

1. При бэггинге новую выборку X_i составляют, генерируя элементы из X с возвращением. При этом объекты в X_i могут повторяться. Будем считать, что число объектов в X и в X_i одинаковое и равно N . Найдите вероятность того, что конкретный объект попадет в выборку.
2. Рассмотрим задачу бинарной классификации, пусть у нас есть три алгоритма, каждый из которых ошибается с вероятностью p . С какой вероятностью будет ошибаться классификатор, построенный по принципу простого голосования?
3. Истинная зависимость имеет вид $y_i = x_i^2 + u_i$, где y_i прогнозируемая переменная, x_i - признак и u_i - ненаблюдаемая случайная составляющая. Величины x_i независимы и равновероятно принимают значения 1 и 2. Величины u_i независимы и равновероятно принимают значения -1 и 1. Разложите ожидание квадрата ошибки прогноза на шум, смещение и разброс, если:
 - а) Вне зависимости от обучающей выборки из-за ошибки в коде в качестве прогноза всегда выдаётся 0.
 - б) Вне зависимости от обучающей выборки из-за ошибки в коде в качестве прогноза равновероятно выдаётся -1 или 1.
 - в) По обучающей выборке строится регрессию на константу.
 - г) В качестве прогноза алгоритм всегда выдаёт последний y из обучающей выборки.
4. Рассмотрим задачу бинарной классификации, пусть у нас есть три алгоритма $b_1(x)$, $b_2(x)$ и $b_3(x)$, каждый из которых ошибается с вероятностью p . Мы строим композицию взвешенным голосованием: алгоритмам присвоены значимости w_1 , w_2 и w_3 , и для вынесения вердикта суммируются значимости алгоритмов, проголосовавших за каждый из классов:

$$a_0 = \sum_{i=1}^3 w_i [b_i(x) = 0],$$

$$a_1 = \sum_{i=1}^3 w_i [b_i(x) = 1].$$

Объект x относится к классу, для которого сумма оказалась максимальной. Например, если первые два алгоритма голосуют за класс 0, а третий за класс 1, то выбирается класс 0, если $w_1 + w_2 > w_3$, и класс 1 в противном случае. Какова вероятность ошибки такой композиции этих трех алгоритмов, если:

a) $w_1 = 0.3, w_2 = 0.4, w_3 = 0.3;$

b) $w_1 = 0.2, w_2 = 0.5, w_3 = 0.2?$