DOI: 10.25728/mlsd.2022.1209

ПРИМЕНЕНИЕ АППРОКСИМАЦИИ ДИСКРИМИНАНТНОЙ ФУНКЦИИ АНДЕРСОНА И АПОСТЕРИОРНЫХ ВЕРОЯТНОСТЕЙ КЛАССОВ ПРИ НЕСБАЛАНСИРОВАННОЙ ОБУЧАЮЩЕЙ ВЫБОРКЕ В МАШИННОМ ОБУЧЕНИИ

Зенков В.В.

Институт проблем управления им. В.А. Трапезникова РАН,

Россия, г. Москва ул. Профсоюзная д.65

zenkov-v@yandex.ru

Аннотация: В машинном обучении при большом дисбалансе классов в обучающей выборке с учителем малый класс может плохо обнаруживаться. Проблема дисбаланса обычно ликвидируется эвристическими методами выравнивания долей классов в обучающей выборке, неприемлемыми для стохастического метода решения задачи классификации. Использование же оценок апостериорных вероятностей классов, получаемых по аппроксимации дискриминантной функции Андерсона, позволяет решать проблему дисбаланса не путем изменения долей классов в выборке, а путем выбора стоимостей ошибок классификации из диапазонов, вычисленных по апостериорным вероятностям в точках малого класса обучающей выборки. Применение этого способа позволяет решать задачу, не выходя за рамки стохастического метода.

Ключевые слова: машинное обучение, несбалансированная обучающая выборка, оценка апостериорной вероятности класса, дискриминантная функция Андерсона, взвешенный метод наименьших квадратов.

Введение

При решении задач классификации стохастическими методами в машинном обучении большую роль играют апостериорные вероятности (АпоВ) классов k, $p(k|x)$, к которым может принадлежать классифицируемый объект с известным вектором признаков x, его характеризующим. Апостериорные вероятности – это название условных вероятностей [1], подчеркивающее, что для классифицируемого объекта вероятности принадлежности к классам получены по результатам его предварительного обследования. Результаты обследования оформляются в виде вектора признаков объекта x. Термин “признаки” является качественным показателем, согласно В.Ю. Кнеллеру [2], и для численного вектора x не совсем уместен, но в машинном обучении термин нашел широкое применение. Да и сам термин машинное обучение не совсем правилен. При поиске работы в интернете по специальности машинное обучение можно получить и приглашение на курсы, например, швеи-мотористки, водителя и т.п. Больше подошел бы термин обучение машин, как считает один из авторов в интернете.

Имея объективные оценки АпоВ классов для отнесения объекта в тот или иной класс, можно принимать решение по классификации объекта с учетом субъективно задаваемых стоимостей ошибок классификации, ограничений, накладываемых на АпоВ классов, применять различные критерии для решения проблемы классификации. Можно выполнять классификацию объектов по-разному, учитывая некоторые особенности объектов, например, пенсионер или состоятельный человек с большим окладом. Процесс классификации, использующий АпоВ классов, состоит таким образом из двух этапов: объективного, в котором находятся оценки АпоВ классов классифицируемого объекта и субъективного, в котором объект относится в класс, удовлетворяющий субъективным условиям и критериям лица, принимающего решение.

Для получения АпоВ классов по обучающей выборке с учителем существуют различные методы. От классических, связанных с получением априорных вероятностей классов P_k и априорных распределений признаков классов $p(x|k)$ и использования формулы Байеса для получения АпоВ классов в заданной точке x пространства признаков [3], до некоторых иных методов. Классические требуют выдвижения гипотез о законах условных распределений признаков классов и достаточно больших объемов обучающих выборок для оценки их параметров. Среди других методов, например, логистическая регрессия дает оценку АпоВ класса в задаче классификации с двумя классами в предположении, что АпоВ класса в пространстве признаков является сигмоидной функцией от вектора x [4]. Такую зависимость имеет АпоВ класса, в частности, при нормальных условных распределениях признаков классов с одинаковыми ковариационными матрицами. Параметры сигмоидной функции находятся методом максимального правдоподобия [5]. В популярном методе опорных векторов [6] строится граница между классами в пространстве признаков классов методом квадратичного программирования с количеством ограничений, равным количеству строк в обучающей выборке. Для получения оценки АпоВ класса необходимо использовать, надстройку для пересчета отступа от границы между классами в АпоВ класса, например, калибратор Платта [7], который в качестве зависимости АпоВ класса от отступа использует сигмоидную функцию и оценивает ее параметры методом максимального правдоподобия. В методе решающих деревьев [8] терминальные вершины можно использовать для оценки АпоВ классов, применив также специальный калибратор [9] для компенсации смешений в оценках.

