**Качество и устойчивость алгоритмов обучения**

План занятия

1. *Общие положения и обозначения*.

При обучении машин существенным моментом является оценивание характеристики обобщения (generalization performance) различных алгоритмов обучения. При заданной тренировочной последовательности важно определить, насколько хорошо конкретный алгоритм обучения будет работать на новых данных, т.е. какова его способность к обобщению. Располагая подобной информацией, можно решать вопрос о выборе алгоритма, модели, значений параметров обучения.

На практике при оценивании характеристики (или ошибки) обобщения используют различные **стратегии перевыбора**, при этом обычно применяется процедура *k*-кратной перекрестной проверки (*k*-fold cross-validation (CV)), и как самостоятельная стратегия, и в составе более сложных стратегий. Наиболее популярной разновидностью CV-ошибки является так называемая **leave-one-out (loo) ошибка** (скользящий контроль), получаемая после применения *l*-кратной перекрестной проверки, где *l* – объём тренировочной последовательности.

Loo-ошибка как **оценка ошибки обобщения** (generalization error) является важной статистической **оценкой качества** алгоритмов обучения. В работах [1, 2, 3] приведены результаты, позволяющие обосновать использование *loo*-ошибки при построении машин.

Понятие **устойчивости алгоритма обучения** формально связывает *loo*-ошибку и ошибку обобщения, например, позволяя получатьверхние границы для ошибки обобщения с использованием *loo*-ошибки.

Введем обозначения.

Пусть *X* – пространство входных объектов ***x*** с фиксированным неизвестным распределением вероятностей *P*(***x***);

*Y* – пространство выходных объектов *y* с фиксированным неизвестным распределением вероятностей *P*(*y*/***x***).

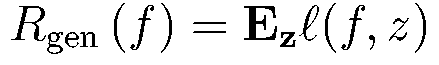
*D* – тренировочная последовательность (ТП) объема *l*, т.е. выборка независимых, одинаково распределенных согласно закону *P*(***x***,*y*) *= P*(***x***)*×P*(*y*/***x***) наблюдений (***x***1, *y*1), … , (***x****l*, *yl*).

*Di* – тренировочная последовательность объема *l*–1,полученная из *D* путем удаления *i*-го образца (***x****i*, *yi*).

*fD* : *X*→*Y* – решение, полученное с использованием данногоалгоритма обучения на основе выборки *D*. Алгоритм обучения называется **симметричным**, если *fD* не зависит от перестановки элементов ТП.

Пусть функция потерь *L* : *R*×*Y*→*R*, *L*( *f*(***x***), *y*) ≡ *L*( *f*, (***x****,y*)) *L*( *f*, *z*), штрафует отклонение оценок *f*(***x***) от наблюдаемых *y*;  *z* = (***x***,*y*).

В качестве **ошибки обобщения** данногоалгоритма обучения относительно функции потерь *L* (∙) рассматривают **ожидаемый риск**

. []

В качестве эмпирической ошибки данногоалгоритма обучения относительно функции потерь *L* можно использовать как **эмпирический риск**

,

так и **leave-one-out*-*ошибку**

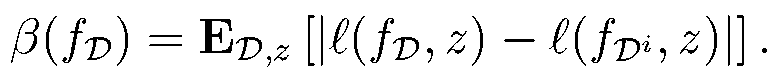
.

Заметим, что в последнем случае образец (***x****i*, *yi*) не используется в процессе обучения, на нем тестируется решение, полученное на основе выборки *Di* объема *l*–1.

2. *Понятия устойчивости алгоритмов обучения*.

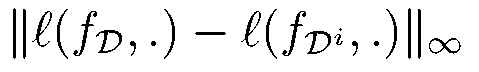
Будем рассматривать **устойчивость симметричного алгоритма обучения** относительно изменений в тренировочной последовательности (ТП), связанных с удалением одного образца.

2.1. Устойчивость гипотезы.



[ См. [2] – п. 3: Определение 3 (стр. 503)]

2.2. Равномерная устойчивость.

по *D*

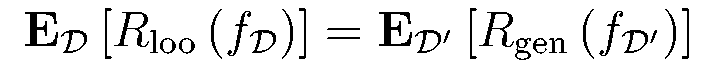
[ См. [2] – п. 3: Определение 6 (стр. 504)]

Из равномерной устойчивости алгоритма следует устойчивость гипотезы.

Равномерную устойчивость можно рассматривать как функцию от объёма тренировочной последовательности *l*. Говорят, что алгоритм обучения устойчив, если его равномерная устойчивость убывает как 1/*l*.

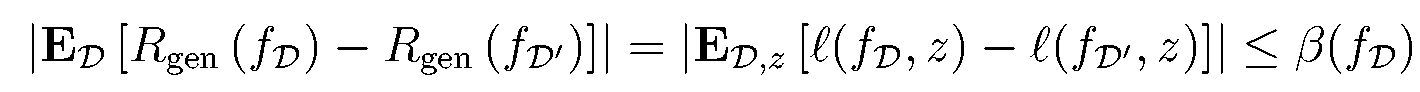
3.  *Cвойства loo-ошибки как статистической оценки ошибки обобщения*.

1) В отличие от эмпирического риска ***l*oo-ошибка почти не смещена** в следующем смысле:

 (Теорема Luntz and Brailovsky).

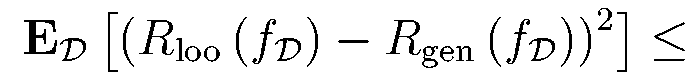
*loo*-ошибка, основанная на выборках объема *l*, дает несмещенную оценку ошибки обобщения после обучения по *l*–1 образцу.

