МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ

УНИВЕРСИТЕТ им. Р.Е.АЛЕКСЕЕВА

Институт радиоэлектроники и информационных технологий

Кафедра «Вычислительные системы и технологии»

Реферат

По теме «Деревья принятия решений - критерий среднеквадратичной ошибки»

по дисциплине «Методы и средства обработки сигналов»

ПРОВЕРИЛ:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Авербух М.Л.

СТУДЕНТЫ:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Колотихин А.И

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Гладич Н.А

17-В-2

Работа защищена «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_

С оценкой \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Нижний Новгород

2020

### Ансамбли

Хорошим примером ансамблей считается теорема Кондорсе «о жюри присяжных» (1784). Если каждый член жюри присяжных имеет независимое мнение, и если вероятность правильного решения члена жюри больше 0.5, то тогда вероятность правильного решения присяжных в целом возрастает с увеличением количества членов жюри и стремится к единице. Если же вероятность быть правым у каждого из членов жюри меньше 0.5, то вероятность принятия правильного решения присяжными в целом монотонно уменьшается и стремится к нулю с увеличением количества присяжных.  
N — количество присяжных  
p — вероятность правильного решения присяжного  
μ — вероятность правильного решения всего жюри  
m — минимальное большинство членов жюри, m=floor(N/2)+1  
CNi — число [сочетаний](https://ru.wikipedia.org/wiki/%D0%A1%D0%BE%D1%87%D0%B5%D1%82%D0%B0%D0%BD%D0%B8%D0%B5) из N по i