На фоне сложных манипуляций для получения оценок АпоВ классов в методах машинного обучения весьма привлекательным выглядит использование способа оценки АпоВ классов на базе аппроксимации дискриминантной функции Андерсона (ДФА) [10].

ДФА определяется объективными апостериорными вероятностями классов и субъективно заданными стоимостями ошибок классификации. Она является по определению регрессионной зависимостью, аппроксимируя которую алгебраически элементарно получаются оценки апостериорных вероятностей классов.

Для аппроксимации ДФА используется имеющаяся обучающая выборка задачи классификации с учителем, которую просто преобразовать в выборку регрессионного анализа заменой номеров классов на соответствующие разности заданных стоимостей ошибок классификации.

Из определения ДФА следует тождественная связь АпоВ первого класса с ДФА в задаче с двумя классами

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

где $p(1|x)$ – АпоВ первого класса, $f_{12}(\cdot)$ – ДФА, p^* – стоимость ошибки классификации, если объект второго класса ошибочно классифицируется как объект первого класса, $0<p^*<1$. Стоимость противоположной ошибки есть 1- p^*, эта ошибка имеет место, когда объект первого класса ошибочно относится во второй класс. Из (1) видно, что p^* также есть АпоВ первого класса на границе между классами, там ДФА принимает нулевое значение, если граница эта существует. Граница может отсутствовать при большом неравенстве долей классов в обучающей выборке и неудачном выборе стоимости ошибки p^*, о чем пойдет речь в дальнейшем. Оценка АпоВ класса (1) от существования границы не зависит. **Не зависит также и получаемая оценка АпоВ по ДФА, определяемой произвольно задаваемыми стоимостями ошибок классификации p^***.

Формула пересчета ДФА в АпоВ первого класса (1) не сложнее формулы пересчета температуры, измеряемой в градусах Фаренгейта, в температуру, измеряемую в градусах Цельсия.

И еще одно достоинство использования аппроксимации ДФА состоит в том, что для аппроксимации ДФА используется метод наименьших квадратов (МНК) [11]. Размерность обращаемой матрицы в методе не зависит от количества строк в обучающей выборке. Она в линейном варианте аппроксимирующей функции лишь на единицу больше размерности пространства признаков классов.

Для аппроксимации ДФА методом МНК во всей области обучающей выборки необходимо, как обычно, задаваться видом аппроксимирующей функции. Однако, можно получить хорошее решение задачи классификации и при линейной аппроксимации ДФА, если аппроксимировать ее в окрестности нулевых значений взвешенным МНК [12]. В работе [13] на 15-ти примерах выполнено сравнение применения для решения задач классификации линейной аппроксимации ДФА в окрестности нулевых значений и линейного метода опорных векторов. Из 15-ти примеров в 14-ти аппроксимация ДФА имела меньшую величину эмпирического риска по сравнению с методом опорных векторов. Однако, результаты сравнения на примерах решения задач не могут дать однозначного ответа на вопрос, какой из методов точнее другого.

Аппроксимация ДФА в окрестности нулей перспективна для решения задачи классификации по критерию минимума средней стоимости ошибок классификации, но для оценки АпоВ классов во всей области обучающей выборки она применяться не может. Для оценки АпоВ классов в широкой области следует использовать непараметрический метод оценки ДФА в заданной точке [10], в котором путем взвешенного МНК изо всей выборки по сути выделяется близкая к точке область точек выборки. И для аппроксимации в этом случае достаточно линейной аппроксимирующей функции или средневзвешенной оценки ДФА в точке методом Надарая-Ватсона [14].

В машинном обучении при решении задач классификации объектов по обучающей выборке с учителем, как правило, доли выборок отдельных классов бывают разными. При слишком большом дисбалансе долей классов в обучающей выборке точки малого класса в итоге либо плохо обнаруживаются, либо не обнаруживаются вообще. Такая ситуация возникает и в случае с многими классами даже при близких объемах выборок в классах, если для решения задачи классификации используется, как обычно, прием один класс против всех остальных, объединяемых в другой класс. Однако, дисбаланс в методе один против всех не играет никакой роли в оценке АпоВ этого класса (1), противопоставляемого всем остальным. При решении самой задачи классификации результат определяется по полученным этим методом оценкам АпоВ всех классов и субъективным предпочтениям лица, принимающего решение.

