

Фамилия:.....

Имя:.....

Группа:.....

Задача 1

Вы изучаете, как поведение (1 – хорошее, 0 – плохое) влияет на вес мешка с конфетами (в килограммах), который индивид получает от Деда Мороза на новый год:

$$\text{Конфеты}_i = \alpha \times \text{Поведение}_i + g(\text{Возраст}_i) + \varepsilon_i,$$

где i – индекс наблюдения (предполагаются i.i.d.), α – структурный параметр, $g(\cdot)$ – некоторая функция, ε_i – случайная ошибка. У вас имеются следующие данные:

Возраст _{<i>i</i>}	20	30	50
Поведение _{<i>i</i>}	1	1	0
Конфеты _{<i>i</i>}	2	4	8

1. Вы прогнозируете поведение индивида в зависимости **только** от возраста с помощью метода ближайших соседей с одним соседом и расстоянием Евклида (прогнозируется хорошее поведение, если оценка его условной вероятности не меньше 0.5). С помощью 3-х частной кросс-валидации оцените точность (ассигасу) прогнозов хорошего поведения. **(10 баллов)**
2. Вы прогнозируете вес мешка с конфетами в зависимости **только** от возраста индивида используя 1 шаг алгоритма градиентного бустинга. В градиентном бустинге скорость обучения равняется 0.5 и остатки прогнозируются с помощью регрессионного дерева глубины 1 (при наличии нескольких в равной степени пригодных порогов разбиения выборки выберите средний из них). Кроме того, функция потерь является квадратичной $L = \frac{1}{n} \sum_{i=1}^n \left(\widehat{\text{Конфеты}}_i - \text{Конфеты}_i \right)^2$. В качестве базовой модели используется выборочное среднее. С помощью 3-х частной кросс-валидации и градиентного бустинга оцените MSE прогнозов веса мешка с конфетами. **(30 баллов)**
3. Найдите оценку параметра α используя метод двойного машинного обучения с разделением выборки на 3 части для кросс-фиттинга. Для оценивания условных математических ожиданий и вероятностей используйте те же методы, что применялись в предыдущих пунктах. **(20 баллов)**

Решение

1. Получим оценки условных вероятностей хорошего поведения и прогнозы для каждого индивида в выборке:

$$\hat{\mathbb{P}}(\text{Поведение}_1 = 1 | \text{Возраст}_1 = 20) = 1 \implies \widehat{\text{Поведение}}_1 = 1$$

$$\hat{\mathbb{P}}(\text{Поведение}_2 = 1 | \text{Возраст}_2 = 30) = 1 \implies \widehat{\text{Поведение}}_2 = 1$$

$$\hat{\mathbb{P}}(\text{Поведение}_3 = 1 | \text{Возраст}_3 = 50) = 1 \implies \widehat{\text{Поведение}}_3 = 1$$

Отсюда следует, что $\text{ACC}_{\text{CV}} = \frac{2}{3}$.

2. Рассмотрим три случая.

Случай 1: обучающая выборка включает наблюдения 2 и 3, прогнозируется наблюдение 1.

Найдем базовую модель:

$$F_0(\text{Возраст}) = \frac{4 + 8}{2} = 6$$

Рассчитаем остатки:

$$r_1 = \frac{1}{2} \times 2 \times (6 - 4) = 2$$

$$r_2 = \frac{1}{2} \times 2 \times (6 - 8) = -2$$

Поскольку для прогнозирования применяется решающее дерево глубины 1, то:

$$h_1(\text{Возраст}) = \begin{cases} -r_1, & \text{если Возраст} \leq 40 \\ r_2, & \text{в противном случае} \end{cases} = \begin{cases} -2, & \text{если Возраст} \leq 40 \\ 2, & \text{в противном случае} \end{cases}$$

В результате получаем прогноз:

$$\widehat{\text{Конфеты}}_1 = F_1(20) = F_0(20) + 0.5 \times h_1(20) = 6 - 0.5 \times 2 = 5$$

Случай 2: обучающая выборка включает наблюдения 1 и 3, прогнозируется наблюдение 2.

