1. Возьмем классификатор вида $H = sign(w^T x + b)$

   $VCdim(H) \geq d + 1$:

   $$X = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ 1 & \vdots & \vdots & \ddots & 0 \\ 1 & 0 & 0 & \dots & 1 \end{pmatrix}$$

   Так как X - обратимая, мы всегда можем выбрать $w = X^T y$, который приведет $w^T x$ в необходимый знак

   $VCdim(H) \leq d + 2$:

   Очевидно, что взяв систему из d+1 уравнений и d+2 переменной мы получим линейную зависимость, а значит не сможем покрыть множество решений своим множеством гипотез

   P.S. Ответ будет d по аналогии, если из классификатора убрать b и соответственно bias в матрице

3. a) Возьмем в качестве C базис пространства, где
   $$x_i^C = (0, \dots, 0, 1^{(i)}, 0, \dots, 0), \forall i = \overline{1, n}$$
   Пусть $y_C = (y_1^C, y_2^C, \dots, y_n^C), y_j^C = \{0, 1\}, \forall j = \overline{1, n}$ - значение произвольной функции из H_C. Покажем, что для получения необходимых значений на наборе C найдется функция из H.

   Выберем множество
   $$I_C = \{i | y_i^C = 1\}$$
   Значит, $h_{I_C}(x_j^C) = 1$, когда $j \in I_C$. Следовательно, семейство H_C таких функций разукрашивает C и
   $$VCdim(H) \geq n$$

   б) $|X| = n$, значит $|H| \leq 2^n$
   $$VCdim(H) \leq log_2(|H|)$$
   $$VCdim(H) \leq n$$

1. ERM-алгоритма над конечным классом H - PAC-learnable только с учетом гипотезы о реализуемости, а No-FLT работает без этого предположения.
2. agnostic PAC-learnability утверждает только что true risk найденной гипотезы больше true risk наилучшей гипотезы не более чем на ε. No-FLT утверждает, что даже для наилучшей гипотезы найдется такое распределение вероятностей, что $L_D(h') \geq 1/8$, что не вызывает противоречий.