ПАРАМЕТРИЧЕСКИЙ И НЕПАРАМЕТРИЧЕСКИЙ СПОСОБЫ РЕШЕНИЯ ЗАДАЧИ КЛАССИФИКАЦИИ В ЭПОХУ ПРОГРЕССА ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ. ВЫБОР СТРУКТУРЫ НЕЙРОСЕТИ ПРИ НЕВОСПРОИЗВОДИМОСТИ ОБУЧЕНИЯ

Зенков В.В.
Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия
zenkov-v@yandex.ru

Аннотация. Прогресс вычислительной техники позволяет использовать более простые, непараметрические модели по сравнению с параметрическими за счет использования обучающей выборки не только в период обучения, но и при использовании обученной модели в последующем. Утверждение продемонстрировано на сравнении нейросетевого метода решения задачи классификации и метода аппроксимации дискриминантной функции Андерсона. Предложено выбирать структуру нейросети с учетом диапазона ошибок невоспроизводимости результатов обучения нейросети.

Ключевые слова: аппроксимация, обучающая выборка, искусственная нейросеть, дискриминантная функция Андерсона, невоспроизводимость обучения нейросети.

Введение

В романтические времена 60-х годов прошлого века активно дискутировался вопрос, может ли машина мыслить, можно ли ее натренировать отличать, например, букву А от буквы В. Многие с увлечением придумывали алгоритмы, как тогда говорили, распознавания образов. Обсуждения различных способов распознавания собирали многочисленные аудитории увлеченных этой проблемой людей.

Я.З. Цыпкин в одночасье лишил эту зарождавшуюся науку романтизма, показав, что задачу надо решать с выбора критерия оптимальности, из условий экстремума которого автоматически будет получен алгоритм решения задачи [1].

Люди, которым довелось работать на компьютерах тех лет, физически ощущают гигантский прогресс вычислительной техники в отличие от современного смартфонного поколения людей, воспринимающих реальность, как будто такой она была всегда, разве что в старину айфон был не той модели.

Тот, кто устанавливал в те годы на тумбу дисковода диск емкостью 7 мегабайт диаметром в полметра, или вешал бобину магнитной ленты в лентопротяжное устройство размером больше своего роста в гудящем от вентиляторов машинном зале, с изумлением воспринимает, что этот айфон в руке прохожего имеет несколько процессоров, оперативную память несколько гигабайт и встроенную память больше сотни гигабайт.

1. Преимущества непараметрических моделей

В соответствии с законом диалектики о переходе количественных изменений в качественные [2] и в математике, инструментально связанной со средствами вычисления, по-видимому, должны происходить некоторые качественные изменения, связанные с прогрессом вычислительной техники. Возможно, в частности, это проявляется в усилении роли непараметрических моделей объектов по сравнению с параметрическими моделями.

Рассмотрим особенности параметрической и непараметрической модели на примере параметрической модели – нейросети и непараметрической модели – дискриминантной функции Андерсона (ДФА) [3-8], применяемых к решению задач построения регрессионных моделей и решения задач классификации в эпоху прогресса вычислительной техники.

Качественное отличие параметрической модели от непараметрической состоит в том, что для параметрической модели задается структура некоей аппроксимирующей функции с точностью до параметров. В частности, для нейросетевой модели с выбранными количествами слоев, количествами нейронов в слоях и функциями активации в слоях параметрами нейросети являются многочисленные коэффициенты связи между нейронами слоев. Параметры эти находятся по обучающей выборке с учителем в период обучения. После обучения нейросеть используется уже без обучающей выборки, если нет необходимости или возможности в её улучшении по новым данным.

Параметрическая модель объекта после получения оценок коэффициентов модели больше не нуждается в обучающих данных. Для анализа и дальнейшей работы используется полученная компактная, удобная параметрическая модель объекта.

Следует иметь в виду, что нейросети присущ один недостаток. Она, строго говоря, не может решать задачу аппроксимации или классификации, поскольку повторное обучение ее на одних и тех же обучающих и тестовых данных дает разные результаты, т.е. нейросетью не обеспечивается воспроизводимость результатов обучения [9].