μ=∑i=mNCNipi(1−p)N−i

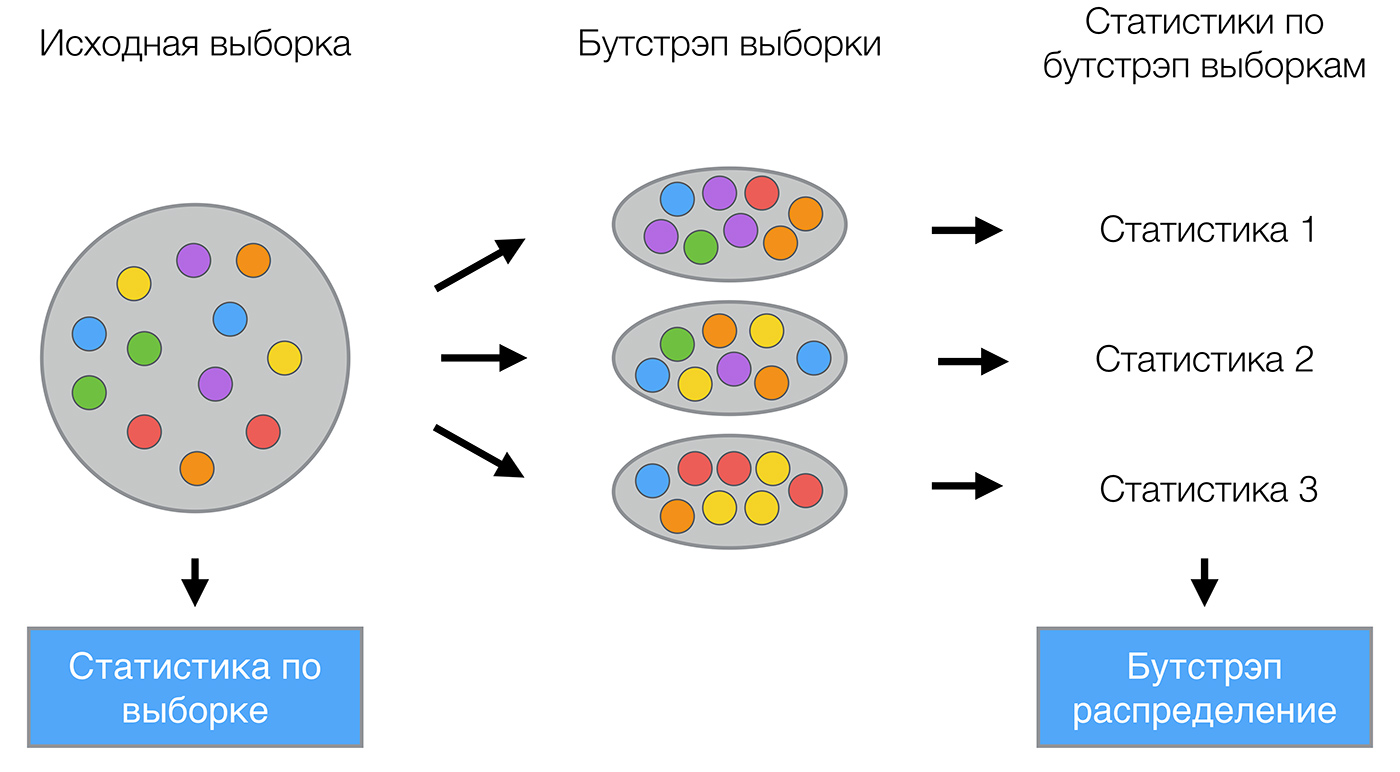
Если p>0.5, то μ>p  
Если N→∞, то μ→1

Давайте рассмотрим ещё один пример ансамблей — "Мудрость толпы". Фрэнсис Гальтон в 1906 году посетил рынок, где проводилась некая лотерея для крестьян.  
Их собралось около 800 человек, и они пытались угадать вес быка, который стоял перед ними. Бык весил 1198 фунтов. Ни один крестьянин не угадал точный вес быка, но если посчитать среднее от их предсказаний, то получим 1197 фунтов.  
Эту идею уменьшения ошибки применили и в машинном обучении.

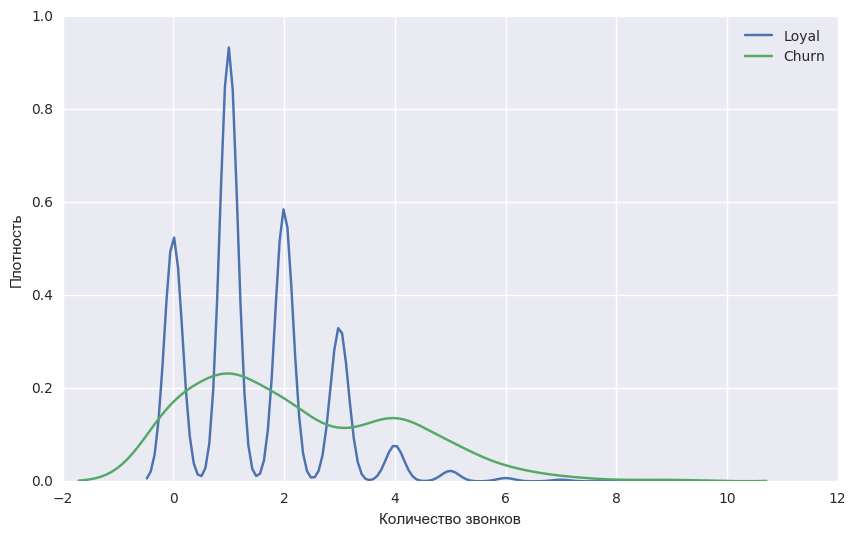
### Бутстрэп

Bagging (от Bootstrap aggregation) — это один из первых и самых простых видов ансамблей. Он был придуман [Ле́о Бре́йманом](https://ru.wikipedia.org/wiki/%D0%91%D1%80%D0%B5%D0%B9%D0%BC%D0%B0%D0%BD,_%D0%9B%D0%B5%D0%BE) в 1994 году. Бэггинг основан на статистическом методе бутстрэпа, который позволяет оценивать многие статистики сложных распределений.

