# Слайд: Специфика вероятности переобучения правил

Обозначения комбинаторного подхода и само определение вероятности переобучения требуют обобщения на случай логических закономерностей. Причина этого – в наличии ошибок первого и второго рода, т.е. числа покрытых чужих, и числа не-покрытых своих объектов.

Первое и довольно очевидное отличие заключается в том, что нам необходимо прогнозировать две вероятности переобучения, отдельно для вероятности большого уклонения частот ошибок первого и второго рода. Мы будем соответственнно называть их ***вероятностью переобучения первого и второго рода.***

Менее очевидное отличие заключается в методе обучения. Классически в комбинаторной теории рассматривается метод минимизации эмпирического риска – то есть, минимизации числа ошибок на обучении. Очевидно, что для логических закономерностей необходимо уточнить, о каком числе ошибок идёт речь (n, n’ или n’’). Однако все три перечисленных варианта плохо соответствуют действительности, поскольку на практике обучение производится с помощью максимизации критерия информативности. Поэтому для анализа вероятности переобучения (как первого, так и второго рода) для критерия максимизации информативности в комбинаторной теории был разработан соответствущий математический аппарат. Это позволило обобщить многие результаты полученные ранее для минимизации эмпирического риск, на более широкий класс монотонных критериев обучения.

[ToDo] На слайде нужна ссылка на на статью Воронцова про монотонные методы обучения, что бы отослать туда заинтересовавшихся слушателей.

# Слайд: Оценки вероятности переобучения

Опишем общие соображения, из которых строятся оценки вероятности переобучения первого и второго рода для правил.

Для произвольного правила $r$ определим множество $\XX\_r$ как множество объектов, на которых правило ошибается.

Введем отношение частичного порядка на множестве правил. Для произвольной пары алгоритмов $v, r$ будем говорить, что **v монотонно-лучше r**, если множество ошибок $v$ вкладывается в множество ошибок $r$. В таких случаях объекты $X\_r \backslash X\_v$ будем называать **разностью** между $v$ и $r$.

Рассмотрим метод минимизации эмпирического риска. Следующая теорема [см слайд] даёт условие на разбиение выборки, при котором из пары алгоритмов минимизация эмпирического риска выбирает строго более плохой $r$. Заметим, что раз $r$ более плохой, то он может быть выбран методом мтолько в том случае, когда на обучающей выборке алгоритмы оказались **неразличимыми** с точки зрения числа ошибок. Это значит, что **разность** лежит в контрольной выборке.

[Пауза – имхо в этом месте нужно дать слушателям подумать, хотя бы 5-10 секунд]

[выбрать из двух альтернатив; мне больше нравится вторая – которая простая]

[Альтернатива 1, сложная] : В предыдущем рассуждении в качестве $v$ можно взять любой монотонно более хороший алгоритм. Объединяя разности $X\_r \backslash X\_v$ по всем таким $v$, получается множество объектов, которое зависит только от алгоритма $r$. Это множество в точности удовлетворяет условию метода порождающих и запрещающих множеств [ToDo: опять нехватает ссылки, что бы отправить туда читетелей], воспользовавшись которой можно получить следующую оценку для вероятности переобучения: [ткнуть в слайд]

[Альтернатива 2, простая]

Таким образом, для каждого правила $r$ учёт его связей со всеми строго-более-хорошими алгоритмами накладывает соответствующие ограничения на условие «алгоритм $r$ получается в результате обучения». Это позволяет существенным образом уточнить оценку вероятности переобучения.

# Слайд: Модифицированный критерий информативности

На прошлом слайде мы научились учитывать связи между алгоритмам и получили оценку вероятности переобучения (как первого, так и второго рода). Обозначим данные оценки через эта-штрих и эта-два-штриха. Рассматривая обе эты как функции от eps, построим обратные функции eps’(eta’) и eps’’(eta’’), и возьмём их в точке 0.5. Для точной оценки вероятности переобучения это соответствовало бы медиане уклонения частоты по всем разбиениям.

Напомним, что по своему «физическому смыслу» eps’(0.5) и eps''(0.5) это среднее уклонение частоты, деленное на длину обучающей выборки. Это соображение позволяет выписать модифицированный критерий информативности в следующем виде.

# Слайд: Схема эксперимента

На прошлых слайдах мы описали, как для фиксированного множества алгоритмов настроить поправки на переобучение eps’(0.5) и eps''(0.5), и как с помощью этих поправок подправить критерий информативности для учёта эффекта преобучения правил. Теперь пришло время проверить, как данный подход работает на реальных данных. Для этого был поставлен следущий эксперимент:

Выбрано 7 задач из репозитория UCI. Для каждой задачи эксперимент состоял из двух больших шагов: шаг первый - настройка поправок, шаг второй – настройка алгоритма классификации с помощью поправок, и оценка его качества.

На первом шаге настройка поправок проозводиась для всех троек признаков, и отдельно для правил первого/второго класса. Для каждой тройки признаков запускался тот же алгоритм поиска правил, что и на каждой итерации алгоритма построения классификатора.

В качестве алгоритма построения классификатора использовался Committee Boosting, с небольшой модификацией: на каждой итерации алгоритму поиска правил запрещалось использовать 30% случайно-выбранных признаков.

Сам алгоритм поиска правил ... [далее по слайду]