Непараметрическая модель, а именно ДФА, в отличие от параметрической, избавлена от необходимости выбора вида структуры модели, если используется обучающая и тестовая выборки для оценки дискриминантной функции или регрессионной зависимости в заданных точках, а не во всей экспериментальной области. При этом для оценки зависимости в заданных точках просто может использоваться средневзвешенная оценка Надарая-Ватсона [10], получаемая с помощью предварительно обученной методом наименьших квадратов (МНК) весовой функции и далее постоянно используемой обучающей выборки.

Можно также использовать оценки касательной [3] в заданных точках обученным взвешенным МНК. Оба метода практически одинаковы по точности, но метод Надарая-Ватсона существенно быстрее работает, чем метод касательной, и не имеет проблем с возможной вырожденностью матрицы во взвешенном МНК.

В качестве платы за отсутствие необходимости выбора вида аппроксимирующей зависимости при использовании непараметрической модели приходится задаваться видом весовой функции и оценивать по обучающей и тестовой выборкам обычным МНК всего лишь один, редко два, ее параметра обучения. Для сравнения, количество оцениваемых коэффициентов связи между нейронами нейросети при её обучении может исчисляться сотнями и миллионами.

Вид весовой функции не требуется менять так часто, как вид аппроксимирующей зависимости в параметрическом случае. Мы, например, в своих работах [3-8] использовали экспоненциальную весовую функцию.

В непараметрическом способе, кроме того, нужно один раз использовать тестовую выборку и многократно обучающую, затрачивая дополнительные ресурсы на вычисление в заданных точках оценок аппроксимирующей зависимости. Обучающая выборка для непараметрической модели используется в двух ипостасях. Во-первых, для получения обычным методом МНК по обучающей и тестовой выборкам одного, редко двух, параметров обучения весовой функции выбранного вида. Во-вторых, затем многократно использовать только обучающую выборку для получения средневзвешенным методом Надарая-Ватсона или методом касательной оценок аппроксимирующей зависимости в заданных точках с помощью уже обученной весовой функции.

В наших примерах применения ДФА использовалась весовая функция – экспонента от евклидова расстояния очередной точки выборки от заданной точки, умноженного на отрицательный скалярный параметр обучения.

Единственный параметр обучения весовой функции позволяет использовать метод простого перебора для поиска лучшего значения этого параметра, гарантирующего при этом и воспроизводимость результатов обучения. В качестве второго параметра можно использовать, например, показатель степени, в которую возводить расстояние между упомянутыми точками и существенно усложнять этим решение задачи и рисковать потерей воспроизводимости обучения, связываясь с градиентными методами поиска экстремума многоэкстремальной функции неизвестного вида.

Нейросетевую задачу из-за невоспроизводимости результата обучения можно решать многократно в надежде получить хороший результат. Для простых задач это может быть приемлемым, но не для сложных, если выборка гигантская и обучение занимает недели, как в лихо именуемых случаях “глубокого, глубинного обучения,” причем, как правило, пытаются еще и её усложнять до получения приемлемого качества обучения за счет попадания в “хороший” экстремум еще более сложной функции.

При желании усложнить нейросеть добавляют слои и увеличивают количество нейронов в слоях. Далее в работе на простом примере показывается, что при усложнении нейросети полезно учитывать диапазоны разброса ошибок от невоспроизводимости результатов обучения нейросети.

2. Взвешенный метод наименьших квадратов – простой метод решения задачи классификации путем аппроксимации ДФА

Судя по заявлениям в интернете, искусственная нейросеть может решать множество типов задач, используя аналог мыслительной деятельности человека в виде слоеного пирога нейронов и связей между нейронами. Это параметрическая модель с большим количеством настраиваемых при обучении коэффициентов связи между нейронами. Коэффициенты связи находятся градиентным методом, применяемым к многомерной критериальной функции со многими экстремумами. В результате метод нейросетей, строго говоря, не может решать, например, задачу классификации, поскольку на одних и тех же обучающей и тестовой выборках повторные попытки решить задачу дают разные и очень разные результаты [9].

Человек, принимая решение на основе своего опыта, по-нашему мнению, не копается в слоеном пироге своих нейронов. Принятие решения человеком в конкретном случае – это решение задачи путем сопоставления конкретного случая с интуитивно близкими случаями из его практики. Чем ближе случай из практики к конкретному случаю, тем с большим весом он используется при принятии решения.

