



Федеральное государственное автономное образовательное учреждение высшего образования "Национальный исследовательский университет "Высшая школа экономики"

Факультет
компьютерных наук

**ОТЗЫВ официального оппонента
о диссертации на соискание ученой степени
доктора физико-математических наук Линке Юлианы Юрьевны на
тему: «Универсальные ядерные оценки в непараметрической регрессии
с приложениями к нелинейным регрессионным моделям»
по специальности 1.1.4. – Теория вероятностей и математическая
статистика**

Диссертационная работа Юлианы Юрьевны Линке посвящена актуальным задачам оценивая в непараметрической и нелинейной регрессии в случае плотных данных. В непараметрической регрессии методы ядерного сглаживания применяются, начиная с 1960-х годов. Актуальность данной работы объясняется тем, что автору удалось предложить новые наиболее общие условия на регрессоры в этих задачах, что существенно расширяет область применения указанных методов в непараметрической регрессии.

В первой главе диссертации обсуждаются задачи непараметрической регрессии. Предлагаются новые локально–постоянные и локально–линейные ядерные оценки для регрессионной функции многомерного аргумента. Отдельно рассматриваются одномерный и многомерный случаи. Вводится условие плотного заполнения регрессорами области задания регрессионной функции. Формулируются и доказываются теоремы о состоятельности и асимптотической нормальности предлагаемых оценок. Полученные результаты применяются в задаче оценивания характеристик случайного процесса. В частности, приводятся результаты о состоятельности новых универсальных оценок для функций среднего и ковариации непрерывного случайного процесса в случае разреженных и плотных данных.

Во второй главе диссертации исследуется задача оценивания конечномерных параметров моделей нелинейной регрессии. Предлагается подход построения состоятельных оценок параметров, основанный на ядерных оценках регрессионной функции. Приводятся результаты о состоятельности оценок. Также предлагается метод построения состоятельных оценок параметров, основанный на аддитивных преобразованиях откликов, исследуются асимптотические свойства полученных оценок.

Третья глава диссертационной работы посвящена асимптотическому анализу одношаговых M -оценок, построенных по выборке, элементы которой не обязательно независимы и одинаково распределены, и распределение которых зависит от некоторого конечномерного параметра. Приводится теорема об асимптотической нормальности одношаговых оценок, теорема о минимальном достаточном условии на точность предварительной оценки, теорема о сходимости к нормальному закону погрешностей оценивания при замене асимптотических дисперсий их оценками. Также приводится алгоритм построения асимптотически эффективных одношаговых оценок неизвестного параметра в случае однородной выборки из однопараметрического семейства распределений и достаточно медленно сходящейся к параметру предварительной оценки.

Теоретические результаты диссертации проверяются на различных численных примерах в размерности $k = 1$ и $k = 2$.

Полученные Юлианой Юрьевной Линке результаты являются новыми, вносят важный вклад в развитие регрессионного анализа. Результаты четко сформулированы и снабжены строгими математическими доказательствами. Все основные результаты опубликованы в 19 статьях в научных журналах, индексируемых в Web of Science или Scopus, и докладывались на многочисленных научных семинарах, всероссийских и международных

конференциях. Несмотря на то, что некоторые результаты диссертации получены совместно с соавторами, можно с уверенностью сказать, что вклад Юлианы Юрьевны Линке был ключевым во всех работах. Автореферат правильно и полно отражает содержание диссертации.

Приведу ряд замечаний и вопросов.

- 1) В одномерном случае прием с упорядочиванием регрессоров известен давно. Например, для оценки энтропии этот прием использовался еще в 70-е (Vasicek, J. R. Stat. Soc., 1976).
- 2) В случае многомерных регрессоров для построения оценки (4) необходимо строить разбиение куба $[0,1]^k$ и вычислять меру Лебега каждого множества. Это может привести к существенным вычислительным затратам. В Главе 1.3.3 проведены численные эксперименты только при $k = 2$. Было бы интересно посмотреть, как алгоритм ведет себя в пространствах большей размерности.
- 3) На стр. 80 написано, что в сравнении алгоритмов участвует линейная регрессия. Но на рисунках с максимальными и среднеквадратичными ошибками эта модель отсутствует.
- 4) В каждой из ядерных оценок выбирается оптимальный размер окна с помощью кросс-валидации. При этом не уточняется, как обучаются обобщенная аддитивная модель (GAM) и случайный лес (RF). Деревья какой глубины используются в случайном лесу и как производился выбор глубины? На основе скольких оценок строится предсказание RF? Сколько функций и какие используются в обобщенной аддитивной модели?

5) Почему использован RF, а не бустинг, который обычно работает лучше? Также можно было бы посмотреть, как с аналогичной задачей справляется простая нейронная сеть (feed-forward NN).

