МАШИННОЕ ОБУЧЕНИЕ. АППРОКСИМАЦИЯ ДИСКРИМИНАНТНОЙ ФУНКЦИИ АНДЕРСОНА И ОЦЕНКА АПОСТЕРИОРНЫХ ВЕРОЯТНОСТЕЙ КЛАССОВ

Зенков В.В.

Институт проблем управления им. В.А. Трапезникова,
Россия, г. Москва, ул. Профсоюзная, д.65
zenkov-v@yandex.ru

Аннотация: Аппроксимация дискриминантной функции Андерсона в заданной точке пространства признаков двух классов по обучающей выборке с учителем позволяет одновременно оценивать и объективные апостериорные вероятности классов – исчерпывающие данные для решения задачи классификации в машинном обучении при субъективно задаваемых стоимостях ошибок и критериях классификации. Обучающая выборка с учителем преобразуется в выборку регрессионного анализа заменой номеров (меток) классов соответствующими разностями стоимостей ошибок. Получающиеся оценки апостериорных вероятностей классов не зависят от выбора стоимостей ошибок классификации, используемых для определения дискриминантной функции. Для аппроксимации дискриминантной функции в точке по обучающей выборке с учителем не требуется задания функции аппроксимации сложнее линейной. Кроме непараметрической модели, использующей обучающую выборку для каждой точки, предложены две параметрические: аппроксимация в окрестности нулевых значений дискриминантной функции и серия аппроксимаций при заданных апостериорных вероятностях на границе между классами. Апостериорные вероятности в заданной точке находятся при этом методом интер-экстраполяции или методом сглаживания. Приведены примеры.

Ключевые слова: Дискриминантная функция Андерсона, аппроксимация. апостериорная вероятность класса, машинное обучение..

Введение

Апостериорные вероятности (АпоВ) классов являются исчерпывающей информацией для решения задачи классификации. Они позволяют относить объект классификации в тот или иной класс с учетом разных матриц стоимостей ошибок классификации, выбираемых пользователем не только для всей задачи в целом, но и для каждого объекта классификации в отдельности, в зависимости, например, от некоторого статуса классифицируемого объекта. Задача решается в два этапа. На первом этапе для объекта классификации находятся АпоВ классов – объективные данные для классификации. На втором этапе объект относится в тот или иной класс с учетом субъективных предпочтений пользователя, проявляющихся в выборе стоимостей ошибок классификации, в методике самого процесса принятия решения, например, с учетом некоторых ограничений, правила отмены принятия решения, условий проведения дополнительных исследований и т.п..

В связи со значимостью АпоВ классов для решения задачи классификации и для интерпретации принимаемых решений при классификации объектов используются надстройки к некоторым уже существующим методам классификации. Так, например, для метода опорных векторов используется калибратор Платта [1], оценивающий АпоВ классов по расстоянию точки в пространстве признаков классов до границы между классами. При этом приходится выдвигать гипотезу о виде зависимости АпоВ классов от расстояния точки до границы между классами, а затем по обучающей выборке подбирать параметры гипотетической зависимости методом максимального правдоподобия.

Дискриминантная функция Андерсона (ДФА), по знаку которой точка в пространстве признаков классов относится в один из двух классов, является регрессионной зависимостью. Она удобна тем, что по ее аппроксимации, получаемой взвешенным методом наименьших квадратов по обучающей выборке с учителем, и по стоимостям ошибок, используемых для ее определения, АпоВ классов алгебраически вычисляются так же просто, как по градусам Фаренгейта получаются градусы Цельсия.

Мы определяем ДФА по байесову решению задачи классификации [2], полученному Теодором Уилбурном Андерсоном, американским математиком, в середине прошлого века. ДФА - это разность двух функций средних потерь от ошибок классификации в пространстве признаков классов при заданных стоимостях потерь от ошибок классификации в бинарной задаче классификации. ДФА, таким образом, по определению является функцией регрессии. Особенностью ДФА является ее тождественная связь с АпоВ двух классов, (сумма равна единице) и со стоимостями ошибок классификации, определяющими ДФА. Это позволяет алгебраически просто вычислять АпоВ классов в заданной точке пространства признаков по аппроксимации ДФА в этой точке.

Аппроксимация ДФА получается по обучающей выборке с учителем, преобразуемой в выборку регрессионного анализа путем замены номеров (меток) классов соответствующими разностями стоимостей ошибок классификации, задаваемыми произвольно для определения ДФА.

