Фамилия:..

Имя:..

Группа:..

Задача №1

У вас имеются следующие данные (обучающая выборка):

Наблюдение	1	2	3	4	5
T_i – покупка	1	1	0	1	0
X_{1i} – скидка	1	1	0	0	1
X_{2i} – друзья	1	0	1	0	2
Y_i – удовлетворенность	200	182	146	270	100

С помощью градиентного бустинга вы прогнозируете покупку пользователем платной подписки на ваше приложение. В качестве признаков используется число друзей индивида, зарегистрированных в приложении, а также факт предложения индивиду скидки. В качестве базовой модели рассматривается метод ближайших соседей с 4 соседями и метрикой расстояния Манхэттен (без нормализации данных). Совершается 1 итерация алгоритма со скоростью обучения 0.6. Применяется логистическая функция потерь:

$$\mathrm{L}(T, F) = \frac{1}{n}\sum_{i=1}^{n} -T_i \ln(F_i) - (1 - T_i)\ln(1 - F_i)$$

Для прогнозирования остатков (домноженных на минус единицу градиентов) используется регрессионное дерево глубины 2, в котором разбиения осуществляются на основании средневзвешенной дисперсии.

Примечание: при взяти градиента функции потерь обратите внимание, что он скорректирован на объем выборки n.

1. Рассчитайте точность (accuracy) базовой модели на обучающей выборке в случае, когда 1 прогнозируется для индивидов с оценкой условной вероятности покупки не меньше 0.5. **(5 баллов)**

2. С помощью описанного градиентного бустинга оцените условную вероятность того, что индивид без скидки и с одним другом приобретет платную подписку. **(10 баллов)**

3. Повторите предыдущий пункт, заменив в градиентном бустинге градиентный спуск на метод Ньютона (догадайтесь, как в таком случае может быть модифицирован метод) и снизив скорость обучения до 0.2. **(15 баллов)**

4. Используя взвешивание на обратные условные вероятности (inverse probability weighting), оцените средний эффект воздействия покупки приложения на удовлетворенность от его использования. В качестве метода оценивания условных вероятностей используйте обученный ранее градиентный бустинг (с градиентным спуском). **(10 баллов)**

Решение

Соберем результаты основных расчетов всех пунктов в таблицу[1]:

Наблюдение	1	2	3	4	5
T_i – покупка	1	1	0	1	0
X_{1i} – скидка	1	1	0	0	1
X_{2i} – друзья	1	0	1	0	2
Y_i – Удовлетворенность	200	182	146	270	100
$F_0(x) = \hat{P}_{\text{KNN}}\left(Y_i = 1 \vert X_{1i} = x_1, X_{2i} = x_2\right)$	0.5	0.75	0.75	0.75	0.5
$\hat{T}_{(0)}$ (прогноз базовой модели)	1	1	1	1	1
L_i'	-0.4	-4/15	0.8	-4/15	0.4
r_{1i}	0.4	4/15	-0.8	4/15	-0.4
$h_1(x_i)$	0	4/15	-0.8	4/15	0
$F_1(x) = \hat{P}_{\text{GB}}\left(T_i = 1 \vert X_{1i} = x_1, X_{2i} = x_2\right)$	0.5	0.91	0.27	0.91	0.5
L_i'' (приблизительно)	0.8	0.36	3.2	0.36	0.8
$r_{1i}^* = -L_i'/L_i''$	0.5	0.75	-0.25	0.75	-0.5
$h_1^*(x_i)$	-0.125	-0.75	-0.125	-0.75	0.5
$F_1^*(x) = \hat{P}_{\text{GB2}}\left(T_i = 1 \vert X_{1i} = x_1, X_{2i} = x_2\right)$	0.475	0.6	0.725	0.6	0.6

1. Обучение базовой модели очевидно. Из результатов обучения, представленных в таблице, следует, что поскольку все оценки условных вероятностей оказались не меньше 0.5, то, с учетом рассматриваемого порога прогнозирования, все прогнозы равны 1. Следовательно, точность базовой модели на обучающей выборке равняется $\text{ACC} = 0.6$, поскольку из 5 наблюдений верно были спрогнозированы 3.

