Домашнее задание №1 по курсу «Машинное обучение»:

Кукуев Михаил

1. Так как $P[X>t]=1-P[X\leqslant t]$, то для неравенства из условия можно выполнить следующие преобразования:

$$\forall t>0:\ P[X>t]\leqslant f(t)\Leftrightarrow$$
$$\forall t>0:\ 1-P[X\leqslant t]\leqslant f(t)\Leftrightarrow$$
$$\forall t>0:\ P[X\leqslant t]\geqslant 1-f(t)$$

Так как f обратима, то $\exists\delta:\ f^{-1}(\delta)=t,\ f(t)=\delta$. Учитывая, что $0\leqslant P[X\leqslant t]\leqslant 1$, получаем:

$$\forall\delta>0:\ P[X\leqslant f^{-1}(\delta)]\geqslant 1-\delta$$

2.

1) Классификатор h_s можно записать следующим образом:

$$h_s(x)=\begin{cases}1, & \textit{если}\ \ \exists i\in[m]:\ x_i=x, y_i=1\\ 0 & \textit{иначе}\end{cases}$$

Для удобства выделим из объектов тренировочной выборки отдельный набор $S^*=\{x_i\,|\,y_i=1, i\in[m]\}$. Если существует полином, возвращающий значения ≥ 0 только для $x\in S^*$, а в остальных случаях < 0, то классификатор из h_p с таким полиномом совпадет с h_s. В качестве такого полинома можно взять

$$P(x)=-(x-x_1^*)^2(x-x_2^*)^2\ldots(x-x_n^*)^2,\quad n=|S^*|,\quad x_j^*\in S^*,\quad j=\overline{1,n}$$

Его корнями являются значения из S^*, поэтому для них он возвращает 0, а в остальных случаях значения < 0.

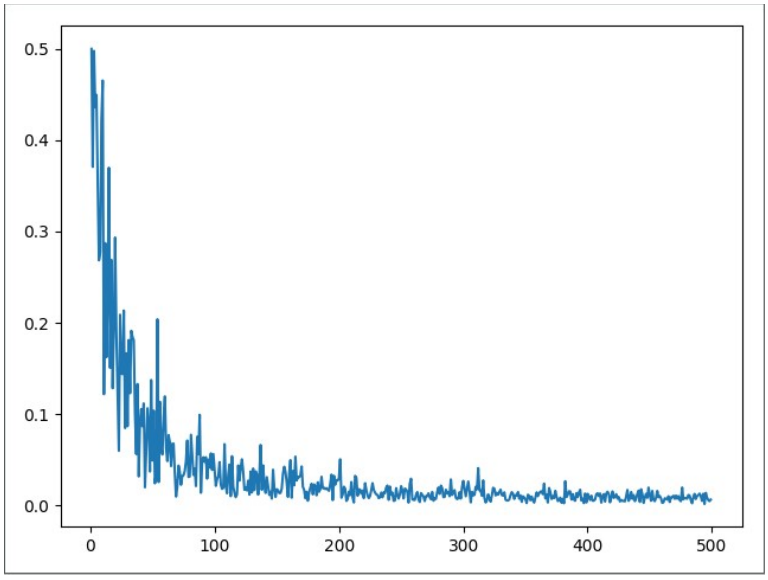
2) Касательно ERM-парадигмы в классе h_p, можно сделать вывод, что раз в нем найдется такой “плохой” ERM-классификатор как h_s, то применение ERM-парадигмы при выборе классификаторов из h_p может приводить к переобучению.

3.

1) С одной стороны, раз выполнено предположение о реализуемости, то существует классификатор h_D с прямоугольником, содержащим все точки положительного класса, и только их (без точек отрицательного). С другой стороны, если алгоритм A находит наименьший прямоугольник, содержащих все точки положительного класса из набора S, то он будет лежать внутри прямоугольника, соответствующего классификатору h_D, и следовательно точно не будет содержать точек отрицательного класса. А это значит, что эмпирический риск классификатора, реализуемого алгоритмом A, равен 0, и это ERM-классификатор.

Решение пунктов 2) – 4) в файле hw1.py, используемая версия python – 3.5. Для получения результатов можно просто запустить код из файла или выполнить сам файл как скрипт.

3) График true risk:



4) Средние размеры выборки, необходимые для достижения значений true risk:
m = 34 для true_risk = 0.1
m = 334 для true_risk = 0.01
m = 2989 для true_risk = 0.001

5)
Формула расчета true risk: $true\,risk = (S_Q - S_A)/S_X$, S_Q ,S_X, S_A – площади прямоугольников Q, X, и возвращаемого алгоритмом A соответственно.

Как ответ на предыдущий пункт должен зависеть от площади X (при неизменной площади Q)?
Из формулы расчета, true risk обратно пропорционален площади X, значит с ростом X будет уменьшаться и m.

От относительной площади Q и X?
Не зависит, снова из формулы, так как $S_A \propto S_Q$, true risk и m не изменятся при росте относительной площади Q и X.

От размерности пространства X?
Пусть q_n, x_n и a_n – длины n-й стороны n-мерных прямоугольников Q, X, и возвращаемого алгоритмом A соответственно, а $S_Q^{(n)}, S_X^{(n)}, S_A^{(n)}$ их площади. Очевидно, что $S_Q^{(n)} = S_Q^{(n-1)} q_n$, для *X* и *A* аналогично. Сравним true risk для *n*-го и *(n-1)*-го измерений:

$$L_D(h,n) - L_D(h,n-1) = \frac{S_Q^{(n)} - S_A^{(n)}}{S_X^{(n)}} - \frac{S_Q^{(n-1)} - S_A^{(n-1)}}{S_X^{(n-1)}} = \frac{S_Q^{(n-1)}}{S_X^{(n-1)}}\left(\frac{q_n}{x_n} - 1\right) + \frac{S_A^{(n-1)}}{S_X^{(n-1)}}\left(\frac{a_n}{x_n} - 1\right) \leqslant [q_n \leqslant x_n, a_n \leqslant q_n] \leqslant 0$$

Значит, при увеличении размерности true risk будет уменьшаться, а следовательно m тоже.

Должен ли зависеть результат от D?
Да, так как выбор распределения вероятности влияет на вероятность выпадения “плохой” (нерепрезентативной) выборки, а от этого зависит и true risk, и m.

Если да, то как объяснить тот факт, что зависимость есть, хотя в определении PAC-learnable класса выборочная сложность не зависит от D?
Выборочная сложность зависит от вероятности получения нерепрезентативной выборки δ, на которую влияет выбранное распределение вероятности D.