Домашнее задание 3

Авласов Владислав

Задание 1

Вычислите VCdim(H), если H - семейство линейных бинарных классификаторов в d-мерном пространстве.

Покажем, что VCdim(H) = d + 1. Будем действовать по схеме из 2 пунктов, описанной в лекции.

1. Сначала покажем, что VCdim(H) $\geq$ d+1. Семейство наших классификаторов выглядит в общем виде так: $H = \operatorname{sign}\left(b + w^T x\right)$
Покажем, что существует набор из d+1 точки, который можно "раскрасить". Пусть X будет квадратной матрицей размерности d+1 на d+1.

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_0^\top \\ \vdots \\ \boldsymbol{x}_d^\top \end{bmatrix} \tag{1}$$

Каждый вектор x_i имеет размерность d+1, где $x_i[0] = 1$ отвечает за смещение(априорное знание), а из остальных d координат d-мерной линейной функции $x_i[i] = 1$, а остальные равны 0. Т.е. у точки x_0 ненулевая только координата смещенияя b. Такая матрица не является вырожденной, т.е. к ней можно построить обратную. В итоге мы можем переписав наше семейство классификаторов с использованием матрицы, имеем:

$$\operatorname{sign}(Xw) = y \tag{2}$$

Т.к. у нас бинарный классификатор, т.е. $y \in$\{-1,+1\}, то (2) $\Leftrightarrow$

$$Xw = y \tag{3}$$

Домножим на обратную матрицу, получим:

$$w = X^{-1} y \tag{4}$$

Т.е. смогли раскрасить.

2. Покажем, что VCdim(H) $<$ d+2. Пусть у нас есть набор из d+2 точек в (d+1)-мерном пространстве. Очевидно, что он линейно зависимый, т.е. существует $x_j = \sum_{i \neq j} a_i x_i$ где, не все a равны 0. Пусть y_j = -1 и $y_i = \operatorname{sign}(a_i) = \operatorname{sign}\left(\boldsymbol{w}^\top \boldsymbol{x}_i\right)$. Если $a_i = 0, y_i$ может быть любым. Подставим и получим:

$$\operatorname{sign}\left(\sum_{i \neq j} a_i w^\top x_i\right) > 0 \neq y_j \tag{5}$$

Т.е. не можем разукрасить d+2 точек. ЧТД.

Задание 3

Условие про гиперкуб, которое я не знаю как быстро и красиво сюда копировать

1. Возьмём $X = E_n$, т.е. единичная м-ца размерности n. Тогда для любого вектора y $\exists I \in \{1, ..n\}$ такое, что $h_I(x_i) = y_i$ для любого $x_i \in X$. I в данном случае строим так: для всех $y_i = 1$ добавляем индекс i в наше I. Все x_i в нашей маске имеют только одну 1, поэтому мы её и выбираем и $y_i = 1$. Т.е. можем разукрасить n-элементное множество точек.

2. Очевидно, что мощность всего множества $|\mathcal{H}| = 2^n$. У нас было красивое свойство в презентации:

$$\text{VCdim}(H) \leqslant \log_2 |H| \tag{6}$$

Согласно ему и нашему очевидному факту получаем, что VCdim(H) $\leq$ n.
Т.е. доказали, что VCdim(H) = n.

Задание 4

Объясните, как согласуются:
1.ERM-алгоритм над конечным классом H - PAC-learnable в случае гипотезы реализуемости и No Free Lunch theorem?
2.ERM-алгоритм над конечным классом H - agnostic PAC-learnable и No Free Lunch theorem?

FLT-теорема говорит нам о том, что какой-то алгоритм не может решать все задачи одинаково хорошо, и нам важно ограничивать множество наших гипотез, опираясь на какие-то априорные данные.
В первом случае, когда у нас есть гипотеза о реализуемости, то $\exists h^* \in H$, что $L_{D,f}(h^*) = 0$. Т.е. у нас есть априорное знание, и ERM-алгоритм при достаточной размере m выборки S сможет выбрать такую гипотезу, при которой тру-риск будет мало отличаться от эмпирического.
Во втором случае ERM-алгоритм над agnostic PAC-learnable классом ничего не говорит о размере ошибки классификации, он только утверждает, что эмпирический риск не будет сильно отличаться от общей ошибки классификации, а эта ошибка может быть какой угодно большой. Т.е. это никак не противоречит FLT.