ДФА по определению является регрессионной зависимостью от признаков **двух классов** в точках признакового пространства, характеризующих объекты классификации. Она представляется разностью двух средних потерь от ошибок в точке пространства признаков от отнесения точки в тот или иной класс. Поэтому в определение ДФА входят и пока неизвестные объективные апостериорные вероятности (АпоВ) этих двух классов, которым может принадлежать объект классификации в точке пространства признаков. Кроме того, в определение ДФА входят и субъективно задаваемые стоимости конкретных ошибок классификации [3].

По знаку ДФА точка в пространстве признаков классов, характеризующая объект классификации, и сам объект относятся в один из двух классов. Так реализуется оптимальный по Байесу метод классификации при двух классах. Для любого другого метода классификации и любого количества классов в задаче используются АпоВ всех классов в точке пространства признаков и субъективно задаваемые стоимости ошибок классификации.

В случае, когда классов больше двух и они взаимно независимы, используется метод последовательного решения задач с двумя классами путем один, целевой класс, АпоВ которого определяется в заданной точке пространства признаков, против всех остальных классов, принимаемых за другой класс.

ДФА имеет одно великолепное свойство, вытекающее из её определения [3]. Аппроксимация ДФА в заданной точке тождественна оцениванию АпоВ двух разделяемых классов в этой точке, сумма которых равна единице, и которые при этом входят в само определение ДФА вместе с субъективно задаваемыми стоимостями ошибок классификации для двух классов.

Второе великолепное свойство ДФА состоит в независимости оценки АпоВ двух классов в заданной точке от субъективно задаваемых стоимостей ошибок классификации этих классов [3].

Оба свойства позволяют применять для определения АпоВ классов в многоклассовой задаче классификации прием: один, целевой класс, АпоВ которого определяется, против всех остальных, принимаемых за другой класс, с произвольно задаваемыми при этом стоимостями ошибок классификации для двух классов.

Возможность получать объективные оценки АпоВ всех классов классифицируемого объекта позволяет решать задачу классификации объектов не только байесовым методом по знаку ДФА при двух классах, но и любыми другими методами и при любом количестве классов при дополнительно заданных, если нужно, ограничениях на АпоВ всех классов.

Простыми являются и математические методы непараметрического решения задачи классификации с помощью ДФА и АпоВ классов. Оценка ДФА в заданной точке может получаться, например, простым средневзвешенным методом Надарая-Ватсона [10] или более сложным методом касательной [3].

Для обучения весовой функции используются обучающая с учителем выборка и тестовая выборка. При обучении должен использоваться тот же самый метод аппроксимации ДФА в точке, какой будет применяться после обучения вместе с обучающей выборкой: средневзвешенный Надарая-Ватсона или метод касательной.

Для получения оценок ДФА в случае с двумя классами в обучающей и тестовой выборках производится замена классов на соответствующие разности стоимостей ошибок классификации. Эти разности являются по определению ДФА оценками ДФА в точках пространства признаков классов [3].

Аналогичная замена производится при решении многоклассовой задачи при оценивании каждой ДФА в методе один класс против всех остальных.

Для поиска параметра обучения весовой функции по обучающей выборке в задаче с двумя классами используются следующие критерии. Критерий МНК для суммы квадратов ошибок аппроксимации ДФА в точках тестовой выборки, полученных с помощью весовой функции и обучающей выборки, байесов критерий или любой другой критерий, использующий оценки АпоВ двух классов по оценкам ДФА в точках тестовой выборки.

При этом можно ограничиться одним лучшим параметром обучения и, следовательно, общей весовой функцией для всех классов и всех ДФА. Единственный параметр обучения весовой функции может находиться методом простого перебора по общему критерию решения задачи классификации.

Наряду с этим возможно в принципе использовать и несколько видов весовых функций, и индивидуальные обучения одной весовой функции для разных классов и групп классов и прочее.

Использование средневзвешенной оценки Надарая-Ватсона по сравнению с методом касательной существенно ускоряет процедуру обучения, а также и дальнейшее получение оценок ДФА и АпоВ классов в заданных точках.