При получении аппроксимации ДФА в точке нет необходимости задаваться видом аппроксимирующей зависимости в пространстве признаков классов. Достаточно по близким точкам выборки к заданной точке строить средневзвешенную линейную аппроксимацию ДФА, или методом Надарая-Ватсона [3] находить средневзвешенное значение ДФА в этой точке.

Задача классификации точки с помощью ДФА решается в два этапа. На первом (объективном) этапе находятся АпоВ классов точки, на втором (субъективном) точка относится в тот класс, который соответствует выбранному методу классификации, использующему полученные АпоВ классов.

ДФА используется только на первом этапе.

В определение ДФА входят стоимости ошибок классификации. Оценки АпоВ классов в точке по полученной аппроксимации ДФА в этой точке не зависят от выбора стоимостей ошибок, используемых в определении ДФА. Можно на первом этапе задать стоимости ошибок равными, например, 1 и 1, или не равными, например, 2 и 5. Аппроксимации ДФА в точке будут разные, но оценки АпоВ в точке будут одинаковыми. Стоимости ошибок классификации нужны так же для преобразования обучающей выборки в выборку регрессионного анализа. Номера классов в первом случае заменяются на -1 и +1, во втором – на -2 и +5. В пересчет полученного значения ДФА в точке в оценки АпоВ классов входят использованные для ДФА стоимости ошибок классификации. Вследствие этого результат оценивания АпоВ не зависит от выбранных на первом этапе стоимостей ошибок классификации. Это важное свойств ДФА.

В случае, когда классов больше двух, задача решается путем последовательного сведения к нескольким задачам по принципу один из классов против остальных. Количество решаемых бинарных задач равно количеству классов и, возможно, минус один класс. Граница между одним из классов в пространстве признаков и остальными классами сложнее, чем между парой классов, но для аппроксимации ДФА в точке по близким точкам это не имеет принципиального значения.

Метод решения задач, в которых количество точек одного класса в выборке значительно больше точек другого класса, остается тем же. Нет смысла устранять дисбаланс классов в выборке с помощью, например, разных стоимостей ошибок классификации что в случае с ДФА не меняет результата оценивания АпоВ классов. Дисбаланс почти всегда имеет место, когда классов больше двух и используется прием - один класс против остальных.

На первом этапе нет смысла выбирать разные стоимости ошибок классификации. Достаточно всегда использовать равные стоимости ошибок, например, 0.5 и 0.5.

Метод решения задачи классификации в точке пространства признаков по обучающей выборке с учителем, когда для каждой классифицируемой точки нужно использовать всю обучающую выборку, является непараметрическим, а используемая модель, в которой как обязательная часть присутствует обучающая выборка, является непараметрической моделью задачи классификации.

1 Дискриминантная функция Андерсона

ДФА определяем как разность двух функций средних потерь от ошибок классификации в задаче с двумя классами:

$$f_{12}\left(x,C\right) = G_1\left(x,C\right)-G_2\left(x,C\right)=\mathrm{M}_{k|x}\left(C_{1k}-C_{2k}\right), \qquad (1)$$

где $G_1(x,C) = C_{12}(1-p(1|x))$, $G_2(x,C) = C_{21}\,p(1|x))$ - средние потери в точке x, если точку отнести к классу 1 или, соответственно, к классу 2; C_{ij} –стоимость ошибки, когда точка из класса j ошибочно относится в класс i, $C_{ij} > 0$, $i \neq j$, $C_{ii} = 0$; $p(k|x)$ – АпоВ класса k в точке x, $p(1|x)+p(2|x)=1$, $p(k|x)=P_k p(x|k)/p(x)$, $p(x)=\sum_k P_k p(x|k)$, P_k – априорные вероятности классов, $p(x|k)$ –условные распределения признаков классов; k – номера классов 1 или 2; $M_{k|x}(\cdot)$ – математическое ожидание по k в точке x.

Если $f_{12}(x,C) \le 0$, то точка x относится в класс 1, иначе – в класс 2.

1.1 Свойства ДФА

1.1.1 *ДФА по определению есть функция регрессии.* Для преобразования обучающей выборки задачи классификации с учителем в выборку регрессионного анализа нужно в соответствии с (1) заменить в выборке номер класса 1 на $-C_{21}$ а номер класса 2 на C_{12}.

1.1.2 *Для АпоВ первого класса и ДФА, определенной для заданных* C_{12} *и* C_{21}*, имеет место тождество*

$$p(1|x) \equiv (C_{12} - f_{12}(x,C)) \;/\; (C_{12} + C_{21}) \qquad (2)$$

Если задавать стоимости ошибок при условии $C_{12} + C_{21} = 1$, что не приводит к потере общности, то тождество (2 упрощается

$$p(1|x) \equiv p^* - f_{12}(x, p^*) \qquad (3)$$

и на границе между классами, где $f_{12}(x,p^*) = 0$, имеет место $p^*=p(1|x)=C_{12}$ и $1\text{-}p^*=C_{21}$. АпоВ первого класса на границе между классами равна p^*.