Для решения проблемы дисбаланса классов в обучающей выборке используются эвристические приемы выравнивая долей классов в обучающей выборке [15,16] либо путем различных способов искусственного увеличения малой выборки (пересэмплирование) либо искусственным сокращением большой выборки (недосэмплирование). При этом, очевидно, прерывается связь выборки с генеральной совокупностью, с возможностью решения задачи в рамках стохастической постановки и с возможностью использования АпоВ классов, как для решения задачи классификации в различных постановках, так и для аргументированного объяснения принимаемых решений.

Использование ДФА и тождественно связанных с ней АпоВ классов дает возможность решать несбалансированную задачу классификации в рамках стохастической постановки, сводя ее к проблеме выбора стоимостей ошибок классификации из диапазона, полученного по оценкам АпоВ классов в точках малого класса обучающей выборки. Решая затем байесовы задачи классификации для разных стоимостей ошибок в полученном диапазоне, можно выбрать субъективно приемлемый вариант стоимостей ошибок, которому соответствуют “достаточно хорошие” ошибки в классах. При этом никаких манипуляций с количествами точек классов обучающей выборки делать не приходится.

Вследствие новизны ДФА для удобства читателя начнем с ее определения и ее свойств, позволяющих получать оценки АпоВ классов по ДФА также просто, как пересчитывать градусы Цельсия в градусы Фаренгейта.

1 Дискриминантная функция Андерсона и апостериорные вероятности классов

ДФА в случае с двумя классами мы определяем как разность двух функций средних потерь от ошибок при отнесении точки *x* пространства признаков классов (объекта классификации) к первому и второму классам. Называем эту функцию, как дискриминантную функцию Андерсона, в честь Теодора Уилбура Андерсона, автора простого способа решения байесовой задачи классификации [17]

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

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 относить всегда к классу 1 или, соответственно, к классу 2, C_{ij} –стоимость ошибки, когда точка из класса *j* ошибочно относится в класс $i, C_{ij} > 0$, $C_{ii} = 0$, $p(k\,|\,x)$– 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, $\mathrm{M}_{k|x}(\cdot)$ – математическое ожидание по *k* в точке *x*.

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

Перечислим свойства ДФА, опустив доказательства из-за их очевидности:

1. *ДФА по определению есть функция регрессии. Для преобразования обучающей выборки задачи классификации с учителем в выборку регрессионного анализа нужно заменить в выборке номер класса 1 на -C_{21}, а номер класса 2 на C_{12}.*
2. *ДФА есть ограниченная функция регрессии,* $-C_{21} < f_{12}(x,C) < C_{12}$.
3. *Для АпоВ первого класса и ДФА, определенной для заданной матрицы стоимостей ошибок классификации С, имеет место тождество, вытекающее из (2),*
$$p(1\,|\,x) \equiv (C_{12} - f_{12}(x,C))\,/\,(C_{12} + C_{21}) \qquad (3)$$
4. *Если задавать стоимости ошибок при условии* $C_{12} + C_{21} = 1$, *что не приводит к потере общности, то тождество (3) не только предельно упростится,*
$$p(1\,|\,x) \equiv p^* - f_{12}(x, p^*) \qquad (4)$$
но и стоимости ошибок обретут удобный для интерпретации смысл
$$C_{12} = p^*, \quad C_{21} = 1 - p^* \qquad (5)$$
где p - это также АпоВ первого класса на границе между классами в пространстве признаков, где* $f_{12}(x, p^*) = 0$, *если граница существует при заданной p* и обучающей выборке.*
5. *Результат вычисления АпоВ по (4)* ***не зависит от выбора значения*** *p* и существования границы между классами. АпоВ класса является объективной величиной, в то время как*

параметр p - величина субъективная, произвольно задаваемая для определения ДФА.*
Можно формально показать независимость АпоВ от p*, подставив выражение ДФА (2) в (4), в результате чего исчезнет p*.

2 Аппроксимация ДФА для оценивания АпоВ классов

Для оценивания АпоВ классов проще использовать непараметрический метод аппроксимации ДФА по обучающей выборке с учителем [10]. Этот метод не связан с выбором аппроксимирующей зависимости, достаточно использовать линейную зависимость или средневзвешенную оценку Надарая-Ватсона [14], поскольку аппроксимация производится в заданной точке пространства признаков классов, а не во всей области обучающей выборки.