Метод касательной может осложняться еще и вырожденными случаями с матрицей решения взвешенного МНК. Когда лучше применять тот или иной метод – вопрос открытый.

При малом объеме обучающей выборки или при отсутствии желания делить обучающую выборку на две части, обучающую и тестовую, можно воспользоваться методом LOO (leave-one-out) [10]. В методе LOO последовательно каждая точка обучающей выборки с возвратом выполняет роль тестовой, а остальные – роль обучающих.

Выбор вида весовой функции и меры расстояния между точками – вопрос открытый. Мы в своих примерах использовали в качестве весовой функции экспоненту от евклидова расстояния точки выборки до заданной точки пространства признаков классов, умноженного на отрицательный коэффициент обучения, обеспечивающий уменьшение веса точки выборки с ростом расстояния ее от заданной точки.
В качестве меры расстояния между точкой выборки и заданной точкой также не обязательно может быть евклидова мера.

3. Выбор структуры нейросети с учетом невоспроизводимости обучения

Структура нейросети, в частности, задается количеством слоев и количествами нейронов в слоях. Рассмотрим на простом примере, как влияет увеличение количества слоев или количества нейронов в слоях на невоспроизводимость обучения нейросети. Чем шире диапазон погрешностей от невоспроизводимости обучения нейросети, тем хуже нейросеть заданной структуры.

Для оценки диапазона повторяем обучение варианта нейросети заданное количество раз и вычисляем по тестовой выборке погрешности оценки АпоВ класса А.

Интуитивно кажется, что усложнение нейросети приводит к улучшению ее качества. Но как это усложнение может сказаться на невоспроизводимости, покажем на простом примере, взятом из [9].

Рассматриваем задачу классификации с классами A, B и C. Пространство признаков классов одномерное. Классы имеют нормальные условные распределения этого признака. Классы B и C расположены по бокам класса A на оси признака. Рассмотрим один этап решения задачи – класс A против классов B и C, воспринимаемых за другой класс. Известны средние и дисперсии условных распределений и априорные вероятности всех трех классов. Рассчитываем истинную АпоВ класса А.

Метод нейросети используем без учета вышеперечисленной априорной информации. Имеем лишь обучающую выборку с учителем, сгенерированную по априорной информации в количестве 120 точек. Точки обучающей выборки класса А создают один класс. Точки классов B и C имеют метку другого класса.

Тестовая выборка задана 30-ю точками, равномерно расположенными на некотором интервале признака классов.

Нейросеть обучается на обучающей выборке и используется для предсказания АпоВ класса А в тестовых точках признака классов. По оценкам нейросети АпоВ класса А в точках тестовой выборки вычисляется средняя квадратичная ошибка обучения варианта нейросети. Находятся также минимальная и максимальная ошибки невоспроизводимости для каждого варианта нейросети.

Нейросеть реализована на Питоне. Она имеет один нейрон на входе, один на выходе. Содержит изначально один скрытый слой с четырьмя нейронами. Функция активации скрытого слоя – relu, выходного – sigmoid.

Произведем сравнение двух стратегий усложнения нейросети: путем увеличения количества нейронов в единственном скрытом слое и путем увеличения количества скрытых слоев с одинаковым заданным количеством нейронов в каждом слое.

Нейросеть каждого варианта структуры обучаем 20 раз невоспроизводимым способом. Воспроизводимость 20-ти вариантов обучения методом [9] не пытаемся обеспечить ввиду зависимости результата от операционных систем и версий пакетов Питона keras и tensorflow.

Вычисляем ошибки оценки нейросетью АпоВ класса А в тестовых точках, находим минимальную и максимальную ошибки оценки для каждой структуры из 20 полученных результатов её обучения. Строим график зависимостей предельных оценок для случая увеличения количества нейронов в единственном скрытом слое, Рис.1, и для случая увеличения количества скрытых слоев с четырьмя нейронами в слоях, Рис. 2 и с 10-ю нейронами в слоях, Рис. 3.

На участках с уменьшающейся минимальной ошибкой нейросети на графике Рис. 1, имеет место и резкое уменьшение максимальной ошибки нейросети – уменьшение диапазона невоспроизводимости нейросети.

На Рис. 2 максимальная ошибка нейросети в вариантах обучения не уменьшается.