Метод бутстрэпа заключается в следующем. Пусть имеется выборка X размера N. Равномерно возьмем из выборки N объектов с возвращением. Это означает, что мы будем N раз выбирать произвольный объект выборки (считаем, что каждый объект «достается» с одинаковой вероятностью 1N), причем каждый раз мы выбираем из всех исходных N объектов. Можно представить себе мешок, из которого достают шарики: выбранный на каком-то шаге шарик возвращается обратно в мешок, и следующий выбор опять делается равновероятно из того же числа шариков. Отметим, что из-за возвращения среди них окажутся повторы. Обозначим новую выборку через X1. Повторяя процедуру M раз, сгенерируем M подвыборок X1,…,XM. Теперь мы имеем достаточно большое число выборок и можем оценивать различные статистики исходного распределения.



Давайте для примера возьмем уже известный вам набор данных telecom\_churn из прошлых уроков нашего курса. Напомним, что это задача бинарной классификации оттока клиентов. Одним из самых важных признаков в этом датасете является количество звонков в сервисный центр, которые были сделаны клиентом. Давайте попробуем визулизировать данные и посмотреть на распределение данного признака.

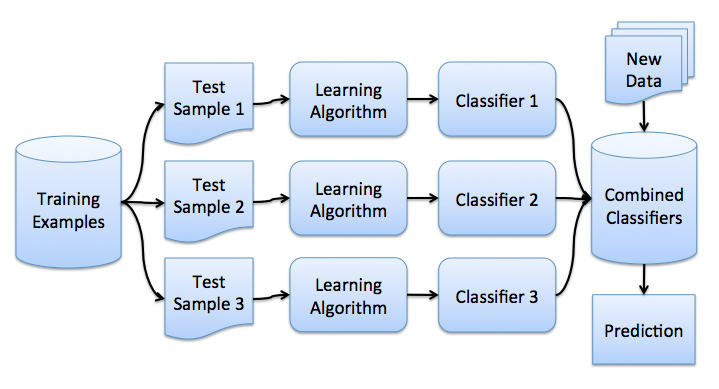


Как вы уже могли заметить, количество звонков в сервисный центр у лояльных клиентов меньше, чем у наших бывших клиентов. Теперь было бы хорошо оценить, сколько в среднем делает звонков каждая из групп. Так как данных в нашем датасете мало, то искать среднее не совсем правильно, лучше применить наши новые знания бутстрэпа. Давайте сгенерируем 1000 новых подвыборок из нашей генеральной совокупности и сделаем интервальную оценку среднего.

В итоге мы получили, что с 95% вероятностью среднее число звонков от лояльных клиентов будет лежать в промежутке между 1.40 и 1.50, в то время как наши бывшие клиенты звонили в среднем от 2.06 до 2.40 раз. Также ещё можно обратить внимание, что интервал для лояльных клиентов уже, что довольно логично, так как они звонят редко (в основном 0, 1 или 2 раза), а недовольные клиенты будут звонить намного чаще, но со временем их терпение закончится, и они поменяют оператора.

### Бэггинг

Теперь вы имеете представление о бустрэпе, и мы можем перейти непосредственно к бэггингу. Пусть имеется обучающая выборка X. С помощью бутстрэпа сгенерируем из неё выборки X1,…,XM. Теперь на каждой выборке обучим свой классификатор ai(x). Итоговый классификатор будет усреднять ответы всех этих алгоритмов (в случае классификации это соответствует голосованию): a(x)=1M∑i=1Mai(x). Эту схему можно представить картинкой ниже.



Рассмотрим задачу регрессии с базовыми алгоритмами b1(x),…,bn(x). Предположим, что существует истинная функция ответа для всех объектов y(x), а также задано распределение на объектах p(x). В этом случае мы можем записать ошибку каждой функции регрессии

εi(x)=bi(x)−y(x),i=1,…,n

и записать матожидание среднеквадратичной ошибки

Ex(bi(x)−y(x))2=Exεi2(x).

