Машинное обучение.
Домашнее задание №3

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

1. В чем идея бустинга с квадратичной функцией потерь? Как настраивается каждый базовый алгоритм?
2. Как устроен градиентный бустинг, если нам известно распределение $p(x, y)$? Как вычисляется градиент в конкретной точке x в этом случае?
3. Как аппроксимируется градиент на всем пространстве объектов в случае, когда известна лишь конечная выборка?
4. В чем заключается сокращение шага в градиентном бустинге? Как число итераций, необходимое для сходимости, зависит от размера шага η?
5. Что такое стохастический градиентный бустинг?
6. Как записывается логистическая функция потерь? Что возвращает алгоритм, настроенный на нее? Как по ответу алгоритма найти вероятность $p(y = 1 \mid x)$?
7. В чем заключается идея переподбора ответов в листьях деревьев на каждом шаге градиентного бустинга?
8. Как градиентный бустинг взвешивает объекты в случае, когда функция потерь является функцией от отступа?
9. Почему логистическая функция потерь более устройчива к выбросам по сравнению с экспоненциальной?

Задача 2. Выпишите формулы для поиска базового алгоритма b_N и коэффициента γ_N в градиентном бустинге с экспоненциальной функцией потерь.

Задача 3. Рассмотрим задачу бинарной классификации, $\mathbb{Y} = \{0, 1\}$. Будем считать, что все алгоритмы из базового семейства $\mathcal{A}$ возвращают ответы из отрезка $[0, 1]$, которые можно интерпретировать как вероятности принадлежности объекта к классу 1. В качестве функции потерь возьмем отрицательный логарифм правдоподобия (negative log-likelihood):

$$L(y, z) = -\big(y \log z + (1 - y) \log(1 - z)\big),$$

где y — правильный ответ, а z — ответ алгоритма.

Выпишите формулы для поиска базовых алгоритмов b_n и коэффициентов γ_n в градиентном бустинге.

Задача 4. При переподборе ответов в листьях деревьев с логистической функцией потерь нужно решить задачу

$$F_j^{(N)}(\gamma) = \sum_{x_i \in R_j} \log\left(1 + \exp\left(-2y_i(a_{N-1}(x_i) + \gamma)\right)\right) \to \min_{\gamma}.$$

Обычно ее решают приближенно, делая один шаг метода Ньютона-Рафсона из начального приближения $\gamma = 0$. В этом случае

$$\gamma_{Nj} = \frac{\partial F_j^{(N)}(\gamma)}{\partial \gamma} \Bigg/ \frac{\partial^2 F_j^{(N)}(\gamma)}{\partial \gamma^2}.$$

Покажите, что данное отношение производных равно

$$-\sum_{x_i \in R_j} g_i^{(N)} \Bigg/ \sum_{x_i \in R_j} |g_i^{(N)}|(2 - |g_i^{(N)}|).$$