Найдем базовую модель:

$$F_0(\text{Возраст}) = \frac{2 + 8}{2} = 5$$

Рассчитаем остатки:

$$r_1 = \frac{1}{2} \times 2 \times (5 - 2) = 3$$

$$r_2 = \frac{1}{2} \times 2 \times (5 - 8) = -3$$

Поскольку для прогнозирования применяется решающее дерево глубины 1, то:

$$h_1(\text{Возраст}) = \begin{cases} -r_1, & \text{если Возраст} \leq 35 \\ r_2, & \text{в противном случае} \end{cases} = \begin{cases} -3, & \text{если Возраст} \leq 35 \\ 3, & \text{в противном случае} \end{cases}$$

В результате получаем прогноз:

$$\widehat{\text{Конфеты}}_2 = F_1(30) = F_0(20) + 0.5 \times h_1(20) = 5 - 0.5 \times 3 = 3.5$$

Случай 3: обучающая выборка включает наблюдения 1 и 2, прогнозируется наблюдение 3.

Найдем базовую модель:

$$F_0(\text{Возраст}) = \frac{2 + 4}{2} = 3$$

Рассчитаем остатки:

$$r_1 = \frac{1}{2} \times 2 \times (3 - 2) = 1$$

$$r_2 = \frac{1}{2} \times 2 \times (3 - 4) = -1$$

Поскольку для прогнозирования применяется решающее дерево глубины 1, то:

$$h_1(\text{Возраст}) = \begin{cases} -r_1, & \text{если Возраст} \leq 25 \\ r_2, & \text{в противном случае} \end{cases} = \begin{cases} -1, & \text{если Возраст} \leq 25 \\ 1, & \text{в противном случае} \end{cases}$$

В результате получаем прогноз:

$$\widehat{\text{Конфеты}}_3 = F_1(50) = F_0(50) + 0.5 \times h_1(50) = 3 + 0.5 \times 1 = 3.5$$

Кросс-валидация

В результате получаем:

$$\text{MSE}_{\text{cv}} = \frac{(5 - 2)^2 + (3.5 - 4)^2 + (3.5 - 8)^2}{3} = \frac{59}{6} \approx 9.83$$

3. Осуществим расчет:

$$\hat{\alpha} = \frac{(5 - 2) \times (1 - 1) + (3.5 - 4) \times (1 - 1) + (3.5 - 8) \times (1 - 0)}{(1 - 1)^2 + (1 - 1)^2 + (1 - 0)^2} = -4.5$$

Задача 2

Вы изучаете, как наличие у снеговика-почтовика ведра на голове влияет на скорость доставки им почты (в часах). У вас имеются следующие данные:

Масса _{<i>i</i>}	2	5
Ведро _{<i>i</i>}	1	0
Время _{<i>i</i>}	2	8

Для оценивания математического ожидания времени доставки письма, условного на массу снеговика (в килограммах) и факт наличия или отсутствия у него ведра, применяется нейронная сеть с 1 скрытым слоем, в котором расположено 100 нейронов. Во всех случаях применяется функция активации ReLU, константы (смещения) отсутствуют, функция потерь является квадратичной, изначально все веса равняются 0.1. При этом обучаются только веса, относящиеся к стрелочкам, идущим от нейронов скрытого слоя к выходному (output) слою. Эти веса обучаются на всей выборке с помощью одного шага алгоритма градиентного спуска со скоростью обучения 1/12.