Следствие.

Из (1-3) следует, что АпоВ классов в точке x не зависят от того, для какого $p^* > 0$ или C_{12} и C_{21} определена ДФА.

1.1.3. *При выборе стоимостей ошибок С для классификации по полученным на первом этапе АпоВ классов на втором этапе возникнет неразличимость классов в пространстве признаков, если*

$$(\min_x \; f_{12}(x,C) \; > \; 0) \vee (\max_x \; f_{12}(x,C) \; < \; 0) \;. \qquad (4)$$

Все точки в этом случае нужно будет относить в один и тот же класс.

1.2 Способы аппроксимации ДФА

Аппроксимировать ДФА можно, например, тремя способами:

- аппроксимировать ДФА в точке пространства признаков классов для получения оценок АпоВ классов в этой точке по (2) или (3) в зависимости от способа выбора стоимостей ошибок классификации. Это способ получения непараметрической модели решения задачи.
- аппроксимировать ДФА во всей области обучающей выборки, задавшись видом аппроксимирующей функции и выбрав метод аппроксимации. Для этой параметрической модели предложен метод аппроксимации ДФА в изначально не известной окрестности ее нулевых значений, чтобы решать точнее задачу классификации для выбранного вида аппроксимирующей зависимости и заданных стоимостей ошибок классификации и не преследовать цель оценивать АпоВ классов в точках пространства признаков классов.
- аппроксимировать несколько ДФА в соответствии с заданными значениям АпоВ первого класса на границе между классами, например, 0.1, 0.3, …, 0.9. и с заданным видом/видами аппроксимирующих зависимостей. Это вариант серии аппроксимаций ДФА. Для заданной точки в пространстве признаков АпоВ классов находятся методом интер- экстраполяции с использованием соседних аппроксимаций ДФА, между которыми находится точка.

Во всех способах размерность обращаемой матрицы при решении задачи, равна размерности искомого вектора параметров и не зависит от количества строк в обучающей выборке. Вид весовой функции - не обязательно экспонента.

Непараметрическая модель работает медленнее, она использует всю обучающую выборку в тестовом режиме. Параметрические быстрее работают, но требуют выбора вида аппроксимирующей зависимости для ДФА в режиме обучения.

1.2.1 Аппроксимация ДФА в точке

Этот вариант строит непараметрическую модель задачи [4]. Ее особенностью является необходимость использования всей обучающей выборки или ее части для оценки АпоВ классов в каждой заданной точке. При этом не требуется выдвижения гипотез о виде аппроксимирующей зависимости ДФА в окрестности заданной точки, достаточно использовать аппроксимирующую зависимость не сложнее линейной.

Критерием аппроксимации ДФА в заданной точке выборки x_j является средневзвешенный квадрат ошибки. На этапе обучения для каждой точки выборки x_j, используемой в качестве тестовой, находится вектор коэффициентов λ_j аппроксимаци ДФА в виде (1, x_j)λ_j при заданных значениях параметров W и S весовой функции:

$$Q(\lambda, x_j) = \min_{\lambda_j} \sum_{n=1, n\neq j}^{n=N} \{[C_{1k_n} - C_{2k_n} - (1, x_n)\lambda_j]^2 \exp(-W \| x_j - x_n \|^S)\}, \qquad (5)$$

где вектор-строка (1, x_n) кроме вектора-строки признаков x_n обучающей выборки содержит компонент - единицу; N – количество строк в выборке без строки x_j на этапе обучения, как того требует метод борьбы с переобучением LOO (leave-one-out); k_n – номер класса 1 или 2 в строке выборки n: x_n – вектор признаков в строке выборки n; $W \geq 0$ – весовой коэффициент, задающий скорость спада весовой функции (в данном случае экспоненты от расстояния точки выборки до заданной точки), S – показатель степени, в которую возводиться расстояние от точки выборки до заданной точки; $C_{1k_n} - C_{2k_n}$ - оценка ДФА в строке выборки n, полученная путем замены номера класса k_n, 1 или 2, в строке на разность стоимостей ошибок $C_{1k_n} - C_{2k_n}$, определяемых по заданному p^*.

