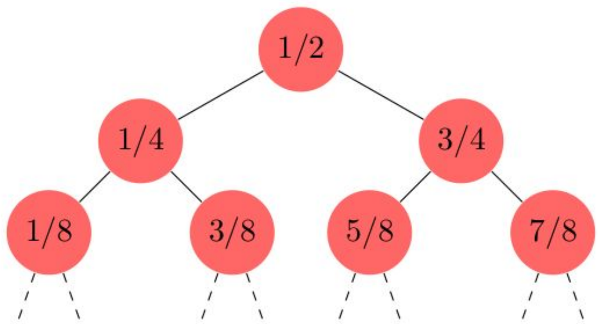
2. доказать, что разница между $\mathrm{Ldim}(H)$ и $\mathrm{VCdim}(H)$ может быть сколь угодно большой

На $\mathcal{X} = [0, 1]$ рассмотрим класс гипотез $\mathcal{H} = \{x \mapsto \mathbb{1}_{[x<a]} : a \in [0, 1]\}$. VCdim(H) = 1, Ldim(H)=$\infty$ (это было показано на лекции и вот тут

Example 21.4 Let $\mathcal{X} = [0, 1]$ and $\mathcal{H} = \{x \mapsto \mathbb{1}_{[x<a]} : a \in [0, 1]\}$; namely, $\mathcal{H}$ is the class of thresholds on the interval $[0, 1]$. Then, $\mathrm{Ldim}(\mathcal{H}) = \infty$. To see this, consider the tree



This tree is shattered by $\mathcal{H}$. And, because of the density of the reals, this tree can be made arbitrarily deep.

)

Следовательно разница между Ldim(H) и VCdim(H) может быть сколь угодно большой.

1. доказать, что $M_{\mathtt{SOA}}(H) = \mathrm{Ldim}(H)$

Доказывается как следствие из двух лемм.

LEMMA 21.6 *No algorithm can have a mistake bound strictly smaller than* $\mathrm{Ldim}(\mathcal{H})$*; namely, for every algorithm, A, we have* $M_A(\mathcal{H}) \geq \mathrm{Ldim}(\mathcal{H})$.

Proof Let $T = \mathrm{Ldim}(\mathcal{H})$ and let $\mathbf{v}_1, \ldots, \mathbf{v}_{2^T-1}$ be a sequence that satisfies the requirements in the definition of Ldim. If the environment sets $\mathbf{x}_t = \mathbf{v}_{i_t}$ and $y_t = 1 - p_t$ for all $t \in [T]$, then the learner makes T mistakes while the definition of Ldim implies that there exists a hypothesis $h \in \mathcal{H}$ such that $y_t = h(\mathbf{x}_t)$ for all t. □

LEMMA 21.7 *SOA enjoys the mistake bound* $M_{SOA}(\mathcal{H}) \leq \mathrm{Ldim}(\mathcal{H})$.

Proof It suffices to prove that whenever the algorithm makes a prediction mistake we have $\mathrm{Ldim}(V_{t+1}) \leq \mathrm{Ldim}(V_t) - 1$. We prove this claim by assuming the contrary, that is, $\mathrm{Ldim}(V_{t+1}) = \mathrm{Ldim}(V_t)$. If this holds true, then the definition of p_t implies that $\mathrm{Ldim}(V_t^{(r)}) = \mathrm{Ldim}(V_t)$ for both $r = 1$ and $r = 0$. But, then we can construct a shaterred tree of depth $\mathrm{Ldim}(V_t) + 1$ for the class V_t, which leads to the desired contradiction. □

3. найти класс H, что алгоритм `Consistent` делает на нём $|H| - 1$ ошибку

Как я понимаю тут надо найти не только класс но и последовательность x_i и правильных ответов.

Построим класс гипотез таким образом:
Все гипотезы кроме одной возвращают 1 при x=1
Все гипотезы кроме одной возвращают 1 при x=2
и т.д.
Только одна гипотеза всегда возвращает 1.

Правильным ответом всегда будет единица.
Построим последовательности x таким образом. До тех пор пока наш алгоритм не вернёт ноль в последовательности будут идти одни единицы. После того как алгоритм вернул ноль (следовательно алгоритм ошибся и одна из гипотез была удалена из списка Vt) в последовательности будут идти двойки опять же до тех пор пока алгоритм не вернёт ноль, и т.д.
Тогда после каждой ошибки будет удалена ровно одна гипотеза и в конечном счёте останется только та гипотеза, которая всегда возвращает единицу.
Следовательно, алгоритм Consistent ошибется |H|-1 раз.

4. найти класс H, что алгоритм `Halving` делает на нём ровно $\log_2(|H|)$ ошибок

Построим класс гипотез таким образом:
Половина гипотез (если кол-во гипотез чётное, то большая из половин) возвращает 0 при x=1
Гипотезы из оставшейся половины возвращают 1 при x=1
и т.д.
Только одна гипотеза всегда возвращает 1.

Правильным ответом всегда будет единица.
Последовательность x будет последовательность натуральных чисел.

Если $|\{h \in V_t : h(\mathbf{x}_t) = r\}|$ будет одинакова при r=0 и r=1, то будем выбирать ответ 0 (и алгоритм ошибется).
На каждом шаге алгоритм будет ошибаться и будет отброшена большая из половин от уже оставшихся гипотез, следовательно алгоритм ошибется ровно log2(|H|) раз.

5. **[бонус]** почему в online perceptron не взять в качестве суррогатной лосс-функции $f_t(w) = 0$, если алгоритм не ошибается и $f_t(w) = 1$, если ошибается? Чем плох такой выбор?

Как я понимаю так мы не уйдём от проблемы, что если

$d \geq 2 \text{ then } \mathrm{Ldim}(\mathcal{H}) = \infty$