2. Сперва найдем производную функции потерь по значениям модели:

$$L_i' = \frac{\partial L}{\partial F_i} = \frac{1}{n}\left[-\frac{T_i}{F_i} + \frac{1 - T_i}{1 - F_i}\right]$$

   Остатки модели рассчитываются как обратные градиенты $r_{1i} = -L_i'$. Например, для первого наблюдения получаем:

$$r_{11} = \frac{1}{n}\left[-\frac{T_1}{F_0(1,1)} + \frac{1-1}{1 - F_0(1,1)}\right] = \frac{1}{5}\left[-\frac{1}{0.5} + \frac{1-1}{1-0.5}\right] = 0.4$$

   Далее обучим регрессионное дерево прогнозировать остатки r_{1i} с помощью признаков X_{1i} и X_{2i}. Найдем оптимальное **первое** разбиение. Обратим внимание, что при разбиении по X_{1i} получаем:

$$\overline{r}_0 = \frac{-0.8 + \frac{4}{15}}{2} = -\frac{4}{15} \qquad \overline{r}_1 = \frac{0.4 + \frac{4}{15} - 0.4}{3} = \frac{4}{45}$$

$$\widehat{\text{Var}}\left(r_i, X_{1i}\right) = \frac{1}{5}\left[\left(0.4 - \frac{4}{45}\right)^2 + \left(\frac{4}{15} - \frac{4}{45}\right)^2 + \left(-0.4 - \frac{4}{45}\right)^2 + \left(-0.8 + \frac{4}{15}\right)^2 + \left(\frac{4}{15} + \frac{4}{15}\right)^2\right] =$$
$$= \frac{632}{3375} \approx 0.1873$$

[1]Расчеты можно найти по ссылке https://colab.research.google.com/drive/1bQ7vBbEWrbiqd0ubsrZzvIEObDNrCeEK#scrollTo=rijD4vXVNaLZ.

Разбиение по X_{2i} возможно путем объединения категорий $0, 1$ или $1, 2$. В первом случае, рассматривая переменную $I(X_{2i} < 2)$, получаем:

$$\overline{r}_1 = \frac{0.4 + \frac{4}{15} - 0.8 + \frac{4}{15}}{4} = \frac{1}{30} \qquad \overline{r}_0 = -0.4$$

$$\widehat{\mathrm{Var}}\left(r_i, I(X_{2i} < 2)\right) =$$

$$= \frac{1}{5}\left[\left(0.4 - \frac{1}{30}\right)^2 + \left(\frac{4}{15} - \frac{1}{30}\right)^2 + (-0.4 + 0.4)^2 + \left(-0.8 - \frac{1}{30}\right)^2 + \left(\frac{4}{15} - \frac{1}{30}\right)^2\right] =$$

$$= \frac{211}{1125} \approx 0.1876$$

В случае с разбиением по переменной $I(X_{2i} < 1)$ имеем:

$$\overline{r}_1 = \frac{\frac{4}{15} + \frac{4}{15}}{2} = \frac{4}{15} \qquad \overline{r}_0 = \frac{0.4 + -0.8 - 0.4}{3} = -\frac{4}{15}$$

$$\widehat{\mathrm{Var}}\left(r_i, I(X_{2i} < 1)\right) =$$

$$= \frac{1}{5}\left[\left(0.4 + \frac{4}{15}\right)^2 + \left(-0.8 + \frac{4}{15}\right)^2 + \left(-0.4 + \frac{4}{15}\right)^2\right] = \frac{56}{375} \approx 0.1493$$

Следовательно, первое разбиение осуществляется по переменной $I(X_{2i} < 1)$, поскольку оно минимизирует средневзвешенную дисперсию.

Перейдем к определению **второго** разбиения. Очевидно, что для наблюдений, у которых $X_{2i} < 1$, дальнейшее разбиение не имеет смысла, поскольку они все равняются одному и тому же значению $\frac{4}{15}$. При разбиении оставшихся наблюдений по X_1 получаем:

$$\overline{r}_0 = -0.8 \qquad \overline{r}_1 = \frac{0.4 - 0.4}{2} = 0$$

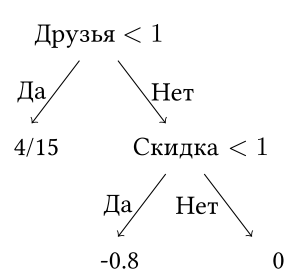
$$\widehat{\mathrm{Var}}\left(r_i, X_{1i}\right) = \frac{1}{3}\left[(0.4 - 0)^2 + (-0.4 - 0)^2\right] = \frac{8}{75} \approx 0.1067$$