Получив вектор коэффициентов λ_j путем минимизации критерия (5) при заданных параметрах метода W и S, вычисляем оценку ДФА в точке x_j, $f_{12}(x_j) = (1, x_j)\,\lambda_j$ и по вышеприведенному тождеству находим оценку АпоВ первого класса в точке x_j. Если количество классов $K>2$, то аналогично получаем АпоВ всех классов в этой точке, используя метод один класс против всех остальных, объединяемых в другой класс.

Лучшие значения параметров W и S подбираются, например, по минимуму критерия

$$R = N^{-1}(C_{12}N_2 + C_{21}N_1), \qquad (6)$$

где N_1 – количество точек первого класса, ошибочно отнесенных во второй класс, N_2 – количество точек второго класса, ошибочно отнесенных в первый класс. Лучшими значениями параметров весовой функции является те, которым соответствуют меньшие потери (6).

Нередко параметр весовой функции S принимается равным единице, а подбирается значение единственного параметра W.

Вид весовой функции выбирается произвольно. Предполагается, что вид весовой функции мало влияет на результат аппроксимации ДФА в точке.

1.2.2 *Аппроксимация ДФА в окрестности нулевых значений*

В некоторых случаях (большой объем обучающей выборки, наличие априорных сведений об условных распределениях признаков классов, о виде разделяющей классы поверхности) целесообразнее построить параметрическую модель. Она может быть в виде дискриминантной функции заданного вида с неизвестными параметрами, определяемыми по обучающей выборке. Для классификации точки можно использовать формулу Байеса для пересчета восстановленных условных распределений признаков классов в АпоВ классов в задаваемых точках и с учетом стоимостей ошибок классификации относить точку в тот или иной класс оптимальным по Байесу способом.

В работе [5] предложен эвристический способ построения параметрической модели задачи классификации по обучающей выборке с учителем, основанный на аппроксимации дискриминантной функции в окрестности ее нулевых значений при заданных стоимостях ошибок классификации. Об оценивании АпоВ классов в этом случае речь обычно не идет.

Обучающая выборка с учителем не содержит сведений о положении нулевых точек ДФА. Задав стоимости ошибок классификации, получают в итоге границу между классами или ее отсутствие, когда все точки следует относить в один и тот же класс.

Для решения задачи используется эвристический метод [5] приближения к границе между классами. Вначале, исходя из заданных стоимостей ошибок классификации, обучающая выборка преобразуется в выборку задачи регрессионного анализа путем замены номеров(меток) классов на соответствующие разности стоимостей ошибок классификации (1). Затем запускаются несколько итераций решения задач аппроксимации ДФА взвешенным методом наименьших квадратов. В качестве весов точек выборки используется весовая функция от расстояния точки до нуля предыдущей полученной аппроксимации ДФА. Мерой расстояния точки до нуля предыдущей аппроксимации служит модуль предыдущей аппроксимации ДФА в этой точке. Критерий решения на каждом шаге с экспоненциальной весовой функцией:

$$Q(\lambda_i) = \min_{\lambda_i} \sum_{n=1}^{n=N} \{[C_{1k_n} - C_{2k_n} - \lambda_i{}'\varphi(x_n)]^2 \exp(-W_i \,|\, \lambda_{i-1}{}'\varphi(x_n)\,|^S)\}, \qquad (7)$$

где i – номер итерации, $i = 1 \div I$. На первом шаге задача решается без весовой функции (W_0=0), на последующих шагах весовая функция придает больший вес точкам, более близким к нулевой области предыдущей аппроксимации ДФА. I – заданное количество итераций, $I < \infty$. W_i – заданный весовой коэффициент на шаге i, $W_i \geq 0$, S – заданный показатель степени, обычно S=1. Весовой коэффициент в итерациях может быть постоянным или изменяться, например $W_i = w*i,\ i = 0 \div I$. Для каждой итерации по полученной аппроксимации ДФА находятся потери (6). Лучшим значением λ является тот вектор, которому соответствуют меньшие потери (6).

1.2.3 Оценка АпоВ в точке по серии аппроксимаций ДФА

В этом параметрическом методе [6] по обучающей выборке строятся несколько аппроксимаций ДФА при заданных при построении стоимостях ошибок классификации, соответствующих ряду АпоВ первого класса на границе между классами, где аппроксимации принимают нулевые значения. На этом заканчивается процесс обучения и обучающая выборка более не используется.

Построенная серия аппроксимаций может использоваться для оценки АпоВ классов в заданной точке пространства признаков двумя методами: интерполяционным методом и методом сглаживания.

