XGBoost (eXtreme Gradient Boosting) — одна из самых популярных и эффективных реализаций алгоритма градиентного бустинга на деревьях на 2019-й год.

Достоинства:

1. Производительность. (в основном за счет распараллеливания)
2. Универсал: может быть использован для решения задач регрессии, классификации, упорядочения и пользовательских задач на предсказание.

   НО!
   В задачах предсказания, которые используют неструктурированные данные (например, изображения или текст), искусственная нейронная сеть превосходит все остальные алгоритмы или фреймворки. Но когда дело доходит до структурированных или табличных данных небольших размеров, в первенстве оказываются алгоритмы, основанные на дереве поиска решений.

   Для справки.

- Конкурентами являются Light GBM (Microsoft) и CatBoost от Яндекса в зависимости от типа задач.
- Часто используется в соревнованиях.

Описание алгоритма

XGBoost est 2016.

В основе лежит алгоритм градиентного бустинга деревьев решений. Градиентный бустинг – это техника машинного обучения для задач классификации и регрессии, которая строит модель предсказания в форме ансамбля слабых предсказыв. моделей, обычно деревьев решений. На каждой итерации вычисл. отклонения предсказаний уже обученного ансамбля на обучающей выборке. След. модель, которая будет доб. в ансамбль будет предсказывать эти отклонения. Таким образом, добавив предск-ия нового дерева к предск. обучен. ансамбля мы можем уменьшить сред. отклон. модели, которое явл. целью оптим. задачи. Новые деревья доб. в ансамбль до тех пор, пока ошибка уменьшается, либо пока не вып-ся критерий остановки.

Иллюстрация бустинга

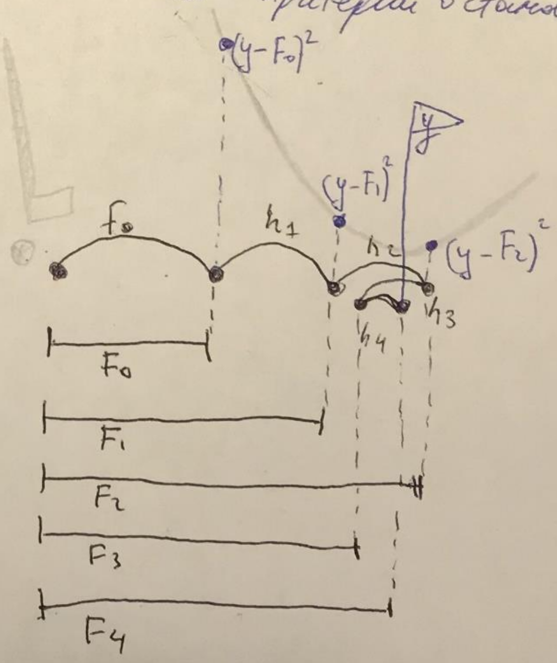
Рассм. поведение модели на одной точке абстрактной задачи лин. регрессии.

] 1-ая модель ансамбля F всегда выдает выборочное среднее предск. величины f_0. Предсказание грубое, поэтому MSE в точке будет большим.

Обучим модель h_1, которая будет „корр." предсказание пред. ансамбля F_0.
Получим ансамбль F_1, предсказание которого будет сумм. из предсказаний моделей f_0 и h_1.

⋮

Продолжая, получаем F_4, предсказание которого сумм. из предск. $f_0 + h_1 \dots h_4$ и предсказывает в точности значение заданного таргета.



Построим ансамбль деревьев

$$f(x) = f_M(x) = f_0 + \sum_{m=1}^{M} h_m(x) \quad , \quad f_m(x) = \underbrace{f_0}_{\text{нач. дерево}}(x) + h_1(x) + \ldots + h_m(x)$$

Штрафная функция + регулятор. = штрафы (ф-ия потерь)

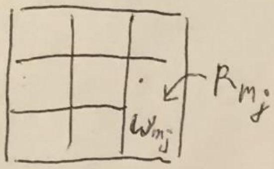
$$R(f) = \sum_{i=1}^{N} \mathcal{L}\left(f(x^{(i)}), y^{(i)}\right) + \sum_{m=1}^{M} \Omega(h_m) \to \min$$

Ω - огр. сложн. дерева, его ветвистость

Сложность будем мерять след. образом:

$$\Omega(h_m) = \gamma T_m + \frac{\lambda}{2} \sum_{j=1}^{T_m} w_{mj}^2 \quad , \quad h_m(x) = \sum_{j=1}^{T_m} w_{mj}\, \mathbb{I}(x \in R_{mj})$$