При разбиении по $I(X_{2i} = 1)$ имеем:

$$\overline{r}_0 = -0.4 \qquad \overline{r}_1 = \frac{0.4 - 0.8}{2} = -0.2$$

$$\widehat{\mathrm{Var}}\left(r_i, I(X_{2i} = 1)\right) = \frac{1}{3}\left[(0.4 + 0.2)^2 + (-0.8 + 0.2)^2\right] = \frac{6}{25} \approx 0.24$$

Следовательно, второе разбиение необходимо делать по X_1. Таким образом, мы получили следующее регрессионное дерево, то есть модель $h_1(x_i)$:



Наконец, получим прогноз:

$$\hat{P}_{\text{GB}}\left(T_i = 1|X_{1i} = 0, X_{2i} = 1\right) = F_1(0,1) = F_0(0,1) + 0.6 \times h_1(0,1) = 0.75 + 0.6 \times (-0.8) = 0.27$$

3. Нетрудно показать, что вторая производная функции потерь по значениям модели будет иметь вид:

$$\text{L}''(T_i, F_i) = \frac{1}{n}\left(\frac{T_i}{F_i^2} + \frac{1 - T_i}{(1 - F_i)^2}\right)$$

В результате решение аналогично предыдущему пункту, за тем лишь исключением, что вместо отрицательных градиентнов прогнозируются:

$$r_{1i}^* = -\frac{\text{L}'(T_i, F_i)}{\text{L}''(T_i, F_i)}$$

В итоге пользуясь посчитанными в таблице значениями получаем:

$$\hat{P}_{\text{GB2}}\left(T_i = 1|X_{1i} = 0, X_{2i} = 1\right) = 0.725$$

4. Используя оцененные ранее условные вероятности получаем:

$$\widehat{\text{ATE}} = \frac{1}{5} \times \left(\frac{200}{0.5} + \frac{182}{0.91} - \frac{146}{1 - 0.27} + \frac{270}{0.9} - \frac{100}{1 - 0.5}\right) = 100$$

Задача №2

Нейтральная к риску фирма прогнозирует уход клиентов (1 - ушел, 0 - остался). С каждого оставшегося клиента фирма получает 5 тысяч рублей, а с ушедшего – ничего. Если фирма считает (на основании прогнозов модели), что клиент хочет уйти, то она дает ему 2 тысячи рублей и в таком случае клиент гарантированно остается.

1. Запишите цены различных видов прогнозов для фирмы. **(5 баллов)**

2. Вы – сотрудник рассматриваемой фирмы. Ваш кот запрыгнул на клавиатуру и случайно (а может и нет) безвозвратно удалил исходные данные, включавшие 100 наблюдений. Однако, вы запомнили рассчитанные по этим данным метрики качества прогнозов модели:

$$\text{precision} = 0.25 \qquad \text{recall} = 0.2 \qquad \text{ACC} = 0.3 \qquad \text{FPR} = 0.6$$

Рассчитайте прибыль от прогнозов модели на удаленных данных. **(15 баллов)**
Подсказка: сперва удобно составить матрицу ошибок (путаницы).

3. По удаленным котом данным (из предыдущего пункта) у вас была составлена таблица с расчетами выигрыша (gain) и подъема (lift). При этом вы разбивали выборку не на 10 (децили), а по аналогии на 5 равных частей. Однако, помахав лапами, кот удалил часть значений в этой таблице.

Часть	Выигрыш	Кумулятивный выигрыш	Подъем	Кумулятивный подъем
1				1.5
2		0.56		
3	0.16			
4			0.8	
5				

Восстановите значения, пропущенные в таблице. **(10 баллов)**
Подсказка: исходя из решения предыдущего пункта определите, сколько всего было ушедших клиентов в данных.

