

Домашнее задание №3 по курсу "Машинное обучение"

Прокопенко Тимофей

Задача 1.

Решение:

Необходимо найти $VCdim(H)$, где H - семейство линейных бинарных классификаторов в d -мерном пространстве. Рассмотрим линейные бинарные классификаторы вида $h(x) = \text{sign}(w^T x)$, $w \in R^d$.

а) Предположим, что $VCdim(H)=d$. Для этого нужно найти S размера d , которое раскрашивается с помощью H .

Возьмем в качестве S базис d -мерного пространства, то есть вектора x_i вида $(0, 0, \dots, 1_i, \dots, 0)$, где $i = \overline{1, d}$.

$y = (y_1, y_2, \dots, y_n)$, $y_i \in \{-1, 1\}$, $i = \overline{1, d}$ - значение произвольной функции из H_C .

В качестве гипотез H_C выберем $h(x_j) = \text{sign}(\sum_{i=1}^d y_i x_i^T x_j)$, где $y_i \in \{-1, 1\}$, $w = \sum_{i=1}^d y_i x_i \in R^d$,

$x_j, x_i \in C$.

Так как $x_i^T * x_j = 0$ при $i, j = \overline{1, d}, i \neq j$, а $x_i^T * x_i = 1$, то $h(x_i) = \text{sign}(y_i)$. Таким образом, H раскрашивает d -мерное пространство.

б) Допустим, что с помощью H можно раскрасить $(d+1)$ -мерное пространство. Но в таком пространстве найдется вектор, который будет являться линейной комбинацией остальных d векторов и, следовательно, его классификация будет предопределена. Таким образом, пространство размерности $d+1$ не может быть раскрашено H , и $VCdim(H)=d$.

Задача 3.

Решение:

Повторим начальные рассуждения из первой задачи. Предположим, что $VCdim(H)=n$. Для этого нужно найти S размера n , которое раскрашивается с помощью H .

Возьмем в качестве S базис n -мерного пространства, то есть вектора x_i вида $(0, 0, \dots, 1_i, \dots, 0)$, где $i = \overline{1, n}$.

Пусть $y = (y_1, y_2, \dots, y_n)$, $y_i \in \{0, 1\}$, $i = \overline{1, d}$ - значение произвольной функции из H_C . Выберем такое множество $I_C = \{i | y_i = 1\}$. Тогда, учитывая характер ранее выбранного множества S , получим

$$h_{I_C}(x_i) = \begin{cases} 1, & \text{если } i \in I_C \\ 0, & \text{иначе} \end{cases}$$

Таким образом, H раскрашивает множество размера n .

Множество размера $n+1$ не может быть раскрашено, т.к. $VCdim(H) \leq \ln(|H|)$, $|H|=2^n \Rightarrow VCdim(H) \leq n$. Значит $VCdim(H) = n$.

Задача 4.

Решение:

1. ERM-алгоритм над конечным классом H - PAC-learnable в случае гипотезы реализуемости рассматривает только те D , для которых гипотеза реализуемости выполняется. В то время как NFL theorem рассматривает все D .
2. ERM-алгоритм над конечным классом H - agnostic PAC-learnable говорит, что для выбранной гипотезы h $L_D(h)$ отклоняется от $L_D(h')$ (h' - лучшая возможная гипотеза) не больше, чем на ϵ с вероятностью $1-\delta$ ($\epsilon, \delta \in (0, 1)$). NFL theorem этому не противоречит, там говорится, что для нашей гипотезы h найдется D , что $L_D(h) \geq 1/8$.