

Машинное обучение в экономике

Семинар 6. Эффекты воздействия

Задание №1

Вы оцениваете эффект воздействия использования больших языковых моделей для написания кода на заработную плату программистов.

Индекс	1	2	3	4	5
Стаж _{<i>i</i>} (X_i)	1	2	5	10	12
Зарплата _{<i>i</i>} (Y_i)	8	4	6	2	12
БЯМ _{<i>i</i>} (T_i)	1	1	0	0	1
Картинки _{<i>i</i>} (Z_i)	1	0	1	0	1

В данных применяются следующие обозначения:

- **Стаж** – число лет работы программистом.
- **Зарплата** – ежемесячная заработная плата программиста в сотнях тысяч рублей.
- **БЯМ** – бинарная переменная, принимающая значение 1, если программист использует большие языковые модели для написания кода и 0 – в противном случае.
- **Картинки** – бинарная переменная, принимающая значение 1, если программист использует генеративный искусственный интеллект (ИИ) для создания изображений и 0 – в противном случае.

В качестве контрольной переменной рассматривается стаж. Для оценивания математических ожиданий, условных на контрольную переменную, применяется метод ближайших соседей с 2 соседями и расстоянием Манхэттен. В качестве альтернативы обычной стандартизации признаков при расчете расстояний вы домножаете бинарную переменную БЯМ на 5. Вас интересует эффект воздействия использования больших языковых моделей на заработную плату программиста.

1. Оцените АТЕ, используя разницу в математических ожиданиях зависимой переменной, условных лишь на факт получения воздействия. Объясните, на какую предпосылку опирается соответствующий подход и почему она может нарушаться в рассматриваемом случае.
2. Обсудите целесообразность домножения переменной БЯМ при расчете расстояний с помощью метода ближайших соседей.
3. Оцените АТЕ с помощью S-learner. Опишите предпосылку, на которую опирается данный метод, а также предположите, почему она может не соблюдаться.
4. Оцените АТЕ используя T-learner.

5. Оцените АТЕ, применив взвешивание на обратные условные вероятности. При этом, если оценка условной вероятности равняется 1 или 0, она заменяется на 0.8 или 0.2 соответственно. Объясните необходимость соответствующих замен.
6. Оцените АТЕ с использованием метода с двойной устойчивостью. Объясните преимущество данного подхода над предыдущими.
7. С помощью S-learner оцените CATE для индивида с 8-летним стажем работы.
8. Выберите инструментальную переменную для БЯМ. Содержательно обоснуйте ее валидность. Оцените LATE с помощью классического подхода без применения контрольной переменной (только с инструментальной). Объясните преимущества и недостатки соответствующего подхода.