В интерполяционном методе для оценки АпоВ первого класса в заданной точке по знакам аппроксимаций ДФА в серии в этой точке находится пара соседних аппроксимаций, между которыми находится точка. Если пара найдена, то методом интерполяции получается оценка АпоВ первого класса. Мерой расстояния до соседних границ с известными по построению АпоВ первого класса на этих границах служат модули значений соседних аппроксимаций ДФА в заданной точке. Если заданная точка выходит за пределы аппроксимаций ДФА, то оценка АпоВ класса находится методом экстраполяции по первой или последней аппроксимации ДФА в серии с проверкой результата на не отрицательность и на не превышение единицы.

В методе сглаживания по серии аппроксимаций ДФА с известными величинами АпоВ на границе классов оценка АпоВ в заданной точке находится по серии аппроксимаций ДФА с соответствующими весами, вычисленными по степени близости к заданной точке. В качестве меры близости к j-аппроксимации ДФА с заданной вероятностью первого класса на границе p^*, $\mathrm{F}_j(x,p_j^*)$, используется модуль значения j-аппроксимации ДФА в заданной точке. В качестве весовой функции используется экспонента

$$w_j(x) = \exp(-W \,|\, \mathrm{F}_j(x, p_j^*)\,|), \quad j = 1 \div J, \qquad (8)$$

где W – параметр, задающий сглаживающие свойства метода, $W > 0$.

Взвешенной оценкой АпоВ в точке x с помощью серии АДФ и (8) будет

$$p(1\,|\,x) = \sum_j p_j^* w_j(x) \;/\; \sum_j w_j(x). \qquad (9)$$

2 Примеры

2.1 Аппроксимация ДФА в точке

Приведен случай, [4] Рис.1, с нормальными условными распределениями двумерных признаков трех классов. Один из них, класс А, расположен между классами В и С. Класс А обозначен номером 1, классы В и С объединим в один класс и обозначены номером 2. Средние классов: $ma = (0,\ 0)$, $mb = (-3,\ 0)$, $mc = (3,\ 0)$. Ковариационные матрицы классов одинаковые $sa = sb = sc = (1,\ \ 0;\ \ 0,\ \ 1)$. Априорные вероятности классов Pa=0,5, Pb=Pc=0,25. Сгенерированая по этим параметром выборка содегжит 120 точек. На рисунке представлены теоретические плотности условных распределений признаков классов, теоретические ДФА и АпоВ первого класса, а также ваборочные оценки АпоВ в заданных точках, полученные при S=1 и W=3 (8). Среднеквадратичные ошибки (СКО) в оценке АпоВ первого класса по отношению к теоретическим составили 0,09 и 0,18 соответственно для линейной и полиномиальной аппроксимации второго порядка ДФА. Линейная аппроксимация ДФА оказалась в этом примере точнее.

Оценки АпоВ первого класса в заданных точках располагаются вблизи выборочной кривой АпоВ первого класса, полученной по оценкам параметров нормальных законов условных распределений признаков (пунктирная кривая рядом со сплошной теоретической кривой АпоВ первого класса).

Поскольку признаков два, то на плоском рисунке представлено сечение плоскостью, проходящей по оси значений первого признака и по оси ординат.

На Рис. 2 представлен пример с одномерным признаком, распределенным по равномерному закону. Классы А, В и С имеет средние $ma = 0$, $mb = -3$, $mc = 3$. Пределы максимальных отклонений от средних в обе стороны одинаковы и равны 2. Классы пересекаются в пространстве признаков. Априорные вероятности классов такие же, как и в предыдущем случае. Обучающая выборка 120 точек, S=1 и W=3. Визуально оценки АпоВ первого класса полиномом второго порядка в этом примере с одномерным признаком кажутся точнее.

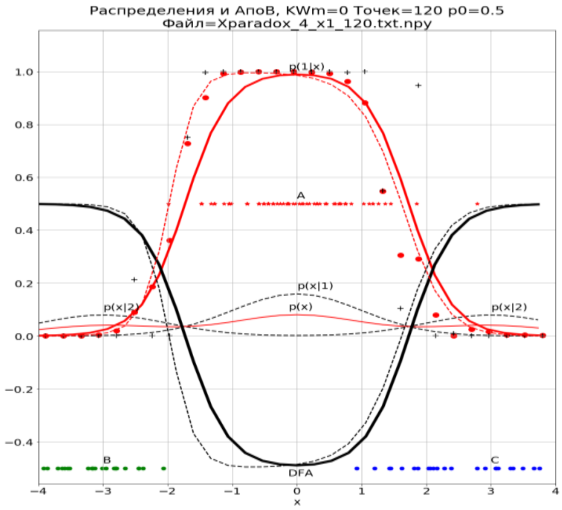

Рис. 1.