Средняя ошибка построенных функций регрессии имеет вид

E1=1nEx∑i=1nεi2(x)

Предположим, что ошибки несмещены и некоррелированы:

Exεi(x)=0,Exεi(x)εj(x)=0,i≠j.

Построим теперь новую функцию регрессии, которая будет усреднять ответы построенных нами функций:

a(x)=1n∑i=1nbi(x)

Найдем ее среднеквадратичную ошибку:

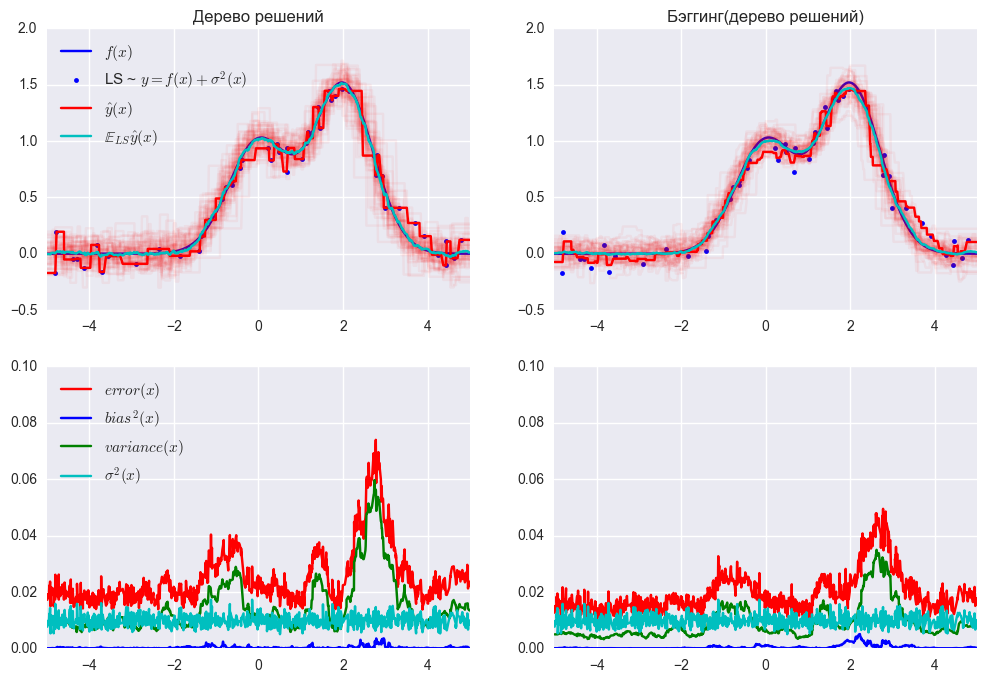
En=Ex(1n∑i=1nbi(x)−y(x))2=Ex(1n∑i=1nεi)2=1n2Ex(∑i=1nεi2(x)+∑i≠jεi(x)εj(x))=1nE1

Таким образом, усреднение ответов позволило уменьшить средний квадрат ошибки в n раз!

Напомним вам из нашего предыдущего [урока](https://habrahabr.ru/company/ods/blog/323890/#razlozhenie-oshibki-na-smeschenie-i-razbros-bias-variance-decomposition), как раскладывается общая ошибка:

Err(x→)=E[(y−f^(x→))2]=σ2+f2+Var(f^)+E[f^]2−2fE[f^]=(f−E[f^])2+Var(f^)+σ2=Bias(f^)2+Var(f^)+σ2

Бэггинг позволяет снизить дисперсию (variance) обучаемого классификатора, уменьшая величину, на сколько ошибка будет отличаться, если обучать модель на разных наборах данных, или другими словами, предотвращает переобучение. Эффективность бэггинга достигается благодаря тому, что базовые алгоритмы, обученные по различным подвыборкам, получаются достаточно различными, и их ошибки взаимно компенсируются при голосовании, а также за счёт того, что объекты-выбросы могут не попадать в некоторые обучающие подвыборки.