Аппроксимация ДФА в заданной точке получается взвешенным (МНК). Чем ближе к заданной точке находится точка выборки, тем с большим весом она используется для получения аппроксимации. В качестве весовой функции можно использовать любую функцию, монотонно убывающую с увеличением расстояния до заданной точки. Мы используем экспоненту от расстояния между двумя точками. Параметрами весовой функции являются показатель степени расстояния и множитель, на который умножается расстояние, возведенное в степень.

Критерий взвешенного МНК для получения линейной аппроксимации ДФА в точке выборки x_j, исключаемой из обучающих точек и играющей роль заданной точки по методу борьбы с переобучением LOO (leave-one-out) [19]:
где $(1,\ x_n)$ –вектор-строка кроме строки признаков x_n обучающей выборки содержит компонент–единицу,

$$Q(\lambda_j, x_j) \ = \min_{\lambda_j} \sum\nolimits_{n=1, n\neq j}^{n=N} \{[C_{1k_n} - C_{2k_n} - \ (1, x_n)\lambda_j]^2 \exp(-W \parallel x_j - x_n \parallel^S)\}\ , \qquad (6)$$

N – количество строк в обучающей выборке, строка x_j из выборки исключается, как того требует метод борьбы с переобучением LOO [19],
k_n – номер класса 1 или 2 в строке выборки n,
x_n – вектор признаков в строке выборки n,
$W \geq 0$ – весовой коэффициент - скорость спада весовой функции по мере удаления точки выборки от заданной точки,
S –степень возведения расстояния от точки выборки до заданной точки.

Получив в (6) вектор коэффициентов λ_j, вычисляем оценку ДФА в точке x_j, $f_{12}(x_j) = (1, x_j)\lambda_j$ и по (4) получаем оценку АпоВ первого класса в точке x_j. Если количество классов $K>2$, то аналогично получаем АпоВ всех классов, используя метод один класс против всех остальных, объединяемых в другой класс.

Решение зависит от параметров весовой функции S и W. Эти параметры, лучшие для обучающей выборки, находятся по минимуму по S и W, например, эмпирического риска путем выполнения классификации точек обучающей выборки по полученным оценкам АпоВ классов при очередных значениях S и W. Обычно параметр S принимается равным единице. Находится только лучшее значение параметра W.

Метод является непараметрическим, потому что для аппроксимации ДФА в каждой заданной точке необходимо использовать всю обучающую выборку.

3 Дисбаланс классов

3.1 Случай с двумя классами

Пусть дисбаланс вызван малым количеством точек первого класса обучающей выборки по сравнению с количеством точек второго класса. Метод оптимальной байесовой классификации при неравных стоимостях ошибок $C_{12} = p^* < C_{21}$ отнесет точку x_n в первый класс, если средние потери от этого решения будут меньше, чем потери от решения отнести ее во второй класс (2,5), т.е.

$$p^*(1 - p(1 \mid x_n)) < (1 - p^*)p(1 \mid x_n),$$

что равносильно неравенству

$$p^* < p(1 \mid x_n). \qquad (7)$$

На первом этапе классификации мы получаем оценки АпоВ классов $p(1 \mid x_n)$ и $p(2 \mid x_n) = 1 - p(1 \mid x_n)$ в точках выборки x_n. На втором этапе нужно выбрать стоимость ошибки p^* так, чтобы точки выборки, принадлежащие первому (малому) классу, обнаруживались в достаточном количестве по субъективной мерке.

Если выбрать p^* в (7) из условия

$$p^* < p_{\min x_n : x_n \in X_1}(1 \mid x_n),$$

где X_1 – это множество точек выборки первого класса, то все эти точки будут обнаружены правильно, как точки первого класса. Если же выбирать p^* из диапазона

$$p_{\min x_n : x_n \in X_1}(1 \mid x_n) \le p^* < p_{\max x_n : x_n \in X_1}(1 \mid x_n), \tag{8}$$

то в первый класс будет относиться лишь некая часть точек выборки из множества X_1. Задавая значения p^* из полученного диапазона (8) и имея количества ошибочно обнаруживаемых точек из первого и второго классов, можно выбрать субъективно приемлемое значение p^* для решения задачи классификации по правилу: точку x нужно от носить в первый класс, если $p(1|x) > p^*$.

Поясним это на одном медицинском примере. Нужно было проверить возможность по результатам входных анализов пациента прогнозировать процент поражения легких, измеряемый компьютерным томографом. В обучающей выборке количество пациентов с наиболее угрожающим уровнем поражения 75% было наименьшим. Нужно было классифицировать пациентов на два класса: 75% - первый класс и ниже – второй.