4. У вас была обучена нейросеть с 2 нейронами во входном слое (признаками), 1 скрытым слоем с 200 нейронами и функцией активации ReLU, а также выходным слоем с сигмоидной функцией активации. Кот благополучно стер веса вашей нейросети. Однако, вы помните, что все смещения (константы) равнялись нулю, а остальные веса равнялись одному и тому же положительному числу. Наконец, вам посчастливилось запомнить, что при подаче на вход нейросети наблюдения со значениями признаков $x_1 = 0.02$ и $x_2 = -0.015$ в выходном слое вы получали значение $\frac{1}{1+e^{-9}}$. Найдите утраченные веса нейросети. **(10 баллов)**

Решение

1. Из содержательной постановки задачи следует, что $P_{\text{TP}} = 3$, $P_{\text{TN}} = 5$, $P_{\text{FP}} = 3$, $P_{\text{FN}} = 0$.

2. Из условия нам известно, что:

$$0.25 = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad 0.2 = \frac{\text{TP}}{\text{TP} + \text{FN}}$$
$$0.3 = \frac{\text{TP} + \text{TN}}{100} \qquad 0.6 = \frac{\text{FP}}{\text{TN} + \text{FP}}$$

Решая соответствующую систему получаем, что $\text{TP} = 10$, $\text{TN} = 20$, $\text{FP} = 30$ и $\text{FN} = 40$, откуда:

$$\text{Profit} = 3 \times 10 + 5 \times 20 + 3 \times 30 + 0 \times 40 = 220$$

3. Общее число ушедших клиентов равняется $\text{TP} + \text{FN} = 10 + 40 = 50$. Через s_k обозначим количество ушедших клиентов, попавших в k-ю часть выборки.

   Из первой строки мы знаем, что кумулятивный подъем равняется 1.5, а значит подъем также равняется 1.5. Кроме того, исходя из принципа расчета подъема мы получаем выигрыш, который очевидно, будет совпадать с кумулятивным выигрышем:

$$\text{lift}(1) = 1.5 = \frac{s_1}{50/5} \implies s_1 = 15 \implies \text{gain}(1) = \frac{15}{50} = 0.3$$

   Из второй строки мы знаем, что кумулятивный выигрыш равняется $\text{cumulative gain}(2) = 0.56$, откуда:

$$\text{gain}(2) = \text{cumulative gain}(2) - \text{cumulative gain}(1) = 0.56 - 0.3 = 0.26$$

   Из полученного результата следует, что:

$$\text{gain}(2) = \frac{s_2}{50} \implies s_2 = 13 \implies$$
$$\implies \text{lift}(2) = \frac{13}{50/5} = 1.3 \qquad \text{cumulative lift}(2) = \frac{15+13}{2 \times (50/5)} = 1.4$$

   По аналогии восстанавливая остальные строки получаем таблицу:

Часть	Выигрыш	Кумулятивный выигрыш	Подъем	Кумулятивный подъем
1	0.3	0.3	1.5	1.5
2	0.26	0.56	1.3	1.4
3	0.16	0.72	0.8	1.2
4	0.16	0.88	0.8	1.1
5	0.12	1	0.6	1

4. Из условия известно, что все веса нейросети равнялись одному и тому же положительному числу, которое мы обозначим как $\omega > 0$.

   Рассчитаем значения нейронов скрытого слоя:

$$h_i = \max(0, \omega x_1 + \omega x_2) = \max(0, \omega \times 0.02 + \omega \times (-0.015)) = 0.005\omega, \text{ где } i \in \{1, \dots, 200\}$$

   Используя информацию о значении выходного слоя получим веса:

$$o = \frac{1}{1 + e^{-(\omega \times 0.005\omega + \ldots + \omega \times 0.005\omega)}} = \frac{1}{1 + e^{-(\omega h_1 + \ldots + \omega h_{200})}} = \frac{1}{1 + e^{-200 \times \omega \times 0.005\omega}} =$$
$$= \frac{1}{1 + e^{-\omega^2}} = \frac{1}{1 + e^{-9}} \implies e^{-\omega^2} = e^{-9} \implies \omega = 3$$

Задача №3

По выборке из независимых и одинаково распределенных наблюдений вы хотите оценить относительный средний эффект воздействия (ratio average treatment effect) и условный относительный средний эффект воздействия (термины были придуманы специально для этой задачи):

$$\text{RTE} = \text{E}\left(\frac{Y_{1i}}{Y_{0i}}\right) \qquad \text{CRTE}_i = \text{E}\left(\frac{Y_{1i}}{Y_{0i}}|X_i\right)$$

Предположим, что при фиксированных контрольных переменных X_i потенциальные исходы Y_{1i}, Y_{0i} и переменная воздействия T_i попарно независимы (все три случайные величины независимы между собой при условии контрольных переменных). Также допустим, что $0 \notin \text{supp}(Y_{0i})$.