Малый диапазон невоспроизводимости варианта нейросети при небольшой его минимальной ошибке свидетельствует в пользу выбора варианта. Малый диапазон невоспроизводимости повышает вероятность хорошего результата от использования такого варианта.

Малая ошибка бэггинга [9], построенного на 5 лучших из 20 обучений, также свидетельствует в пользу выбора варианта.

Так из Рис. 1 следует, что при одном скрытом слое хороший результат будет при 100 нейронах в слое.

Из Рис. 2 следует ожидать хорошего результата при 3-7 скрытых слоях с четырьмя нейронами. Из Рис.3 следует ожидать хорошего результата при 4-6 скрытых слоях с 10 нейронами в скрытом слое. Большая ошибка бэггинга в варианте с 10 слоями на Рис. 2 не рекомендует рассматривать этот вариант нейросети как предпочтительный.

Для сравнения на графиках изображены и горизонтальные прямые погрешностей непараметрического метода касательной и метода Надарая-Ватсона, использованных для решения той же задачи [3]. Их погрешности также можно использовать в качестве верхних ориентиров при выборе структуры нейросети.

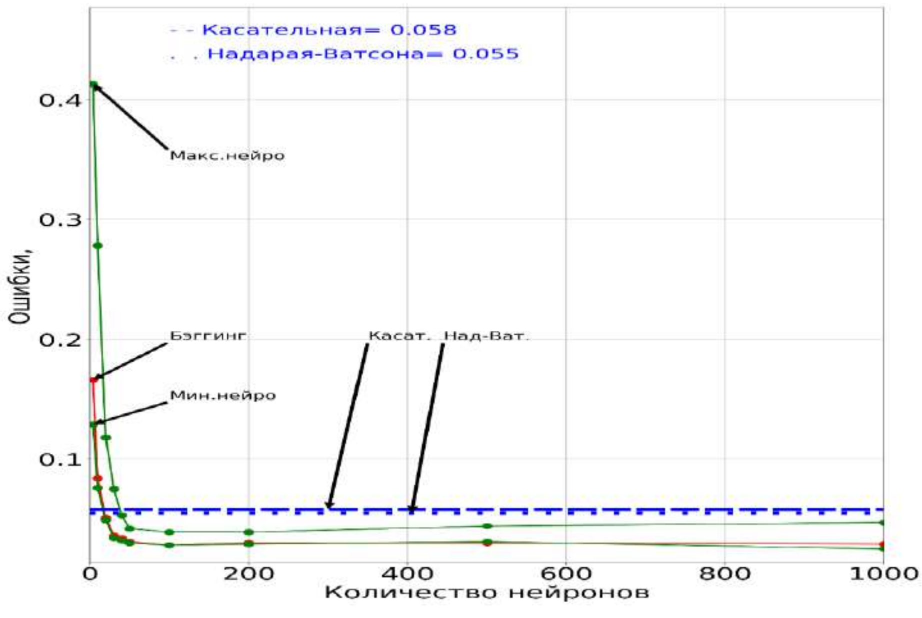

Рис. 1. Один слой с разным количеством нейронов в слое

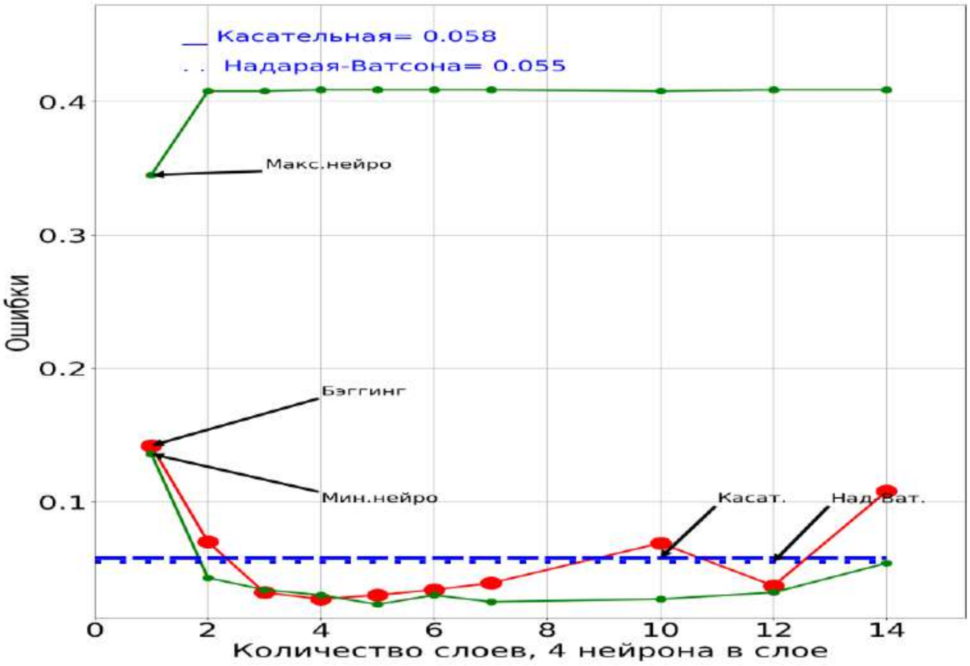

Рис. 2. Разное количество слоёв с четырьмя нейронами в слое

В случае с одним слоем для улучшения обучения нейросети следует увеличивать количество нейронов в слое вместо добавления новых слоев.

Не следует забывать, что рекомендация эта носит весьма частный характер. Она получена в условиях невоспроизводимости результатов обучения на компьютере. Демонстрационный пример очень простой, не требующий заметных затрат времени на многократные прогоны обучения.