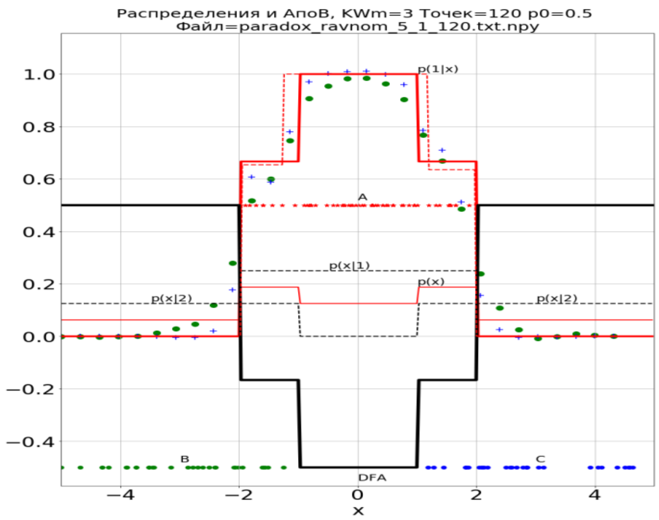

Рис. 2.

Пояснения к рисункам 1 и 2. Точки выборки трех классов А, В и С расположены на уровнях p^* и 1-p^*. Условные распределения признаков классов 1 и 2 (один против объединения двух других) – это пунктирные линии $p(x|1)$ и $p(x|2)$ на рисунках. $p(x)$ – плотности распределения x – сплошные тонкие линии. ДФА – это теоретические ДФА (сплошные толстые линия выпуклостью вниз). $p(1|x)$ – теоретические АпоВ первого класса (сплошные толстые линия выпуклостью вверх). Близкие к теоретическим пунктирные линии – это построенные по выборкам ДФА и АпоВ, для известных условных распределений признаков классов и заданных стоимостей ошибок классификации p^* На Рис. 2 выборочная ДФА не показана. Она – зеркальное отражение АпоВ. Круглые точки возле АпоВ первого класса– оценки АпоВ путем аппроксимации ДФА в заданных точках плоскостью; крестики – оценки АпоВ с использованием для аппроксимации ДФА полинома второго порядка.

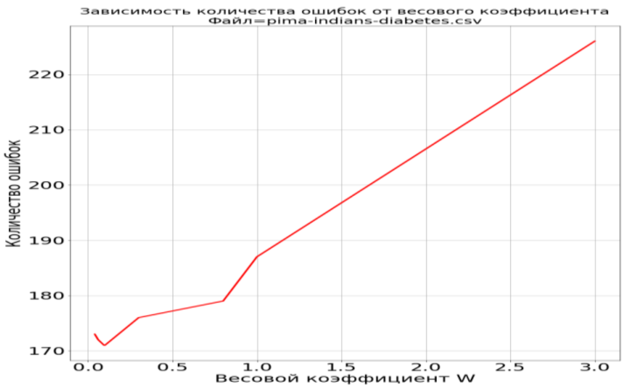

Рис. 3.

Для сравнения на примере нашего способа с нейросетевым методом был использован пример из интернета [7], в котором приведено решение с помощью нейросети (структура не сообщается) задачи по предсказанию заболевания диабетом женщин из одного индийского племени. Данные взяты из базы данных - репозитория UCI. Классов два. Признаков восемь. Обучающая выборка – 768 точек. Автор [7] сообщает о результате правильных ответов 87,9% без исключения тестовых точек из выборки, т.е. без кросс-валидации. В комментариях приводится полученный результат 73% с кросс-валидацией, но не сказано какой. В нашем случае без кросс-валидации при W=10, S=1 правильные ответы составили 93%. При бОльших W результат будет еще выше, влияние не удаляемых из выборки тестовых точек еще повысится. При кросс-валидации по методу LOO правильные ответы составили 77,7 % при коэффициенте W=0,08. Выбор лучшего значения коэффициента W находился простым перебором. Зависимость количества ошибок от W представлена на Рис. 3. Не исключено, что автор [7] при решении этой задачи нейросетевым методом не использовал все возможности метода.

В работе [8] представлено решение задачи оценки АпоВ классов для дифференциальной диагностикеи трех близких по симптоматике заболеваний. Особенность задачи – малый объем обучающей выборки, всего 36 точек. Без кросс-валидации ошибок не было. С кросс-валидацией методом LOO было 5 ошибок классификации по полученным АпоВ классов.

