

Предикт с коэффициентами

29 мая 2023 г.

Поясним нашу формулу

На каждом этапе мы смотрим отношение $PI(\text{Факт})$ и $PI(\text{план})$ и делим его на количество прошедших этапов. Так мы делаем итерационно. Что это значит? Допустим, если мы посчитали это отношение в конце первого этапа, то мы поделим его на 1, т.к. прошел только один этап. В конце второго этапа мы сложим полученный коэффициент с 1-го этапа с текущим коэффициентом (заново пересчитаем $PI(\text{факт}) / PI(\text{план})$) и поделим сумму на 2, т.к. прошло два этапа и т.д.

В итоге мы получили модель, которая делает предсказание, но не очень хорошо. Хочется каждый этап в предикте домножать на какой-то коэффициент, который можно получить с помощью машинного обучения только по завершившимся проектам.

Введем коэффициенты

Итак, давайте делать предикт не просто смотря на каждом этапе на отношение $PI(\text{факт})$ и $PI(\text{план})$, а будем на разных этапах брать это отношение с разными коэффициентами, которые мы получили в результате линейной регрессии.

Немного математики

Что такое линейная регрессия? По сути мы хотим найти такие коэффициенты, на которые мы умножим предсказание и получим $PI(\text{факт})$.

Математически мы ищем вектор x в следующей задаче:

$$\left\| \begin{pmatrix} a_1 \\ \dots \\ a_n \end{pmatrix}^T * \begin{pmatrix} x_1 \\ \dots \\ x_n \end{pmatrix} - \begin{pmatrix} f_1 \\ \dots \\ f_n \end{pmatrix} \right\| \rightarrow \min$$

где a_i - это предсказанные часы на этапы, а f_i - фактические часы, затраченные на соответствующие этапы.

Для решения этой задачи будем использовать линейную регрессию, присутствующую в модулях numpy. Далее, находясь на каждом этапе будем предсказывать коэффициенты для следующих этапов и брать их среднее значение.

Покажем в таблице значения коэффициентов на каждый этап.

Например: мы находимся в конце этапа анализа, смотрим на столбец "момент предикта" и выбираем строку с анализом. В сумме будем брать предикт на этапы, умноженный на коэффициенты в данной строке.

Как будем считать:

$$C = \frac{1}{2} \cdot \left(\frac{TSM_1}{TPM_1} + \frac{TSM_2}{TPM_2} \right)$$

$$\text{Предикт} = TSM_1 + TSM_2 + C \cdot 0.95 \cdot TPM_3 + C \cdot 1.47 \cdot TPM_4 + C \cdot 1.744 \cdot TPM_5$$

Момент предикта	инициация	анализ	реализация	внедрение	завершение
инициация	-	-	-	-	-
анализ	1	1	0.95	1.47	1.744
реализация	1	1	1	1.47	1.83
внедрение	1	1	1	1	1.596