Задание номер 3

Н. К. Животовский
nikita.zhivotovskiy@phystech.edu

27 марта 2017 г.

Задание принимается до 2.00 утра 17 апреля по адресу slt.fupm.2017@gmail.com. Не забываем, что в начале текста задания *обязательно* указывается:

- С кем вы делали это задание.
- Какие источники (кроме материалов лекций) вы использовали.

Задание оформляется в формате pdf (текст набирается в latex/Word) и в таком виде, чтобы ваши коллеги могли разобрать текст решения. Задания, оформленные не в соответствии с указанными правилами, не принимаются. Желательно оставлять зазоры между задачами для пометок.

Упражнение 1. Рассмотрим бесшумную задачу классификации. Пусть $\hat{f}_S$ — правило классификации, полученное по выборке S, а S^i — выборка с удаленным i-ым объектом. Определим ошибку Leave-one-out

$$LOO(S, n) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{I}[\hat{f}_{S^i}(X_i) \neq f^*(X_i)].$$

Рассчитав математическое ожидание данной величины, получите оценку обобщающей способности SVM в разделимом случае, выраженную через среднее число опорных векторов.

Упражнение 2. [Оценка Полларда]

- Докажите, что для любого множества $A \subset \{0, 1\}^n$ выполнено

$$R_n(A) \leq \inf_{\alpha\in[0,1]}\left(\alpha + \sqrt{\frac{2\log(\mathcal{N}(A, \alpha))}{n}}\right).$$

- Подставьте оценку числа покрытий для случая конечной VC размерности и сравните с результатом, полученным с помощью интеграла Дадли.

Указание. В этом упражнении все обозначения как в лекциях.

■

Упражнение 3. [Теорема Дворецкого-Кифера-Вольфовитца] Пусть $F(x)$ — функция распределения, а $F_n(x)$ — эмпирическая функция распределения. Докажите, что для некоторых абсолютных констант $c_1, c_2 > 0$ для $t > 0$ выполнено

$$P(\sup_{x \in \mathbb{R}} |F(x) - F_n(x)| \geq t) \leq c_1 \exp(-c_2 n t^2).$$

Указание. Нужно применить полученные нами VC оценки.

■

Упражнение 4. [Semi-supervised онлайн модель] Рассмотрим модель онлайн обучения без шума, но с дополнительным предположением, что мы всегда заранее знаем все точки, которые необходимо будет классифицировать (но не их лейблы). Обучаемость определим стремлением к нулю в худшем случае отношения суммарного числа ошибок за n шагов к числу шагов. Докажите, что в такой постановке обучаемость класса $\mathcal{F}$ эквивалентна конечности VC размерности (а не размерности Литтлстоуна).

Упражнение 5. Рассмотрим задачу классификации с классами $\{1, -1\}$ и бинарной функцией потерь. Обозначим $\eta(x) = \mathbb{E}[Y|X = x]$ и $f^*(x) = \operatorname{sign}(\eta(x))$, а $(\ell \circ \mathcal{F})^*$ обозначает класс избыточных потерь.

- Докажите, что для всех классификаторов f имеет место соотношение (мы использовали его на лекции):

$$L(f) - L(f^*) = \mathbb{E}\left(|\eta(X)|\mathbf{I}[f(X) \neq f^*(X)]\right).$$

- Зафиксируем $\alpha \in [0, 1]$. Докажите эквивалентность двух условий:

  1. Существует константа B, такая что для всех $g \in (\ell \circ \mathcal{F})^*$ выполнено

  $$\mathbb{E}g^2 \leq B(\mathbb{E}g)^\alpha.$$

  2. Существует константа β, такая что для всех $t \geq 0$ выполнено

  $$P(|\eta(X)| \leq t) \leq \beta t^{\frac{\alpha}{1-\alpha}}.$$

  Любое из двух условий называется условием малого шума Цыбакова. Эти условия являются обобщением условий Массара.

Упражнение 6. [Выбор самого живучего классификатора] Пусть $\mathcal{F}$ — семейство классификаторов. Пусть также известно, что имеется консервативный онлайновый алгоритм делающий не более M ошибок на любой конечной выборке. Предположим, что получена $i.i.d.$ выборка $(X_i, f(X_i))$ длины n для некоторого неизвестного $f \in \mathcal{F}$. Запустим последовательно онлайн алгоритм на элементах выборки и выберем классификатор $\hat{f}$ (получаемый с помощью онлайн алгоритма), который дольше всех последовательно не ошибался на выборке. Докажите, с вероятностью $1-\delta$ выполнено

$$P(\hat{f}(X) \neq f(X)) \leq \left(\frac{M\log(M) + M\log(\frac{1}{\delta})}{n}\right),$$

где C — некоторая абсолютная константа.

Задача 1. [Essential support vectors] Докажите, что в линейно разделимом случае в $\mathbb{R}^d$ для любого набора опорных векторов можно выделить не более $d+1$ *необходимых* опорных векторов, так что результат применения SVM к ним точно такой же как и ко всей выборке. Выведите оценку обобщающей способности для SVM в линейно разделимом случае.

Задача 2. [Обучение конечных классов в условиях Цыбакова] Докажите, что в условиях Цыбакова (См. упражнение 4) для конечного класса $\mathcal{F}$ классификаторов для минимизатора эмпирического риска с вероятностью $1-\delta$ выполнено

$$L(\hat{f}) - L(f^*) \leq C\left(\frac{\log(N)}{n} + \frac{\log(\frac{1}{\delta})}{n}\right)^{\frac{1}{2-\alpha}},$$

где C зависит только от параметров, участвующих в определении условий Цыбакова.

Задача 3. Докажите, что с точностью до абсолютных констант порядок $\frac{k\log(\frac{n}{k})}{n}$ для схем сжатия выборок неулучшаем, то есть существует такой класс $\mathcal{F}$ со схемой сжатия размера k и распределение P_X и $f^* \in \mathcal{F}$ такие, что с некоторой фиксированной вероятностью $P(\hat{f}(X) \neq f^*(X)) = \Omega(\frac{k\log(\frac{n}{k})}{n})$.
Указание. Вам могут очень пригодиться идеи Теоремы 6 из http://jmlr.org/proceedings/papers/v40/Simon15a.pdf

■

Задача 4*. Постройте схему сжатия в $O(d\log(n))$ точек для класса VC размерности d и выборки длины n.
Указание. Здесь стоит очевидным образом переопределить схемы сжатия чтобы допускать зависимость от n. Это утверждение все еще слабее гипотезы о сжатии в d точек. Убедитесь, что задача решается очень легко, если бы неразмеченные точки сжимаемой выборки оставались доступны нам после сжатия.

■