В интернете на сайте kaggle.com представлена конкурсная задача – предсказание выживаемости пассажиров Титаника при аварии. В конкурсе принимают участие 17000 команд. Решение задачи нашим методом с результатом правильных ответов 79.904% в рейтинге занло 760-е место.

2.2 *Аппроксимация ДФА в окрестности нулевых значений*

При выборе вида аппроксимирующей ДФА зависимости можно использовать тот факт, что ДФА является регрессионной зависимостью от признаков классов, и отбирать используемые в аппроксимации признаки, а так же функции от признаков, по коэффициентам корреляции их с искомой величиной – оценкой ДФА в виде разностей стоимостей ошибок классификации, которые используются для замены номеров (меток классов) при преобразовании обучающей выборки с учителем задачи классификации в выборку задачи регрессионного анализа. Так в примере [6] с размерностью пространства признаков 217 обучающая выборка насчитывала всего 252 строки. Были отобраны три признака с коэффициентом корреляции с оценками искомой величины не ниже 0.64 и со взаимной корреляцией не выше 0.66. Ошибка классификации по аппроксимации ДФА в окрестности нулей по линейной аппроксимации составила 6.8 %, по полиному 2 порядка – 4.8 %. Для сравнения, метод классификации, основанный на гипотезе условных нормальных распределений отобранных признаков классов, составил по полиному и по линейной модели 10.3 % и 9.9 % соответственно. Использованы для этого методы классификации quadratic и linear из библиотеки МАТЛАБ.

В работе [9] приведено сравнение на 15-ти примерах качества решения задач классификации методом аппроксимации ДФА в окрестности нулей и метода опорных векторов для линейной аппроксимации ДФА и для аппроксимации полиномом второго порядка при равных стоимостях ошибок классификации. В 14-ти примерах первый метод дал меньшую ошибку классификации, чем второй. Несколько примеров не могут быть достаточным основанием для суждения о преимуществах одного метода над другим.

2.3 Оценка АпоВ в точке по серии аппроксимаций ДФА

В качестве модельного примера [6] рассматривалась задача классификации с двумя классами с нормальными условными распределениями признаков классов в двумерном пространстве с разными ковариационными матрицами.

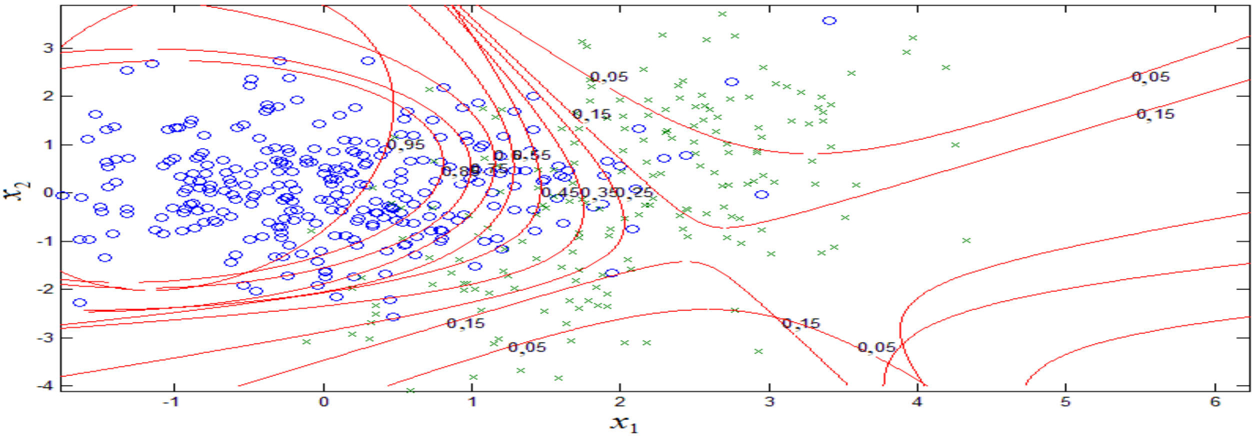

Рис. 4.

Средние значения двух признаков первого класса – $m_1 = (0,\ 0)$, ковариационная матрица – $S_1 = ((1,0);(0,1))$. Для второго класса: $m_2 = (0,\ 2)$, $S_2 = ((4,1);(1,1))$. Априорная вероятность первого класса $P_1 = 0,6$, второго – $P_2 = 0,4$.

По исходным данным и по сгенерированной выборке в 500 точек вычислялись АпоВ следующими способами:

1. точным – по известным распределениям по формуле Байеса;
2. выборочным - по выборочным оценкам параметров распределений с использованием формулы Байеса для вычисления АпоВ первого класса;
3. интер-экстраполяционным по сериям АДФ. АДФ аппроксимировались полиномом второго порядка;
4. методом сглаживания по серии АДФ.

