

A decorative graphic on the left side of the slide consisting of a network of thin, dark blue lines. These lines branch out and connect to small, open circles, resembling a circuit board or a neural network diagram. The lines and circles are arranged in a way that suggests a flow or connection from the top left towards the bottom left.

Временные ряды - 2

Кантонистова Е.О.

ВШЭ, 2021

МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ

- Часто затраты на настройку моделей (ARMA, ARIMA, SARIMA и др.) не окупаются, поэтому имеет смысл попробовать применить методы машинного обучения к предсказанию временных рядов.
- Можно, например, использовать линейную регрессию, в качестве признаков для которой использовать лаговые признаки (значения признака в предыдущие периоды времени). Кроме того, из признака времени можно выделить признаки дней недели, часов и т.д. Модель приобретает больший смысл, если кроме самих значений временного ряда у нас есть и другие признаки.

ЛИНЕЙНАЯ МОДЕЛЬ АВТОРЕГРЕССИИ

- признаки - n предыдущих наблюдений ряда:

$$\hat{y}_{t+1}(\mathbf{w}) = \sum_{j=1}^n \mathbf{w}_j y_{t-j+1}, \quad \mathbf{w} \in \mathbb{R}^n$$

- объекты - $t - n + 1$ моментов в истории ряда:

$$F_{\ell \times n} = \begin{pmatrix} y_{t-1} & y_{t-2} & y_{t-3} & \dots & y_{t-n} \\ y_{t-2} & y_{t-3} & y_{t-4} & \dots & y_{t-n-1} \\ \dots & \dots & \dots & \dots & \dots \\ y_n & y_{n-1} & y_{n-2} & \dots & y_1 \\ y_{n-1} & y_{n-2} & y_{n-3} & \dots & y_0 \end{pmatrix}, \quad y_{\ell \times 1} = \begin{pmatrix} y_t \\ y_{t-1} \\ \dots \\ y_{n+1} \\ y_n \end{pmatrix}$$

$$Q(\mathbf{w}) = \sum_{i=n}^t (\hat{y}_i(\mathbf{w}) - y_i)^2 = \|F\mathbf{w} - \mathbf{y}\|^2 \rightarrow \min_{\mathbf{w}}$$

МАШИННОЕ ОБУЧЕНИЕ: ПРИЗНАКИ

- 1) Лаги – значение временного ряда 1, 2, ... периодов назад
- 2) Агрегированные признаки по дате (среднее значение таргета для каждого дня недели, часа и т.д.)
- 3) Другие характеристики

КОДИРОВАНИЕ СРЕДНИМ ЗНАЧЕНИЕМ

Пример:

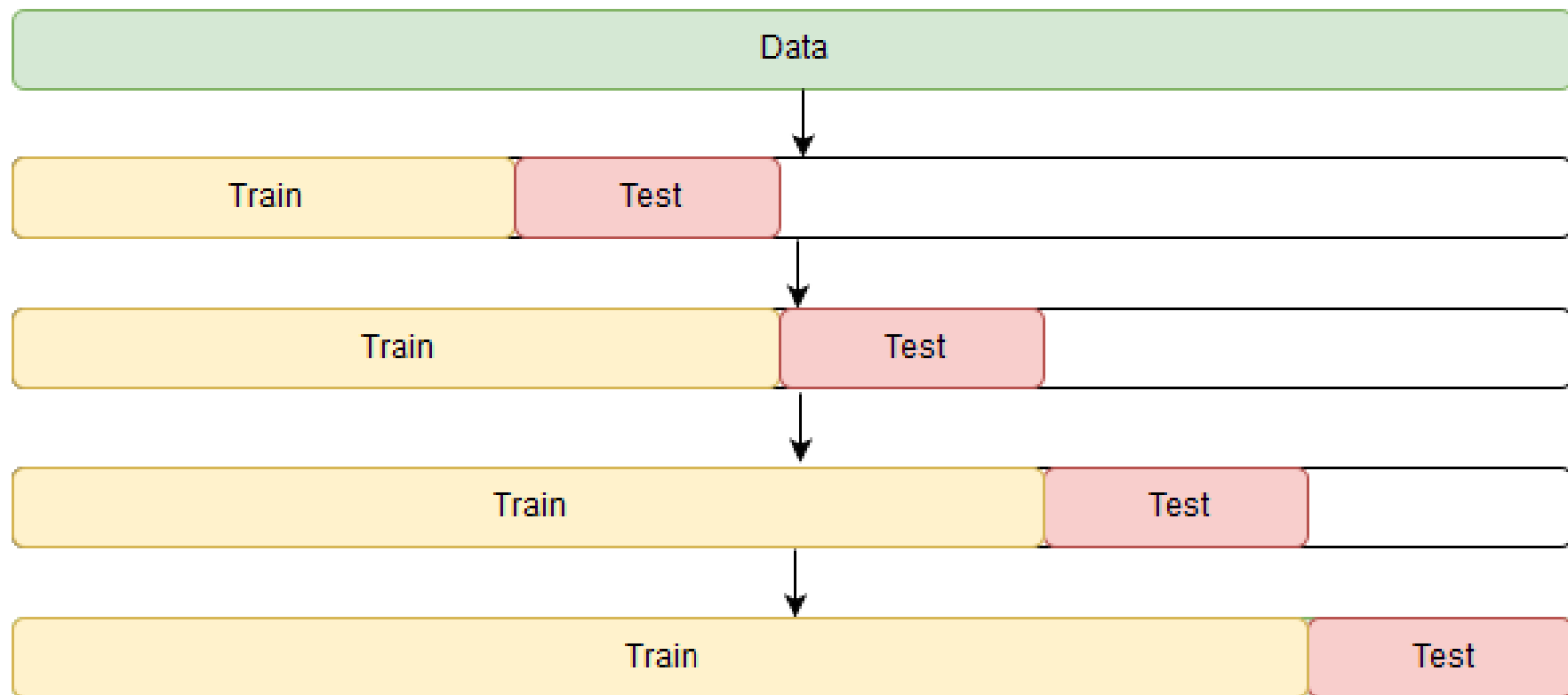
id	job	age	target
1	Doctor	54	1
2	Doctor	35	0
3	Doctor	28	1
4	Doctor	75	0
5	Teacher	29	1
6	Teacher	37	1
7	Engineer	60	0
8	Engineer	38	1
9	Waiter	31	1
10	Driver	22	0

КОДИРОВАНИЕ СРЕДНИМ ЗНАЧЕНИЕМ

Пример:

id	job	job_mean	target
1	Doctor	0,50	1
2	Doctor	0,50	0
3	Doctor	0,50	1
4	Doctor	0,50	0
5	Teacher	1	1
6	Teacher	1	1
7	Engineer	0,50	0
8	Engineer	0,50	1
9	Waiter	1	1
10	Driver	0	0

КРОСС-ВАЛИДАЦИЯ НА ВРЕМЕННЫХ РЯДАХ



ЛИТЕРАТУРА

- www.machinelearning.ru – всё по временным рядам, в частности:
- ✓ http://www.machinelearning.ru/wiki/images/archive/e/e7/20150323154210%21Psad_corr.pdf
- <https://www.coursera.org/lecture/data-analysis-applications/arma-fXTrB> и остальные лекции этого курса по теме
- Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов. Финансы и статистика, 2003.