Задание1.
Вычислите VCdim(H), если H — семейство линейных бинарных классификаторов в d-мерном пространстве
Решение.
Покажем, что ответ задачи d+1.
Рассмотри d+1 точку из d-мерного пространства. И запишем их в виде матрицы

$$X = \begin{matrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ & & \dots & & \\ 1 & 0 & 0 & & 1 \end{matrix}$$

Тут будет d+1 строка и столбец.
Определитель этой матрицы равен 1 => она обратимая.

Классификатор будет определять два класса из мн-ва {1,-1}. Надо показать, что существует вектор w, такой что для любого вектора y вида:

$$y= \begin{matrix} y1 \\ y2 \\ .. \\ yd+1 \end{matrix}$$

где yi = 1 или -1. Выполняется sign(Xw)=y
Это можно сделать так:
Xw = y, и так как X обратима, то w = $X^{-1}y$. Следовательно мы можем раскрасить d+1 точку.
Покажем, что не можем раскрасить никакие d+2 точки.
Размерность пространства X после избавления от неоднородности для бинарного классификатора будет d+1. Возьмём d+2 точки $(x_1, .., x_{d+2})$. Так как кол-во точек больше размерности пр-ва, то они должны линейно выражаться друг через друга. Т.е.:

$$x_j = \sum_{i \neq j} a_i * x_i$$, где не все ai одновременно равны нулю.

Тогда

$$w^T * x_j = \sum_{i \neq j} a_i * \ w^T * x_i$$

Так как надо раскрасить всё множество,то должен существовать такой набор yi:

yi = sign($w^T * x_i$) = sign(ai), тогда $a_i * \ w^T * x_i$ > 0. Тогда $y_j = sign(w^T * x_j) \ = \ +1$
Т.е. получается, что из-за линейной зависимости резултат на xj будет зависеть от результата от всех остальных xi и множество всех точек не сможет быть раскрашено полность.
Следовательно, VCdim(H) = d + 1.

Задание 3.

Пусть X — булев гиперкуб размерности n. Для множества $I \in \{1, 2, \dots, n\}$ и объекта $x \in X$, $x = (x_1, x_2, \dots, x_n)$ зададим функцию $h_I(x) = (\sum_{i \in I} x_i) \mod 2$. Чему равна VCdim таких множества всех таких функций?

Пусть Hn = $\{h_I : I \subseteq \{1, 2, ..., n\}\}$. VCdim такого множества и надо найти.
Рассмотрим случай n = 1
В этом случае Hn состоит из двух ф-ий:
h0 = 0
h1 = x1 mod 2
Возьмем множество X =(x1), где x1=1, тогда
h0(x)=0
h1(x)=1
Т.е. это множество мы смогли раскрасить.VCdim=n=1
Теперь рассмотрим случай n=2. В этом случае Hn состоит из 4-ёх ф-ий:
h0=0
h1=x1 mod 2
h2=x2 mod 2
h3=(x1+x2) mod 2
Возьмем X=(x1,x2), где x1=10, x2 =01
(ho(x1),ho(x2)) = (0,0)
(h1(x1),h1(x2)) = (1,0)
(h2(x1),h2(x2)) = (0,1)
(h3(x1),h3(x2)) = (1,1)
Т.е. это множество мы смогли раскрасить.VCdim=n=2
Ну, и рассмотрим случай n=3. Hn будет:
h0=0
h1=x1 mod 2
h2=x2 mod 2
h3=x3 mod 2
h4=(x1+x2) mod 2
h5=(x1+x3) mod 2
h6=(x2+x3) mod 2
h6=(x1+x2+x3) mod 2
Возьмем X=(x1,x2,x3). x1=(1,0,0),x2=(0,1,0),x3=(0,0,1). Тогда
(ho(x1),ho(x2),h0(x3)) = (0,0,0)
(h1(x1),h1(x2),h1(x3)) = (1,0,0)
(h2(x1),h2(x2),h2(x3)) = (0,1,0)

(h3(x1),h3(x2),h3(x3)) = (0,0,1)
(h4(x1),h4(x2),h4(x3)) = (1,1,0)
(h5(x1),h5(x2),h5(x3)) = (1,0,1)
(h6(x1),h6(x2),h6(x3)) = (0,1,1)
(h7(x1),h7(x2),h7(x3)) = (1,1,1)
Т.е. это множество мы смогли раскрасить.VCdim=n=3

Мощность множества Hn равна количеству подмножеств множества из n элементов. Т.е. в множестве Hn будет 2^n ф-ий. Количество всевозможных способов раскрасить множество из n элементов равно 2*2*..*2 n раз, т.е 2^n. Тогда нашими функциями мы сможем раскрасить множество из n элементов (иначе у нас просто не хватит ф-ий)
Покажем, что эта оценка является достижимой.

Рассмотрим единичную матрицу размерности n. Каждая строка этой матрицы будет репрезентировать элемент из множества X. И покажем, что такое множество может быть раскрашено. Т.е. C = {(100..0),(010..0),(000..1)}. |C|=n
Для случаев n=1,2,3 пример приведен выше.

Множество Hn можно описать как множество множество ф-ий, каждая из которых суммирует некоторые биты в аргументе ф-ии. Причём Hn будет содержать всевозможные такие суммирования
$I \subseteq \{1, 2, ..., n\}$. Возьмём I ={1,2,3}, тогда $h_I = (x1 + x2 + x3) \, mod \, 2$. И у нас будут все подмножества множества $\{1, 2, ..., n\}$.
В каждом x из множества, которое было введено выше, содержит ровно 1 единицу, тогда если посчитать h_I на каждом элементе из множества C (h_I (c1), h_I (c2),.., h_I (cn)), то полученный вектор будет содержать единицы на i-ой позиции ,если xi входит в сумму в h_I . А так как Hn содержит все возможные варианты выборок xi из n элементов, то получим все возможные вектора из n элементов. Следовательно VCdim = n

Задание 4

4. Объясните, как согласуются:

- ERM-алгоритм над конечным классом H — PAC-learnable в случае гипотезы реализуемости и No Free Lunch theorem?
- ERM-алгоритм над конечным классом H — agnostic PAC-learnable и No Free Lunch theorem?

No free lunch теорема говорит, что в случае небольшого размера выборки (меньшего, чем половина всего домена), существует такое распределение и такая ф-ия (которая не ошибается на этом распределении), что любой алгоритм с довольно большой вероятностью будет ошибаться на данной распределении. А то что алгоритм является PAC-learnable (в случае гипотезы реализуемости), говорит, что для любого распределения и для любой ф-ии при некоторой уверенности в обучающей выборке в случае размера выборки большей чем значение некоторой функции (зависящей от необходимой точности алгоритма и степени уверенности в алгоритме) алгоритм с большой вероятностью ошибается очень мало (но для этого видимо понадобится большой размер выборки, больший чем половина домена (если он меньше и мы можем достигнуть любой точности на ней, то это противоречит теореме)).

Свойство agnostic PAC-leanable не противоречит No Free Lunch теореме, так как в случае если класс является agnostic PAC-leanable, то это значит что в классе гипотез H существует гипотеза, которая может с некоторой точностью ε решать задачу лучше любой другой гипотезы из этого класса при условии некоторой уверенности в обучающей выборке σ, но это не значит, что эта гипотеза не может с вероятностью не меньшей 1/7 ошибаться более чем на $\frac{1}{8}$, при условии No free lunch theorem.