Подсказка: если случайные величины A и B независимы, то независимы и случайные величины $f(A)$ и $g(B)$, где $f(.)$ и $g(.)$ – функции, определенные на носителях A и B соответственно. Даже если A и B независимы, то в общем случае $\text{E}\left(\frac{A}{B}\right) \neq \frac{\text{E}(A)}{\text{E}(B)}$.

1. Напомним, что условный средний эффект воздействия можно записать как функцию от двух условных математических ожиданий, первое из которых зависит (не считая контрольных переменных) только от Y_{1i}, а второе – только от Y_{0i}:

$$\text{CATE}_i = \text{E}\left(Y_{1i} - Y_{0i}|X_i\right) = \underbrace{\text{E}\left(Y_{1i}|X_i\right)}_{\text{зависит только от } Y_{1i}} - \underbrace{\text{E}\left(Y_{0i}|X_i\right)}_{\text{зависит только от } Y_{0i}}$$

   По аналогии, опираясь на предпосылки, описанные в условии задачи, запишите CRTE_i как функцию от двух условных математических ожиданий, одно из которых зависит только от Y_{1i}, а другое – только от Y_{0i}. **(2 баллов)**

2. Напомним, что при допущении об условной независимости $\text{E}(Y_{0i}|X_i) = \text{E}(Y_i|X_i, T_i = 0)$. Опираясь на предпосылки, описанные в условии задачи, получите аналогичные выражения для $\text{E}(Y_{1i}|X_i)$ и $\text{E}(1/Y_{0i}|X_i)$, а также подробно распишите их вывод. **(3 баллов)**

3. Опираясь на предпосылки, описанные в условии задачи, а также на результаты предыдущих пунктов, предложите и опишите аналог T-learner для оценивания CRTE_i. Обоснуйте состоятелость оценки описанного вами метода. **(10 баллов)**

4. Опираясь на результаты предыдущего пункта и предпосылки, описанные в условии задачи, предложите и опишите аналог T-learner для оценивания RTE. Обоснуйте состоятелость оценки описанного вами метода. **(5 баллов)**

Решение

1. Поскольку Y_{1i} и Y_{0i} независимы при фиксированных X_i, то независимыми при соответствующем условии являются также Y_{1i} и $1/Y_{0i}$, откуда:

$$\text{E}(Y_{1i}/Y_{0i}|X_i) = \text{E}(Y_{1i} \times (1/Y_{0i})|X_i) = \text{E}(Y_{1i}|X_i) \times \text{E}(1/Y_{0i}|X_i)$$

2. Поскольку при фиксированных X_i потенциальные исходы Y_{1i} и Y_{0i} не зависят от T_i, получаем:

$$\text{E}(Y_{1i}|X_i) = \text{E}(Y_{1i}|X_i, T_i = 1) = \text{E}(T_iY_{1i} + (1 - T_i)Y_{0i}|X_i, T_i = 1) = \text{E}(Y_i|X_i, T_i = 1)$$
$$\text{E}(1/Y_{0i}|X_i) = \text{E}(1/Y_{0i}|X_i, T_i = 0) = \text{E}(1/(T_iY_{1i} + (1 - T_i)Y_{0i})|X_i, T_i = 0) = \text{E}(1/Y_i|X_i, T_i = 0)$$

3. Сперва методами машинного обучения необходимо оценить условные математические ожидания. Во-первых, оценить $\hat{\mathrm{E}}(Y_i|X_i, T_i = 1)$ по выборке из таких i, что $T_i = 1$. Во-вторых, оценить $\hat{\mathrm{E}}(1/Y_i|X_i, T_i = 0)$ по выборке из таких i, что $T_i = 0$, в качестве целевой переменной рассматривая $1/Y_i$. Далее с помощью моделей, обученных оценивать эти условные математические ожидания, для каждого наблюдения в выборке можно получить оценку:

$$\widehat{\mathrm{CRTE}}_i = \hat{\mathrm{E}}(Y_i|X_i, T_i = 1) \times \hat{\mathrm{E}}(1/Y_i|X_i, T_i = 0)$$