1. Нарисуйте используемую нейросеть и рассчитайте значение ее функции потерь, усреднив функции потерь, посчитанные для каждого наблюдения. **(5 баллов)**
2. Найдите веса нейросети после обучения. **(5 баллов)**
3. С помощью обученной нейросети оцените средний эффект воздействия наличия ведра на время доставки. **(10 баллов)**
4. С помощью обученной нейросети методом *S*-learner оцените условный средний эффект воздействия наличия ведра на время доставки для снеговика весом 10 килограмм. **(10 баллов)**
5. Оцените условный средний эффект воздействия из предыдущего пункта с помощью *T*-learner, заменив нейросеть на метод ближайших соседей с одним соседом и расстоянием Евклида. **(10 баллов)**

Решение

1. Найдем значение выходного слоя для каждого наблюдения до обучения весов:

$$\begin{aligned}o_1 &= 0.1 \times (2 \times 0.1 + 1 \times 0.1) \times 100 = 3 \\o_2 &= 0.1 \times (5 \times 0.1 + 0 \times 0.1) \times 100 = 5\end{aligned}$$

Вычислим значение функции потерь для каждого наблюдения:

$$\begin{aligned}l_1 &= (3 - 2)^2 = 1 \\l_2 &= (5 - 8)^2 = 9\end{aligned}$$

В итоге получаем:

$$L = \frac{1 + 9}{2} = 5$$

2. Вычислим производную функцию потерь по каждому из обучаемых весов:

$$\begin{aligned}\frac{\partial l_1}{\partial w} &= 2 \times (3 - 2) \times (2 \times 0.1 + 1 \times 0.1) = 0.6 \\ \frac{\partial l_2}{\partial w} &= 2 \times (5 - 8) \times (5 \times 0.1 + 0 \times 0.1) = -3 \\ \frac{\partial \mathbf{L}}{\partial w} &= \frac{0.6 - 3}{2} = -1.2\end{aligned}$$

Следовательно, после обновления данные веса будут равняться:

$$w^* = 0.1 - \frac{1}{12} \times (-1.2) = 0.2$$

3. Оценим условные математические ожидания:

$$\begin{aligned}\hat{\mathbb{E}}(\text{Время} | \text{Масса} = 2, \text{Ведро} = 1) &= 0.2 \times (2 \times 0.1 + 1 \times 0.1) \times 100 = 6 \\ \hat{\mathbb{E}}(\text{Время} | \text{Масса} = 2, \text{Ведро} = 0) &= 0.2 \times (2 \times 0.1 + 0 \times 0.1) \times 100 = 4 \\ \hat{\mathbb{E}}(\text{Время} | \text{Масса} = 5, \text{Ведро} = 1) &= 0.2 \times (5 \times 0.1 + 1 \times 0.1) \times 100 = 12 \\ \hat{\mathbb{E}}(\text{Время} | \text{Масса} = 5, \text{Ведро} = 0) &= 0.2 \times (5 \times 0.1 + 0 \times 0.1) \times 100 = 10\end{aligned}$$

В результате получаем оценку:

$$\widehat{\text{ATE}} = \frac{(6 - 4) + (12 - 10)}{2} = 2$$

4. Оценим условные математические ожидания:

$$\begin{aligned}\hat{\mathbb{E}}(\text{Время} | \text{Масса} = 10, \text{Ведро} = 1) &= 0.2 \times (10 \times 0.1 + 1 \times 0.1) \times 100 = 22 \\ \hat{\mathbb{E}}(\text{Время} | \text{Масса} = 10, \text{Ведро} = 0) &= 0.2 \times (10 \times 0.1 + 0 \times 0.1) \times 100 = 20\end{aligned}$$

В результате получаем оценку:

$$\widehat{\text{CATE}} = 22 - 20 = 2$$

5. Обратим внимание, что в данном случае независимо от массы:

$$\begin{aligned}\hat{\mathbb{E}}(\text{Время} | \text{Масса}, \text{Ведро} = 1) &= 2 \\ \hat{\mathbb{E}}(\text{Время} | \text{Масса}, \text{Ведро} = 0) &= 8\end{aligned}$$

Отсюда следует, что:

$$\widehat{\text{CATE}} = 2 - 8 = -6$$