2) Смещение *loo*-ошибки по модулю ограничено устойчивостью гипотезы алгоритма обучения

 .

3) Теорема Devroye et al (для классификаторов).

Пусть 0*L*( *f*, *z*) M , тогда :

  .

Для получения малой дисперсии *loo*-ошибки устойчивость гипотезы алгоритма обучения должна быть мала.

4) Вычисление *loo*-ошибки требует больших временных затрат, поэтому многие исследователи пытались получить для нее верхние границы.

Верхние границы для *loo*-ошибки, предполагающие построение **только одной** машины опорных векторов на основе исходной тренировочной последовательности объёма *l*, приведены в работах [1,3].

Для задачи бинарной классификации с использованием функции потерь *θ*(–*y*⋅*f*(***x***)), где *θ*(⋅) – функция Хевисайда, *loo-*оценка ядерной машины *fD* , которая минимизирует регуляризованный риск на основе тренировочной последовательности *D*, ограничена сверху:

 .

[ См.: [1] – п. 6.4 .]

В работе [3] предложена верхняя граница leave-one-out*-*оценки ошибки обобщения для стандартного SV-классификатора (оценка Joachims’а):

.

Здесь: α*i* − множитель Лагранжа, ξ*i* − ослабляющая переменная, соответствующая *i-*му образцу;  − верхняя граница разности *k*(***x***,***x***)–*k*(***x***,***x***′) для всех допустимых входных векторов ***x***, ***x***′, где *k* – ядерная функция, (в случае линейного и гауссовского ядра =1). Доказано, что в среднем эта оценка превышает истинную долю ошибок классификации.

5. *Верхние границы для ошибки обобщения (ожидаемого риска)*.

Различные понятия устойчивости приводят к различным верхним границам для ошибки обобщения в терминах как эмпирического риска, так и *loo*-ошибки.

Выбор меры устойчивости определяется желаемым типом границы: устойчивость гипотезы ведёт к чебышёвскому типу границ; равномерная устойчивость – к экспоненциальному типу границ.

(\*)

[См.: [2] – п. 4: лемма 7; теоремы 11, 12. ]

6. *Устойчивость алгоритмов регуляризации*.

Такие алгоритмы, какSVM,минимизирующие некоторую регуляризованную целевую функцию (алгоритмы регуляризации), имеют следующее важное свойство: их равномерная устойчивость контролируется параметром регуляризации.

(\*\*)

Это обстоятельство даёт возможность получать верхние границы ожидаемого риска, явно не зависящие от меры сложности класса функций, из которого выбирается решение. Такие верхние границы ожидаемого риска **зависят явно от параметра регуляризации**.

Таким образом, **выбор модели можно рассматривать как процесс выбора подходящего уровня сложности для оценивающей функции**.

[См.: [2] – п. 5.2: определение 19 (σ-допустимая функция потерь) стр. 512; теорема 22 (стр. 514); примеры верхних границ ожидаемого риска, не зависящих явно от меры сложности класса функций, из которого выбирается решение, а зависящих явно от параметра регуляризации.]

7.*Оценка качества бинарной классификации*.

Качество SV-классификатора, естественно, может быть исследовано и с использованием ROC-анализа.

ROC-анализ (Receiver Operator Characteristic) представляет собой графический метод для представления результатов бинарной классификации при машинном обучении, отражающий качество данного классификатора или сравнительную эффективность нескольких классификаторов.

ROC-кривые весьма полезны в случае асимметричного распределения входных векторов и неодинаковой цены ошибок классификации. Эти показатели становятся особенно важными при несбалансированных классах.

Чтобы определять качество классификации отдельно для каждого класса (один класс – класс с положительными образцами, второй – с отрицательными образцами), используют показатели чувствительности и специфичности.

**Чувствительность** отражает эффективность работы бинарного классификатора и определяет долю истинно положительных наблюдений относительно количества фактически положительных. Классификатор, обладающий высокой чувствительностью, обеспечивает большую вероятность правильного распознавания положительных образцов.

**Специфичность** отражает точность работы бинарного классификатора и определяется как отношение истинно отрицательных наблюдений к числу фактически отрицательных. Модель, обладающая высокой специфичностью, обеспечивает большую вероятность правильного распознавания отрицательных образцов.

Практическое руководство по корректному использованию ROC-кривых и основных показателей качества классификации можно найти в работе [4].

Литература

1. Elisseeff A., Pontil M. Leave-one-out Error and Stability of Learning Algorithms with Applications. In Advances in Learning Theory: Methods, Models and Applications. Ed. by Suykens J.A.K. et al., Chapter 6, 2003. P. 111.
2. *Olivier Bousquet, Andre Elisseeff*. Stability and Generalization. Journal of Machine Learning Research **2** (2002) 499-526.
3. *Joachims T.* Estimating the Generalization Performance of a SVM Efficiently. [International Conference on Machine Learning - ICML](http://academic.research.microsoft.com/Conference/35/icml-international-conference-on-machine-learning) , 2000. P. 431.
4. *Fawcett T.* ROC Graphs: Notes and Practical Considerations [for](http://www.basegroup.ru/glossary_ajax/definitions/hypersurface) Researchers. Kluwer Academic Publishers, 2004. P. 1.

Р**егуляризованный риск**:



Параметр регуляризации λ > 0 согласовывает малую эмпирическую ошибку со степенью гладкости (сложности) решающей функции. Параметр регуляризации *C*  для стандартного SVM-обучения связан с λ соотношением *C* = 1/(2*l* λ).

Показано, что функция, минимизирующая регуляризованный риск, единственна и представляет собой **ядерную машину**, т.е. имеет вид: , где вектор весов признаков **β**∈*Rl.*