Если оценки условных математических ожиданий состоятельны, то с учетом введенных предпосылок по теореме Слуцкого оценка $\widehat{\mathrm{CRTE}}_i$ также окажется состоятельной, поскольку:

$$\widehat{\mathrm{CRTE}}_i \xrightarrow{p} \mathrm{E}(Y_i|X_i, T_i = 1) \times \mathrm{E}(1/Y_i|X_i, T_i = 0) = \mathrm{E}(Y_{1i}|X_i) \times \mathrm{E}(1/Y_{0i}|X_i) = \mathrm{E}(Y_{1i}/Y_{0i}|X_i)$$

4. Состоятельную оценку $\widehat{\mathrm{RTE}}$ можно получить усреднив состоятельные оценки $\widehat{\mathrm{CRTE}}_i$:

$$\widehat{\mathrm{RTE}} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\mathrm{CRTE}}_i \xrightarrow{p} \mathrm{E}\left(\mathrm{E}(Y_{1i}/Y_{0i}|X_i)\right) = \mathrm{E}(Y_{1i}/Y_{0i}) = \mathrm{RTE}$$

Проверка в R:

```
# Число наблюдений
n <- 100000

# Вспомогательные переменные
u1 <- rpois(n, lambda = 1)
u2 <- rpois(n, lambda = 1)
u3 <- rpois(n, lambda = 1)
b1 <- rbinom(n, 1, 0.5)
b2 <- rbinom(n, 1, 0.5)

# Контрольные переменные
x1 <- pmin(b1 * u1 + (1 - b1) * u2, 3)
x2 <- pmin(b2 * u1 + (1 - b2) * u3, 3)

# Случайные ошибки
eps1 <- abs(rnorm(n))
eps0 <- abs(rnorm(n))

# Потенциальные исходы
y1 <- 1 + x1 + 0.5 * x2 + eps1
y0 <- 2 + 0.5 * x1 + 0.5 * x2 + eps0

# Переменная воздействия
t <- rbinom(n, 1, pt(x2 / max(x2) - 2 * x1 / max(x1), df = 5, ncp = 0.1))

# Целевая переменная
y <- y1 * t + y0 * (1 - t)

# Обартное значение
```

```
y_inv = 1 / y

# Значения
x1_vals <- unique(x1)
x2_vals <- unique(x2)
t_vals  <- unique(t)

# Число значений
n_x1 <- length(x1_vals)
n_x2 <- length(x2_vals)
n_t  <- length(t_vals)

# Оценки условных математических ожиданий
y_est     <- rep(NA, n)
y_inv_est <- rep(NA, n)
for (i in 1:n_x1)
{
  for (j in 1:n_x2)
  {
    cond           <- (x1 == x1_vals[i]) & (x2 == x2_vals[j])
    y_est[cond]     <- mean(y[cond & (t == 1)])
    y_inv_est[cond] <- mean(y_inv[cond & (t == 0)])
  }
}

# Пункт 1
cond <- (x1 == 2) & (x2 == 3)
c('E(y1/y0|X)' = mean((y1 / y0)[cond]),
  'E(y1|X) * E(1/y0|X1, X2)' = mean(y1[cond]) * mean(1 / y0[cond]))

# Пункт 2
cond <- (x1 == 2) & (x2 == 3)
c('E(y1|X)' = mean(y1[cond]),
  'E(y|X, T = 1)' = mean(y[cond & (t == 1)]))
c('E(1/y0|X)' = mean((1 / y0)[cond]),
  'E(1/y|X, T = 0)' = mean(1 / y[cond & (t == 0)]))

# Пункт 3
cond <- (x1 == 2) & (x2 == 3)
CRTE <- mean((y1 / y0)[cond])
CRTE_est <-  mean(y[cond & (t == 1)]) * mean((1 / y)[cond & (t == 0)])
c("CRTE" = CRTE, "CRTE_est" = CRTE_est)

# Пункт 4
c("RTE" = mean(y1 / y0), "RTE_est" =  mean(y_est * y_inv_est))
```