Имелась таким образом несбалансированная обучающая выборка с учителем для оценки принадлежности объектов-пациентов к одному из двух классов по нескольким признакам этих объектов. Количество точек выборки первого класса 8, Второго класса – 127. Для каждой точки обучающей выборки находилась оценка АпоВ первого класса по методу [10]. Далее при заданных равных стоимостях ошибок, т.е. $p^* = 0.5$, решалась задача классификации и подсчитывалось количество точек выборки первого класса, попавших в первый класс. Из 8 точек первого класса в первый класс при классификации попала лишь одна точка. Это было неприемлемо. Поэтому было решено отказаться от равной стоимости ошибок классификации и проверить несколько меньших значений p^* из диапазона (8).

Значения АпоВ первого класса по 8-ми точкам первого класса обучающей выборки лежали в пределах 0.00099901-0.555138. Количества ошибок по двум классам выборки для различных произвольно задаваемых величин p^* из этого предела представлены в Таб. 1.

Таблица 1. Количества ошибок в классах

№	**p***	**Ошибки 1 класса**	**Ошибки 2 класса**
1	0.000999	0	126
2	0.05	2	46
3	0.1	3	27
4	0.2	3	11
5	0.25	4	6
6	0.3	5	6
7	0.4	6	2
8	0.5	7	1
9	0.555	7	1

Лучшим вариантом принят вариант $p^* = 0.2$, при котором 3 точки первого класса выборки ошибочно попали во второй класс, а 11 точек второго класса ошибочно – в первый. Величины ошибок C_{12} и C_{21} не обязательно задавать так, чтобы они в сумме давали единицу. Главное, чтобы различие между ними C_{21}/C_{12} в этом случае было четырехкратным, 0.8 / 0.2.

3.2 Случай с количеством классов больше двух

Если один из классов (первый) представлен значительно меньшим количеством точек в выборке по сравнению с количествами точек других классов, то можно, применив принцип один класс против всех остальных, решить проблему аналогично предыдущему случаю с двумя классами, добавив для простоты решения задачи несколько условий.

Примем стоимости ошибок классификации больших классов в случае ошибочного отнесения точки в первый, малый, класс для всех остальных, вошедших во второй, (большой) класс, одинаковыми $C_{1k} = p^*, k = 2 \div K$. Для ошибок при занесении точки из малого класса в большие классы так же используем одинаковую стоимость $C_{k1} = 1 - p^*, k = 2 \div K$. Стоимости ошибок при неверной классификации между большими классами считаем тоже равными, $C_{jk} = d, \ j > 1, k > 1, \ j \ne k$.

При байесовой классификации точка x будет относиться в первый класс с меньшими средними потерями при условии

$$\sum_{k>1}^{K} C_{1k}\, p(k \mid x) < \min_{j>1} \sum_{k\neq j}^{K} C_{jk}\, p(k \mid x) = \min_{j>1}[C_{j1} p(1 \mid x) + \sum_{k\neq j,k>1}^{K} C_{jk}\, p(k \mid x)],$$

что при сделанных ограничениях равносильно неравенству

$$p^*(1 - p(1 \mid x)) < (1 - p^*)\, p(1 \mid x) + d(1 - p(1 \mid x) - \max_{j>1} p(j \mid x))$$

и после упрощения неравенству

$$p^* < p(1 \mid x) + d(1 - p(1 \mid x) - \max_{j>1} p(j \mid x)). \qquad (9)$$

Для выполнения условия (9) для отнесения точки в первый класс требуется меньшая вероятность $p(1|x)$ за счет второго положительного слагаемого в правой части неравенства, в котором должны быть известны $p(j|x)$. Если же мы примем d=0, или ослабим (9)

$$\max_{j>1} p(j \mid x) < \sum_{j>1}^{K} p(j \mid x) = 1 - p(1 \mid x), \qquad (10)$$

то получим неравенство (7), которое потребует большей, чем в (9), апостериорной вероятности для отнесения точки в первый класс.

Далее, построим вышеприведенную таблицу, считая малай класс первым и выполняя классификацию по правилу (7). При этом количество точек, помещаемых в первый класс, будет несколько заниженным по сравнению с правилом (9) при d>0, а количество ошибок несколько завышенным. По второму классу результат будет противоположным.

По результатам, представленным в таблице, субъективно выбираем p^*.