Ошибка дерева решений

0.0255(Err)=0.0003(Bias2)+0.0152(Var)+0.0098(σ2)

Ошибка бэггинга

0.0196(Err)=0.0004(Bias2)+0.0092(Var)+0.0098(σ2)

По графику и результатам выше видно, что ошибка дисперсии намного меньше при бэггинге, как мы и доказали теоретически выше.

Бэггинг эффективен на малых выборках, когда исключение даже малой части обучающих объектов приводит к построению существенно различных базовых классификаторов. В случае больших выборок обычно генерируют подвыборки существенно меньшей длины.

Следует отметить, что рассмотренный нами пример не очень применим на практике, поскольку мы сделали предположение о некоррелированности ошибок, что редко выполняется. Если это предположение неверно, то уменьшение ошибки оказывается не таким значительным. В следующих лекциях мы рассмотрим более сложные методы объединения алгоритмов в композицию, которые позволяют добиться высокого качества в реальных задачах.

**Критерий - среднеквадратичная ошибка**

*Критерий среднеквадратичной ошибки* широко используется для систем, находящихся под воздействием стационарных случайных процессов, но он учитывает ошибку лишь в установившемся режиме.

Оптимальной передаточной функцией при использовании *критерия среднеквадратичной ошибки* ( СКО) является такая передаточная функция системы, при которой средне квадр этическая ошибка имеет минимум.

Задача идентификации сводится к определению коэффициентов разложения ядер, минимизирующих *критерий среднеквадратичной ошибки* между выходами объекта и модели. Применение в качестве изучающего сигнала, сигнала, близкого по своим свойствам к нормальному белому щуму, позволяет значительно упростить алгоритм идентификации и повысить его точность.

Преобразование Карунена - Лоэва является оптимальным преобразованием для представления сигналов по отношению к *критерию среднеквадратичной ошибки*.

Винер в качестве критерия использовал минимизацию среднеквадратичной ошибки, причем, Винер указал , что возможны другие критерии, но математическая трактовка упрощается при использовании *критерия среднеквадратичной ошибки*. Минимизация среднеквадратичной ошибки соответствует минимизации мощности сигнала ошибки, и если дальнейшее кодирование не производится, то это - разумный критерий для случая кодирования с предсказанием. Однако в схеме, приведенной на рис. 1, сигнал ошибки до передачи кодируется и его мощность может быть существенно изменена в процессе кодирования. Больше всего мы заинтересованы в минимизации емкости канала, которая потребуется для передачи сигнала ошибки.

Рассмотрим также характер использовавшихся нами критериев. Как *критерий среднеквадратичной ошибки*, так и критерия разброса представляют собой математические ожидания некоторых квадратичных функций переменных.

Ясно, что в смысле второй постановки линейное безынерционное преобразование бесцельно, в то время как улучшение приближения в смысле среднеквадратичной ошибки в этом случае оказывается значительным. Поэтому в задачах, где существенно выделение формы полезного сигнала, *критерий среднеквадратичной ошибки* может оказаться неприемлемым.

При передаче информации др. видов более удобными для оценки помехоустойчивости оказываются др. вероятностные критерии, в частности *критерий среднеквадратичной ошибки* на выходе системы при заданном О. Такой критерий используется при анализе систем автоматич. В ТУ широко пользуются критериями подавления и возникновения ложной команды, н др. случаях - иными критериями оптимальности. При применении этих критериев необходимо определять более сложные статистич. Все эти критерии находятся при определенном О.

При передаче информации др. видов более удобными для оценки помехоустойчивости оказываются др. вероятностные критерии, в частности *критерий среднеквадратичной ошибки* на выходе системы при заданном О. Такой критерий используется при анализе систем автоматич. В ТУ широко пользуются критериями подавления и возникновения ложной команды, и др. случаях - иными критериями оптимальности. При применении этих критериев необходимо определять более сложные статистич. Все эти критерии находятся при определенном О.

