MultiOutputRegressor в Scikit-learn

`MultiOutputRegressor` – мета-оценщик из библиотеки `scikit-learn`, который позволяет превратить любой регрессор, предназначенный для одномерного выхода, в регрессор с многомерным выходом. Реализуется следующим образом - для каждой целевой переменной создается отдельная модель, которая обучается на исходном наборе данных.

Пример кода:

model = LinearRegression()

multi\_output\_model = MultiOutputRegressor(model)

multi\_output\_model.fit(X, y)

В данном случае целевая переменная должна содержать минимум два значения:

(y = [

[500, 12],

[520, 11],

[480, 13],

...])

Физический смысл: пример - есть данные о погоде, и необходимо предсказать не только температуру, но и влажность на следующий день. В этом случае есть две целевые переменные, и `MultiOutputRegressor` позволяет обработать их одновременно, обучив две отдельные модели на одних и тех же входных данных.

Преимущества:

1. Удобство: нет необходимости вручную управлять созданием и обучением нескольких моделей. Обучение и прогнозирование можно выполнять через единый интерфейс.

2. Параллелизм: если базовый регрессор поддерживает параллельное выполнение, обучение может происходить одновременно для всех целевых переменных.

3. Гибкость: подходит для любого регрессора в `scikit-learn`, позволяя легко интегрировать многомерный выход в существующие рабочие процессы.

Недостатки:

1. Невозможность индивидуальной настройки: `MultiOutputRegressor` не позволяет делать разные настройки гиперпараметров для разных целевых переменных.

2. Ресурсоемкость: в общем случае при одновременном обучении нескольких моделей для каждой целевой переменной потребуется большая пиковая загрузка памяти.

Примеры сфер применения:

- Экология и климат: Прогнозирование нескольких погодных параметров на основе исторических данных.

- Агрономия: Предсказание различных характеристик урожая на основе данных о состоянии почвы и погодных условиях.

- Финансы: Прогнозирование нескольких экономических показателей на основе текущих данных рынка.

- Медицина: Прогнозирование нескольких показателей состояния пациента на основе медицинских измерений.

В контексте нашей задачи с подбором дозы варфарина применение MultiOutputRegressor не требуется, поскольку у нас в датасете нет еще одной целевой переменной, которую надо предсказать.

---