- T_m – кол-во листьев в дереве
- w_{mj} – вес (значение) в листе (ящике) R_{mj} $\to$



- γ, λ – гиперпараметры
- m – номер дерева, j – номер ящика

Т.е штрафуем за „сложность" дерева и за веса w_{mj}

Жадный алгоритм:

$$R_m(f) = \sum_{i=1}^{N} \mathcal{L}\Big(\underbrace{f_{m-1}(x^{(i)})}_{\substack{\text{уже найдены,}\\ \text{часть ансамбля}\\ \text{есть — мы жадины!}}} + \underbrace{h_m(x^{(i)})}_{\text{хотим дальше добавлять для min}}, y^{(i)}\Big) + \Omega(h_m) \longrightarrow \min$$

Разложим $\mathcal{L}(\cdot)$ в ряд Тейлора (до членов 2-го порядка), в окрест. точки $f_{m-1}(x^{(i)})$ N-мерного пр-ва:

$$\begin{pmatrix} f_{m-1}(x^{(1)}) \\ f_{m-1}(x^{(2)}) \\ \vdots \\ f_{m-1}(x^{(N)}) \end{pmatrix}$$

Т.е по принципу $f(x+\delta x) \simeq f(x) + f(x)'\delta x + \frac{1}{2} f(x)''(\delta x)^2$ получим:

$$R_m(f) \simeq \sum_{i=1}^{N} \left(\mathcal{L}\left(f_{m-1}(x^{(i)}), y^{(i)}\right) + h_m(x^{(i)}) \frac{\partial \mathcal{L}}{\partial f_{m-1,i}} + \frac{h_m^2(x^{(i)})}{2} \frac{\partial^2 \mathcal{L}}{\partial f_{m-1,i}^2} \right) + \Omega(h_m)$$

Для того, чтобы не писать квадратичную форму с част. произв., предположим, что матр H полож. опред., поэтому можно занулить

Внедиагональные эл-ты Гессиана.

- Перепишем еще раз, добавив регул. ф-ию с параметрами:

$$R_m(f) \approx \sum_{i=1}^{N} \left(\mathcal{L}\left(f_{m-1}(x^{(i)}), y^{(i)}\right) + h_m(x^{(i)}) \frac{\partial \mathcal{L}}{\partial f_{m-1,i}} + \frac{h_m^2(x^{(i)})}{2} \cdot \frac{\partial^2 \mathcal{L}}{\partial f_{m-1,i}^2} \right) + \gamma T_m + \frac{\lambda}{2} \sum_{j=1}^{T_m} w_{mj}^2$$

теперь по методу автора преобразуем, уйдя от суммы по объектам к сумме по ящикам R_{mj}, первое слагаемое останется без изменений:

$$\sum_{i=1}^{N} \mathcal{L}(f_{m-1},(x^{(i)}, y^{(i)}) + \sum_{j=1}^{T_m} \Bigg(w_{mj} \underbrace{\sum_{i \in R_{mj}} \frac{\partial \mathcal{L}}{\partial f_{m-1,i}}}_{G_{mj}} + \frac{1}{2} \Big(\lambda + \underbrace{\sum_{i \in R_{mj}} \frac{\partial^2 \mathcal{L}}{\partial f_{m-1,i}^2}}_{H_{mj}} \Big) w_{mj}^2 \Bigg) + \gamma T_m \to \min_{R_{mj}, w_{mj}}$$

Обозначим суммы по ящикам как G_{mj} и H_{mj} и перейдем к задаче $\min\limits_{R_{mj}} \left(\min\limits_{w_{mj}} \right)$

[G_{mj} и H_{mj} можно найти, выбрав подход. ф-ию штрафа и посчитать эти частн. производные]

Если уже знаем топологию дерева, т. е знаем R_{mj}, то для min второго слагаемого (т.к 1-ое и 3-ее будут const) то нужно рассм. каждый ящик:

$$w_{mj} \cdot G_{mj} + \frac{1}{2} (\lambda + H_{mj}) \cdot w_{mj}^2 \longrightarrow \min$$

это кв ф-ия, поэтому ищем вершину параболы:

$$G_{mj} + (\lambda + H_{mj}) w_{mj} = 0 \implies w_{mj}^* = - \frac{G_{mj}}{H_{mj} + \lambda}$$

теперь перепишем ф-ию штрафа, подставив w_{mj}^* и преобразуем:

$$R_m^* = -\frac{1}{2} \sum_{j=1}^{T_m} \frac{G_{mj}^2}{H_{mj} + \lambda} + \gamma T_m$$

Строим дерево.