Предполагается, что вероятность попадания входного значения х на границы многогранников Вороного пренебрежимо мала. И) Множитель 1 / п вводится для того, чтобы честно сравнивать квантизаторы разных размерностей, ( iii) *Критерий среднеквадратичной ошибки* - это лишь один из многих возможных способов измерить искажение; его преимущества состоят в широкой употребительности и математической простоте. При применениях к обработке речи и изображения правильный выбор меры искажения является трудной задачей.

Степень искажений принятого сигнала обычно оценивают по верности его воспроизведения. Под верностью часто понимают среднеквадратичную ошибку, отличающую принятый сигнал от переданного. Необходимо подчеркнуть, что *критерий среднеквадратичной ошибки* является информационно неверным. Принципиально более правильно оценивать искажения сигнала как появление ложной, ошибочной информации, не поддающейся устранению с помощью коррекции. Искажения, поддающиеся коррекции, как было показано, не изменяют количества информации, а потому здесь не рассматриваются.

В книге коллектива американских авторов под редакцией Джеймса, Ни-колса и Филипса Теория следящих систем, выпущенной в 1947 г., дан метод построения следящих систем на базе заданного показателя колебательности, а также на основе *критерия среднеквадратичной ошибки*, предложенного в СССР А. А. Харкевичем в 1937 г. и в США Холлом в 1943 г. Основываясь на идеях А. Н. Колмогорова, высказанных в 1941 г. и развитых им в 1949 г., Винер разрабатывает метод наилучшего линейного фильтра, удовлетворяющего критерию наименьшей среднеквадратичной ошибки.

В книге коллектива американских авторов под редакцией Джеймса, Ни-колса и Филипса Теория следящих систем, выпущенной в 1947 г., дан метод построения следящих систем на базе заданного показателя колебательности, а также на основе критерия среднеквадратичной ошибки, предложенного в СССР А. А. Харкевичем в 1937 г. и в США Холлом в 1943 г. Основываясь на идеях А. Н. Колмогорова, высказанных в 1941 г. и развитых им в 1949 г., Винер разрабатывает метод наилучшего линейного фильтра, удовлетворяющего *критерию наименьшей среднеквадратичной ошибки*.

Данный результат позволяет глубже проникнуть в суть процедуры, обеспечивающей решение по методу наименьшей квадратичной ошибки. Аппроксимируя g0 ( x), разделяющая функция а у дает непосредственную информацию относительно апостериорных вероятностей Р ( со. К сожалению, *критерий среднеквадратичной ошибки* в основном распространяется не на точки, близкие к поверхности решения go ( x) 0, а на точки, для которых значение р ( х) велико. Таким образом, разделяющая функция, которая наилучшим образом аппроксимирует разделяющую функцию Байеса, не обязательно минимизирует вероятность ошибки. Несмотря на данный недостаток, решение по методу наименьшей квадратичной ошибки обладает интересными свойствами и широко распространено в литературе.

В данной главе показано, что ортогональные преобразования можно использовать для сжатия данных. Показано, что ПКЛ является оптимальным преобразованием для сжатия данных по отношению к *критерию среднеквадратичной ошибки*.

Важной областью применения ортогональных преобразований является сжатие данных. Если дискретный сигнал содержит N отсчетов, то его можно рассматривать как точку Af-мерного пространства. Тогда каждый отсчет является координатой Af-мерного вектора данных X, который представляет собой сигнал в этом пространстве. Для более эффективного представления можно осуществить ортогональное преобразование X, что приводит к YTX, где Y и Т - вектор коэффициентов преобразования и матрица преобразования соответственно. Целью сжатия данных является выбор подмножества М координат вектора Y, где М существенно меньше N. Остальные ( N - М) координат можно отбросить, не вызывая существенной ошибки при восстановлении сигнала по М координатам вектора Y. Следовательно, сравнивать ортогональные преобразования следует в соответствии с некоторым критерием ошибки. Одним из часто используемых критериев является критерий среднеквадратичной ошибки.  **[**