как показывает практика, в реальных задачах эти требования часто бывают трудновыполнимыми, что неизбежно приводит к появлению неопределённостей, которые оказывают значительный негативный эффект на работоспособность традиционных подходов и приводят к существенному ухудшению качества работы обученной модели. Как следствие, возникает необходимость в разработке новых подходов к обучению и оптимизации в описанных условиях, что обуславливает актуальность темы исследования, выбранной диссертантом.

### **Теоретическая значимость и научная новизна**

В диссертационной работе получен ряд содержательных результатов, определяющих её теоретическую значимость и научную новизну. В частности, диссертантом впервые предложен и обоснован рандомизированный алгоритм стохастической аппроксимации для кластеризации в условии смеси гауссовских распределений, а также получена его модификация для случая смеси гауссовских распределений с разреженными параметрами и рандомизированный метод обучения адаптивного классификатора по малому количеству размеченных примеров.

К основным результатам диссертации относятся:

1. Разработка рандомизированного алгоритма стохастической аппроксимации для кластеризации в модели смеси гауссовских распределений, а также анализ качества и скорости работы алгоритма при неизвестных, но ограниченных помехах.
2. Описание и обоснование модификации предложенного рандомизированного алгоритма стохастической аппроксимации для кластеризации применительно к условиям смеси гауссовских распределений с разреженными параметрами.

3. Разработка и обоснование подхода к обучению адаптивного классификатора по малому количеству размеченных примеров, основанного на оптимизации параметров в многозадачной функции потерь с помощью алгоритма стохастической аппроксимации с рандомизацией на входе.

### **Практическая значимость**

Практическая значимость полученных результатов состоит в том, что они могут быть использованы для исследования задач кластеризации в условиях неопределенности, а также задач обучения и классификации по малому количеству данных. В частности, диссертантом разработан набор прикладных программ, реализующих предложенные в ходе диссертационного исследования методы, который убедительно показал свою состоятельность применительно к задачам классификации рукописных букв и верификации авторства средневековых арабских манускриптов.

### **Достоверность результатов**

Достоверность научных результатов диссертационной работы обеспечивается корректным использованием математического аппарата, строгостью математических постановок задач и доказательств утверждений, а также апробацией полученных результатов на авторитетных международных научных конференциях (в том числе – под эгидой IEEE) и наличием публикаций в рецензируемых научных изданиях. Все положения диссертации, выносимые на защиту, строго обоснованы.

### **Замечания**

По диссертационной работе имеются следующие замечания:

1. В разделе 1.1.3 дано определение смеси гауссовских распределений с разреженными параметрами и выделены три типа поведения кластеров при такой модели данных; однако четкое обоснование выделения этих типов отсутствует (поэтому они имеют скорее иллюстративный характер).
2. Было бы полезно обсудить, насколько ограничительны требования, содержащиеся в Предположениях 1-4 в разделе 2.1.1.
3. Выбор параметров рандомизированного алгоритма стохастической аппроксимации для кластеризации для численных экспериментов в разделе 3.1.4 нуждается в большем обосновании. Возможно, имело бы смысл провести эксперименты с несколькими различными наборами параметров.
4. Вводимые в диссертационной работе понятия и термины имело бы смысл оформлять в виде отдельных определений (помимо единственного Определения 1).
5. Работа содержит неизбежные опечатки («точность ... деградируют», стр. 14 и др.) и стилистические неточности («существует много возможных разбиений», стр. 18; «методу нужны все данные», стр. 22 и др.), а рисунки 3.5-3.7 не вполне удачны с визуальной точки зрения.

Указанные замечания в основном носят редакционный характер, они не снижают научной значимости и содержательности полученных теоретических и практических результатов и

не влияют на общее положительное впечатление, которое производит диссертационная работа.

Диссертация Боярова Андрея Александровича является законченным научным исследованием, выполненным на высоком уровне, и представляет несомненный теоретический и практический интерес. Диссертационная работа грамотно и аккуратно оформлена, написана ясным языком; ее отличает четкая логика и продуманность последовательности изложения представленного материала, а также степени его детализации. При написании диссертации Бояров А.А. продемонстрировал глубокую профессиональную эрудицию, понимание мотивации поставленных задач. Список литературы достаточно полон, представленный в диссертации ее обзор адекватно отражает состояние дел в исследуемой области, цитирование работ проведено аккуратно.

Основные результаты диссертационной работы изложены в девяти работах, из которых одна опубликована в издании из списка ВАК, три опубликованы в изданиях, индексируемых в международных наукометрических базах Web of Science и Scopus; отметим также наличие свидетельства о государственной регистрации программы для ЭВМ.

Диссертационная работа Боярова Андрея Александровича на тему «Рандомизированный подход к обучению в условиях отсутствия разметки и малого количества данных» соответствует основным требованиям, установленным Приказом от 01.09.2016 № 6821/1 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете», а ее автор Бояров Андрей Александрович заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 01.01.09 – Дискретная математика и математическая кибернетика. Пункт 11 указанного Порядка диссертантом не нарушен.

Член диссертационного совета

Доктор физико-математических наук,  
профессор РАН,  
главный научный сотрудник,  
и.о. заведующего Лабораторией адаптивных и робастных систем им. Я.З. Цыпкина  
ФГБУН Институт проблем управления РАН им. В.А. Трапезникова



Хлебников Михаил Владимирович

«13» августа 2020 г.



Хлебникова М.В.