Заключение

1. Аппроксимация дискриминантной функции Андерсона – это простой и легкий способ оценки апостериорных вероятностей классов по обучающей выборке с учителем. Использование апостериорных вероятностей классов позволяет решать задачу классификации в стохастической постановке в два этапа. На первом, объективном, этапе для объекта классификации по его признакам находятся апостериорные вероятности классов, к которым он может принадлежать. На втором, субъективном, этапе с учетом заданных стоимостей ошибок классификации и прочих условий выполняется собственно классификация объекта.
2. Обучающая выборка считается несбалансированной, если доля точек одного класса настолько меньше доли точек другого/других классов, что объекты малого класса либо субъективно плохо обнаруживаются, либо не обнаруживаются вообще. В таком случае предложено на втором этапе классификации использовать такие стоимости ошибок классификации, при которых объекты малого класса обнаруживаются достаточно эффективно с точки зрения субъекта классификации.
3. Для случая с двумя классами приведен способ расчета по обучающей выборке с учителем диапазона стоимостей ошибок классификации, в пределах которого объекты малого класса по субъективным оценкам становятся достаточно обнаруживаемыми при классификации.
4. Метод не нарушает долей классов в обучающей выборке и позволяет решать задачу в рамках стохастического метода оптимизации, в отличие от методов борьбы с несбалансированностью, использующих эвристические приемы выравнивания долей классов в обучающей выборке.
5. Приведен пример.

Литература

1. *Колмогоров А.Н.* Основные понятия теории вероятностей (Серия: “Теория вероятностей и математическая статистика”). – М.: Физматгиз, 1974. – 120 стр.
2. *Кнеллер В. Ю.* Об определении и специфике автоматического контроля // АиТ. 1962, № 4. С. 509–518.
3. *Гмурман В. Е.* Теория вероятностей и математическая статистика. – М.: Высшее образование, 2005.
4. Логистическая регрессия и ROC-анализ — математический аппарат | Loginom URL: https://loginom.ru/blog/logistic-regression-roc-auc (дата обращения: 13.07.2022).
5. *Магнус Я. Р., Катышев П. К., Пересецкий А. А.* Эконометрика. Начальный курс. – М.: Дело. 2007, – 504 стр. — ISBN 978-5-7749-0473-0.
6. *Вьюгин В.* Математические основы теории машинного обучения и прогнозирования. – М.: МЦМНО. 2013, – 390 стр. — ISBN 978-5-4439-0111-4.
7. *Platt J.* Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods (PDF) // Advances in large margin classifiers. 2000. V. 10. No. 3. P. 61–74.

8. Руководство к использованию деревьев решений в машинном обучении и науке о данных | by Margarita. M | NOP: Nuances of Programming | Medium URL: https://medium.com/nuances-of-programming/руководство-к-использованию-деревьев-решений-в-машинном-обучении-и-науке-о-данных-c10030f05349 (дата обращения: 13.07.2022).
9. *Chawia N.V, Cieslak D.A.* Evaluating probability estimates from decision trees. URL: https://www.researchgate.net/publication/228965727_Evaluating_probability_estimates_from_decision_trees (дата обращения: 13.07.2022).
10. *Зенков В.В.* Машинное обучение. Аппроксимация дискриминантной функции Андерсона и оценка апостериорных вероятностей классов // Труды 14-й Международной конференции «Управление развитием крупномасштабных систем» (MLSD'2021, Москва). С. 1535-1543
11. *Айвазян С. А.* Прикладная статистика. Основы эконометрики. Том 2. – М.: Юнити-Дана, 2001. 432 стр. — ISBN 5-238-00305-6.
12. Использование взвешенного метода наименьших квадратов при аппроксимации дискриминантной функции цилиндрической поверхностью в задачах классификации // АиТ. 2017. № 9, С. 145–158.
13. *Зенков В.В.* Применение аппроксимации дискриминантной функции Андерсона и метода опорных векторов для решения некоторых задач классификации // АиТ. 2020. № 1, С. 147–160.
14. *Хардле В*. Прикладная непараметрическая регрессия. – М.: Мир, 1993.
15. *Дьяконов А.Г.* Дисбаланс классов URL: https://dyakonov.org/2021/05/27/imbalance/ (дата обращения: 13.07.2022).
16. Class-imbalance-problem URL http://www.chioka.in/class-imbalance-problem/ (дата обращения: 13.07.2022).
17. *Anderson T. W.* An Introduction to Multivariate Statistical Analysis. Third Edition. – John Wiley & Sons, 2003.
18. *К.В.Воронцов* Машинное обучение (курс лекций) URL: http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_(курс_лекций,_К.В.Воронцов) (дата обращения: 13.07.2022).