Машинное обучение.
Домашнее задание №2

Задача 1. Убедитесь, что можете ответить на следующие вопросы:

1. Какая функция потерь оптимизируется в AdaBoost?
2. Какому требованию должно удовлетворять семейство базовых алгоритмов?
3. Как вычисляются веса объектов в AdaBoost? Как веса с текущей итерации зависят от весов с предыдущей итерации?
4. Как в AdaBoost строится очередной базовый алгоритм?
5. Почему говорят, что AdaBoost неустойчив к выбросам?
6. С какой скоростью ошибка AdaBoost на обучающей выборке стремится к нулю?
7. Что такое решающий пень? Как использовать его в AdaBoost в качестве базового алгоритма?
8. Какая функцию потерь используется в многоклассовом AdaBoost? Как в нем устроены базовые алгоритмы?

Задача 2. На семинаре был рассмотрен пример с бутстрэппингом, где строилась композиция из нескольких функций регрессии. Откажемся от предположения о несмещенности и некоррелированности ошибок.

Пользуясь неравенством Йенсена, покажите, что среднеквадратичная ошибка композиции не превосходит среднюю ошибку отдельных алгоритмов:

$$E_n \leqslant E_1.$$

Задача 3. Пусть $X^\ell = \{x_1, \ldots, x_\ell\} \subset \mathbb{R}$ — произвольная одномерная обучающая выборка. Покажите, что при любых ответах $(y_i)_{i=1}^\ell$ на этих объектах существует композиция вида

$$a(x) = \operatorname{sign} \sum_{n=1}^{N} \gamma_n b_n(x)$$

над решающими пнями, не допускающая ошибок на обучающей выборке X^ℓ. Покажите, что в ней будет не более $2\ell + 2$ различных классификаторов (классификаторы считаются одинаковыми, если они дают одинаковые ответы на обучающей выборке).

Задача 4. Пусть $X^\ell = \{x_1, \ldots, x_\ell\} \subset \mathbb{R}^d$ — произвольная d-мерная обучающая выборка. Рассмотрим *радиальные* базовые функции:

$$b(x; i) = y_i \exp\left(-\beta \|x - x_i\|^2\right).$$

Параметром такой функции является индекс i объекта обучающей выборки; величина $\beta > 0$ является гиперпараметром и считается фиксированной в нашей задаче. Покажите, что при любых ответах $(y_i)_{i=1}^\ell$ существует взвешенная композиция радиальных функций

$$a(x) = \operatorname{sign} \sum_{n=1}^{N} \gamma_n b_n(x),$$

не допускающая ошибок на обучающей выборке X^ℓ.

Задача 5. Пусть известно распределение на объектах и ответах $p(x, y)$. Ответы на объектах принадлежат множеству $\mathbb{Y} = \{-1, +1\}$. Рассмотрим алгоритм, который возвращает ответ, минимизирующий матожидание экспоненциальной функции потерь в данной точке:

$$a^*(x) = \arg\min_{a \in \mathbb{R}} \mathbb{E}_{y \,|\, x} e^{-ya} = \arg\min_{a \in \mathbb{R}} \sum_{y \in \{-1, +1\}} p(y \,|\, x) e^{-ya}.$$

Найдите в явном виде алгоритм $a^*(x)$.