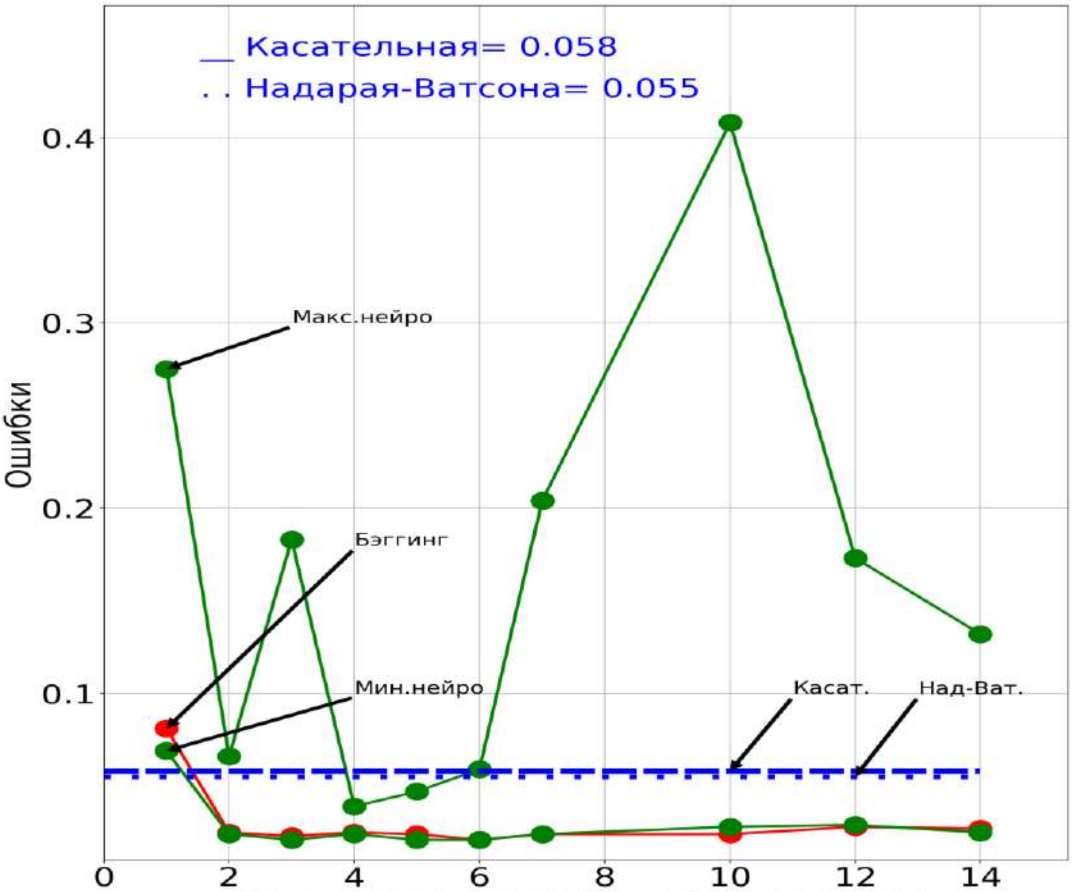

Рис. 3. Разное количество слоёв с десятью нейронами в слое

4. Заключение

Прогресс вычислительной техники усиливает роль непараметрических моделей, например, в задачах классификации и построения регрессионных зависимостей, избавляя от необходимости задаваться видом аппроксимаций за счет возможности аппроксимировать регрессионную зависимость или дискриминантную функцию в заданной точке пространства признаков. Но при этом приходится использовать обучающую выборку всегда, а не только в период обучения.

Выполнено сравнение на простом примере проблем построения нейронной сети, как параметрической модели с одной стороны, и дискриминантной функцией Андерсона, как непараметрической модели с другой стороны, использующей метод аппроксимации функции в заданной точке пространства признаков, избавляющий от необходимости задаваться видом аппроксимирующей функции, но затрачивать дополнительные вычислительные ресурсы на многократное использование обучающей выборки.

При поиске структуры нейросети рекомендовано учитывать такие характеристики невоспроизводимости ее обучения, как диапазон погрешностей от невоспроизводимости, оценку наименьшей погрешности дискриминантной функции Андерсона и погрешность бэггинга, построенного на лучших результатах обучения.

Можно использовать при выборе варианта нейросети в качестве верхней границы ошибки варианта также ошибку аппроксимации дискриминантной функции Андерсона методом касательной или Надарая-Ватсона или ошибку оценки апостериорной вероятности класса, получаемых в точке непараметрическими методами без необходимости выбора вида аппроксимирующей зависимости и лишенных недостатка невоспроизводимости результатов обучения.

Литература

1. *Цыпкин Я.З.* Основы теории обучающихся систем. – М.: Наука, 1970. – 252 с.
2. Закон перехода количественных изменений в качественные. https://ru.wikipedia.org/wiki (дата обращения 24.06.2025).
