Моделирование временных рядов
Семинар 2
Табличные задачи

Борис Демешев, Матвей Зехов

1 Подход к моделированию

Основные подходы к моделированию временных рядов делятся на два направления. Первый предполагает рассматривать ряд именно как последовательность и пытаться моделировать породивший его случайный процесс в виде стохастического разностного или дифференциального уравнения. Под это направление попадают все основные классические методы: ETS, ARIMA, GARCH и т.д. Второй подход предлагает свести задачу к табличному виду и попытаться решить её классическими моделями машинного обучения. В текущей главе мы рассмотрим как раз второй подход и различные стратегии прогнозирования, которые он позволяет использовать.

2 Стратегии прогнозирования табличных данных

Предположим, что мы хотим решить задачу одношагового прогнозирования. Используем модель авторегрессии порядка k и одной внешней переменной x_t:

$$\hat{y}_{t+1} = \hat{f}(y_t, \ldots, y_{t-k}, x_t, \ldots, x_{t-p}), t \in \overline{1, T}$$

Под функцией f будем подразумевать любую стандартную модель регрессии. Подготовим обучающую выборку для $k = 3$, $p = 1$ и $T = 8$. Так как момент времени $t = 8$ уже наступил, нам известны значения y_8 и x_8. Оценим эту модель. Построить прогноз на один шаг не составит труда:

t	$\hat{y}_t$	y_t	y_{t-1}	y_{t-2}	y_{t-3}	x_{t-1}
4	-	4	3	2	1	15
5	-	5	4	3	2	20
6	-	6	5	4	3	12
7	-	7	6	5	4	17
8	-	8	7	6	5	30
9	9.5	-	8	7	6	50
10	-	-	?	8	7	?

Однако для прогноза $\hat{y}_{10}$ у нас недостаточно данных, неизвестны y_9 и x_9. Рассмотрим подходы, которые помогут обойти эту проблему.

§2.1 Рекурсивная стратегия

Предположим, что наша модель f оценивается очень долго и мы не можем себе позволить оценивать дополнительные модели. Значит необходимо обойтись уже оценённой одношаговой моделью. Попытаемся аппроксимировать неизвестный y_9 наиболее очевидным образом, то есть полученным на предыдущем шаге прогнозом. Стратегию можно записать следующим образом.

$$\hat{y}_{t+h} = \hat{f}(\tilde{y}_{t+h-1}, \dots, \tilde{y}_{t+h-k}, \tilde{x}_{t+h}, \dots, \tilde{x}_{t+h-p}), t \in \overline{1, T}$$

$$\tilde{z}_t = \begin{cases} z_t & \text{если } t <= T \\ \hat{z}_t & \text{если } t > T \end{cases}$$

Из определения следует, что для внешней переменной x_t тоже необходимо каким-то образом получать прогнозы. Обычно для этого строят несложную вспомогательную модель.

t	$\hat{y}_t$	$\hat{x}_t$	y_t	y_{t-1}	y_{t-2}	y_{t-3}	x_{t-1}
4	-	-	4	3	2	1	15
5	-	-	5	4	3	2	20
6	-	-	6	5	4	3	12
7	-	-	7	6	5	4	17
8	-	-	8	7	6	5	30
9	9.5	55	-	8	7	6	50
10	10.5	40	-	9.5	8	7	55

Среди преимуществ рекурсивной стратегии можно назвать относительно низкий разброс. В случае, если у нас мало внешних регрессоров и мы не оцениваем на них большое количество вспомогательных моделей, итоговый прогнозный алгоритм получается довольно простым. Как следствие, оценить модель и построить прогноз можно довольно быстро. К недостаткам можно отнести высокое смещение. Сама модель $\hat{f}$ по построению обучается как одношаговая. Рекурсивные подстановки прогнозов будут очень быстро накапливать ошибку. Эта стратегия довольно плохо подходит для прогнозов на далёкие горизонты. Однако она может хорошо послужить для краткосрочного прогнозирования или в качестве бенчмарка с небольшим количеством регрессоров.

§2.2 Прямая стратегия

Теперь предположим, что модель f оценивать не очень дорого. Чтобы избежать рекурсивного прогнозирования, построим для каждого горизонта отдельную модель.

$$\hat{y}_{t+h} = \hat{f}_h(y_t, \dots, y_{t-k}, x_t, \dots, x_{t-p}), t \in \overline{1, T}$$

Для этого необходимо сформировать h обучающих выборок. Для $h = 2$ выборки будут выглядеть следующим образом:

t	y_t	y_{t-1}	y_{t-2}	y_{t-3}	x_{t-1}
4	4	3	2	1	15
5	5	4	3	2	20
6	6	5	4	3	12
7	7	6	5	4	17
8	8	7	6	5	30

t	y_{t+1}	y_{t-1}	y_{t-2}	y_{t-3}	x_{t-1}
4	5	3	2	1	15
5	6	4	3	2	20
6	7	5	4	3	12
7	8	6	5	4	17

Легко заметить, что мы просто сдвигаем вектор целевой переменной на один шаг вниз. С каждой итерацией мы теряем одно наблюдение, но на достаточно больших датасетах это не критично. Таким образом модель для горизонта h изначально обучается для прогнозов на h шагов вперёд. Прогноз для каждой модели будет рассчитываться из последней доступной точки:

t	h	$\hat{y}_{t+h}$	y_t	y_{t-1}	y_{t-2}	x_t
8	1	9.5	8	7	6	50
8	2	10.5	8	7	6	50
8	3	11	8	7	6	50