6) Универсальный выбор h для всех точек может быть неоптимальной стратегией, особенно в тех примерах, где производная оцениваемой функции значительно меняется. Интересно было бы сравнить выбор оптимального окна с помощью кросс-валидации и адаптивных процедур, например, из (Belomestny, Spokoiny, Ann. Statist., 2007).

7) Рассмотрен только случай неотрицательного ядра. В работе (Samworth, Ann. Statist., 2012) обсуждается, что использование отрицательных весов в методе k ближайших соседей может привести к значительному уменьшению ошибки предсказания в задаче непараметрической классификации. Наблюдается ли подобный эффект в задаче непараметрической регрессии?

8) Всюду предполагается, что регрессоры лежат в кубе $[0,1]^k$. Как следует из работы (Klein, Gadat, Marteau, Ann. Statist., 2016), тяжелые хвосты регрессоров могут значительно усложнить задачу непараметрической классификации, а следовательно, и непараметрической регрессии. Было бы интересно посмотреть, как ведут себя предложенные оценки при более общих предположениях о распределении регрессоров. Также стоит обратить внимание на поведение случайной величины δ_n из раздела 1, когда плотность не отделена от нуля.

9) Не рассмотрены случаи, когда распределение регрессоров или регрессионная функция обладают дополнительной структурой. Например, известно, что оценка локально полиномиальной регрессии адаптируется к внутренней размерности данных (Bickel, Li, IMS Lecture Notes Monogr. Ser.,

2007). В работе (Schmidt-Hieber, Ann. Statist., 2020) рассматривается ситуация, когда f является композицией гладких функций от малого числа переменных. В этом случае эффективная размерность задачи оказывается значительно ниже, чем размерность пространства регрессоров, благодаря чему получается оценить f с гораздо лучшей точностью, чем в общем случае. В качестве оценки в (Schmidt-Hieber, Ann. Statist., 2020) рассматривается нейронная сеть. Способны ли ядерные оценки адаптироваться к такой структуре?

10) Было бы интересно добавить сравнение с более современными алгоритмами и обсудить, что будет, если данные лежат на многообразии низкой размерности.

Вместе с тем, указанные замечания не умаляют значимости диссертационного исследования. Диссертация отвечает требованиям, установленным Московским государственным университетом имени М. В. Ломоносова к работам подобного рода. Содержание диссертации соответствует специальности 1.1.4. – Теория вероятностей и математическая статистика (по физико-математическим наукам), а также критериям, определенным пп. 2.1 – 2.5 Положения о присуждении ученых степеней в Московском государственном университете имени М. В. Ломоносова, и оформлена согласно требованиям Положения о совете по защите диссертаций на соискание ученой степени доктора наук Московского государственного университета имени М. В. Ломоносова.

Таким образом, соискатель Юлиана Юрьевна Линке заслуживает присуждения ученой степени доктора физико-математических наук по специальности 1.1.4. – Теория вероятностей и математическая статистика.

Официальный оппонент:

доктор компьютерных (физико-математических) наук,
заведующий международной лабораторией стохастических алгоритмов и
анализа многомерных данных Института искусственного интеллекта и
цифровых наук факультета компьютерных наук Федерального
государственного автономного образовательного учреждения высшего
образования «Национальный исследовательский университет «Высшая
школа экономики»

Наумов Алексей Александрович



«5» июня 2024 г.

Контактные данные:

тел.: 7(495)772–95–90, e-mail: anaumov@hse.ru

Специальность, по которой официальным оппонентом
защищена диссертация:

1.2.1. – Искусственный интеллект и машинное обучение

Адрес места работы:

101000, г. Москва, ул. Мясницкая, д. 20,

Федеральное государственное автономное образовательное учреждение
высшего образования «Национальный исследовательский университет
«Высшая школа экономики», Факультет компьютерных наук, Институт
искусственного интеллекта и цифровых наук, Международная лаборатория
стохастических алгоритмов и анализа многомерных данных

Тел.: 7(495)771–32–32; e-mail: hse@hse.ru

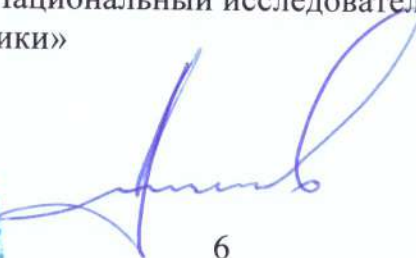
Подпись А. А. Наумова удостоверяю:

Профессор, доктор физико-математических наук, декан

Факультет компьютерных наук

Федеральное государственное автономное образовательное учреждение
высшего образования «Национальный исследовательский университет
«Высшая школа экономики»

И.В. Аржанцев



«5» июня 2024 г.