3. *Зенков В.В.* Дискриминантная функция Андерсона или нейросеть для оценки апостериорных вероятностей классов при решении задач классификации в машинном обучении // Управление развитием крупномасштабных систем (MLSD’2022): труды 15-й Международной конференции. – М.: ИПУ РАН, 2022. – С. 1199–1208.
4. *Zenkov V.* Estimation of the Posterior Probabilities of Classes by the Approximation of the Anderson Discriminant Function // Advances in Systems Science and Applications. – 2022. – Vol. 22, № 4. – P. 233–241.
5. *Зенков В.В.* Применение аппроксимации дискриминантной функции Андерсона и метода опорных векторов для решения некоторых задач классификации // Автоматика и телемеханика. – 2020. – № 1. – С. 147–160.
6. *Зенков В.В. Молочкова Ю.В., Молочков В.А., Хлебникова А.Н., Глазков А.А.* Оценка апостериорных вероятностей классов в задаче дифференциальной диагностики заболеваний слизистой оболочки полости рта // Управление развитием крупномасштабных систем (MLSD’2020): труды 13-й Международной конференции. – М.: ИПУ РАН, 2020. – С. 1793–1801.
7. *Зенков В.В.* Диагностика пневмонии по рентгеновским снимкам с помощью дискриминантной функции Андерсона // Управление развитием крупномасштабных систем (MLSD’2023): труды 16-й Международной конференции. – М.: ИПУ РАН, 2023. – С. 1390–1397.
8. *Зенков В.В.* Корреляционный метод свертки изображений // Управление развитием крупномасштабных систем (MLSD’2023): труды 16-й Международной конференции. – М.: ИПУ РАН, 2023. – С. 1398–1402.
9. З*енков В.В.* Бэггинг на отсутствии воспроизводимости обучения в нейросети // Управление развитием крупномасштабных систем (MLSD’2024): труды 17-й Международной конференции. – М.: ИПУ РАН, 2024. – С. 1100–1105.
10. *Воронцов К.В.* Математические методы обучения по прецедентам (теория обучения машин). http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf (дата обращения 24.06.2025).