Серия аппроксимаций ДФА, Рис.4, построена для значений p^* от 0,95 до 0,05 АпоВ первого класса на соответствующих границах между классами с шагом 0.1. Кружки – точки первого класса, крестики – второго. Среднеквадратичная ошибка оценки АпоВ первого класса по обучающей выборке интерполяционным методом по серии аппроксимаций ДФА составила 0.055. Ошибка ее оценки методом сглаживания составила 0.06. Ошибка оценки АпоВ первого класса по выборочным оценкам распределений признаков с использованием формул Байеса составила 0.02, т.е. в три раза меньше. Но в оправдание других методов следует помнить, что они не требуют знания законов распределения признаков классов. При малом объеме обучающей выборки затруднительно обоснованно выдвигать гипотезы о законах распределений признаков классов.

Заключение

1. Аппроксимация дискриминантной функции Андерсона по обучающей выборке с учителем позволяет одновременно находить оценки апостериорных вероятностей классов для классифицируемых объектов в машинном обучении. Апостериорные вероятности классов представляют собой исчерпывающую информацию для решения задачи классификации по различным критериям.
2. Обучающая выборка задачи классификации преобразуется в выборку задачи регрессионного анализа путем замены в обучающей выборке номеров (меток) классов на соответствующие разности стоимостей ошибок классификации, задаваемые произвольно на этапе обучения для

определения аппроксимаций дискриминантной функции. На втором этапе решения задачи по полученным по аппроксимациям дискриминантной функции оценкам объективных апостериорных вероятностей классов выполняется собственно классификация объекта по субъективно задаваемому критерию/критериям классификации.

3. Предложены методы решения задачи, связанные с использованием непараметрического и параметрического подходов. При непараметрическом подходе для аппроксимации дискриминантной функции Андерсона в точке пространства признаков классов, представляющей классифицируемый объект, используется обучающая выборка для каждой классифицируемой точки, что может быть неприемлемо при больших объемах выборки и при определенных требованиях к быстродействию. Но достоинством непараметрического подхода, использующего взвешенный метод наименьших квадратов, является то, что в качестве аппроксимирующей функции достаточно использовать зависимость не сложнее линейной.
4. В параметрических подходах также используется взвешенный метод наименьших квадратов либо для получения одной аппроксимации дискриминантной функции Андерсона выброанного вида в пространстве признаков при заданных стоимостях ошибок классификации. Взвешенный метод наименьших квадратов используется для получения более точной аппроксимации функции в окрестности ее нулевых значений. В другом параметрическом подходе строится серия аппроксимаций дискриминантной функции Андерсона для заданного набора апостериорных вероятностей первого класса на границе межу классами. Затем для классифицируемой точки находятся апостериорные вероятности классов методом интер-экстраполяции или методом сглаживания.
5. Методы демонстрируются примерами.

Литература

1. *Platt J.* Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods (PDF). – Advances in large margin classifiers. 10 (3): 61–74.
2. Anderson T. W. An Introduction to Multivariate Statistical Analysis. Third Edition. John Wiley & Sons, 2003. 721 p.
3. *Хардле В.* Прикладная непараметрическая регрессия. – М.: Мир, 1993. – 349с.
4. *Зенков В.В.* Оценка вероятности принадлежности точки классу по аппроксимации одной дискриминантной функции // АиТ. 2018, № 9. С. 46–58.
5. *Зенков В.В.* Использование взвешенного метода наименьших квадратов при аппроксимации дискриминантной функции цилиндрической поверхностью в задачах классификации // АиТ. 2017. № 9. С. 145–158.
6. *Зенков В.В.* Оценка апостериорной вероятности класса по серии дискриминантных функций Андерсона // АиТ. 2019, № 3. С. 68–82.
7. Нейронная сеть на Python в 15 строк кода для диагностики диабета. Библиотека программиста. Обучающие материалы для разработчиков. https://proglib.io/
8. *Зенков В.В. и др.* Оценка апостериорных вероятностей классов в задаче дифференциальной диагностики заболеваний слизистой оболочки полости рта // Тринадцатая международная конференция “Управление развитием крупномасштабных систем” (MLSD’2020). Труды. С.1793-1801, под общей редакцией С.Н.Васильева, А.Д.Цвиркуна.
9. *Зенков В.В.* Применение аппроксимации дискриминантной функции Андерсона и метода опорных векторов для решения некоторых задач классификации // АиТ. 2020, № 1. С. 147–160.