1 Strategies of matrix construction

Suppose a large set of time series $\mathfrak{D} = \{\mathbf{s}\}$ is given. The "object-feature" matrix $\mathbf{X}^*$ for the multiscale autoregressive problem statement is composed of row-vectors

$$\mathbf{s}'_i = [\mathbf{y}'_i, \mathbf{x}'_i] = [\underbrace{s(t_i), \dots, s(t_i - \Delta t_\mathrm{r})}_{\mathbf{y}'_i}), \underbrace{\dots, s(t_i - \Delta t_\mathrm{r} - \Delta t_\mathrm{p})}_{\mathbf{x}'_i}],$$

where $s(t)$ is an element of time series $\mathbf{s}$. Consider several strategies to decompose time series $\mathbf{s}$ into segments $\Delta t_i = (t_i, \dots, t_i - \Delta t_\mathrm{r} - \Delta t_\mathrm{p})$ to construct matrix $\mathbf{X}^*$.

1. Row vectors $\mathbf{s}_i$ cover time series without intersections. Let $\{T_\mathrm{max}, \dots, 1\}$ be the set of indices of tim series $\mathbf{s}$, then thee strategy of selecting t_i holds the following:

$$\{T_\mathrm{max}, \dots, 1\} = \bigsqcup_{i=1}^{M} \Delta t_i. \tag{1}$$

   It follows from (1) that $|t_{i+1} - t_i| > \Delta t_\mathrm{r} + \Delta t_\mathrm{p}$ for any $i = 1, \dots, M-1$.

2. Row vectors $\mathbf{s}_i = [\mathbf{y}_i, \mathbf{x}_i]$ overlap, but target parts $\mathbf{y}_i$ do not intersect:

$$\{T_\mathrm{max}, \dots, 1\} = \bigsqcup_{i=1} M - 1\{t_i, \dots, t_i - \Delta t_\mathrm{r}\} \Rightarrow |t_{i+1} - t_i| > \Delta t_\mathrm{r}. \tag{2}$$

3. For each time stamp t_i of the least frequent regular sampling there is correspondent row vector $\mathbf{s}_i$ in $\mathbf{X}^*$:

$$\{T_\mathrm{max}, \dots, 1\} = \bigcup_i t_i.$$

4. Time intervals Δt_i are selected randomly.

5. Potentially, other sensible strategies are possible.

Vector $\boldsymbol{\varepsilon} \in \mathbb{R}^{\Delta t_\mathrm{r}}$ of model residuals at time stamp t_i is given by

$$\boldsymbol{\varepsilon}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i.$$

Dependent on the way the matrix $\mathbf{X}^*$ is designed, there might be dependencies between components of subsequent vectors $\mathbf{y}_i$, $\mathbf{y}_{i+1}$. If there are such i, $i' \in \mathcal{B}$ that $|t_i - t_{i'}| < \Delta t_\mathrm{r}$, vectors $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\varepsilon}_{i'}$ overlap or contain residuals for the same time stamp. In this case define the test vector of residuals as

$$\boldsymbol{\varepsilon}(\mathcal{B}) = \left\{ \bar{\varepsilon}_t \,\middle|\, t \in \bigcup_{i \in \mathcal{B}} \{i - \Delta t_\mathrm{r}, \dots, i\} = \{t_{i_\mathrm{min}} - \Delta t_\mathrm{r}, \dots, t_{i_\mathrm{max}}\} \right\},$$

where $\bar{\varepsilon}_t$ is the average residual for the time stamp t.

To avoid these issues, we fix the second strategy of the $\mathbf{X}$ construction.

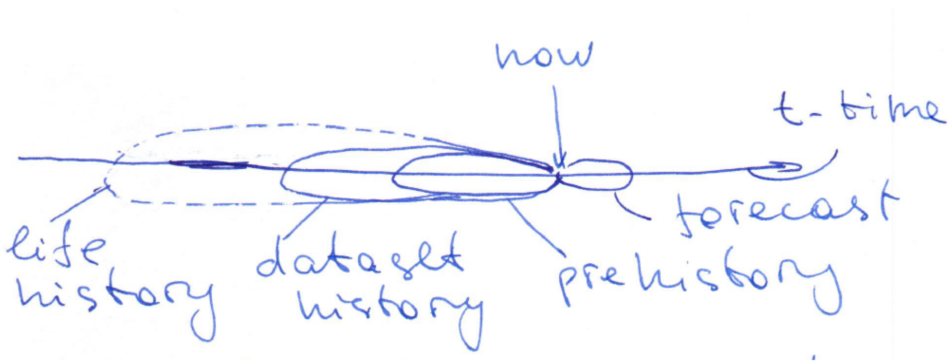

Figure 1: Online forecasting paradigm.

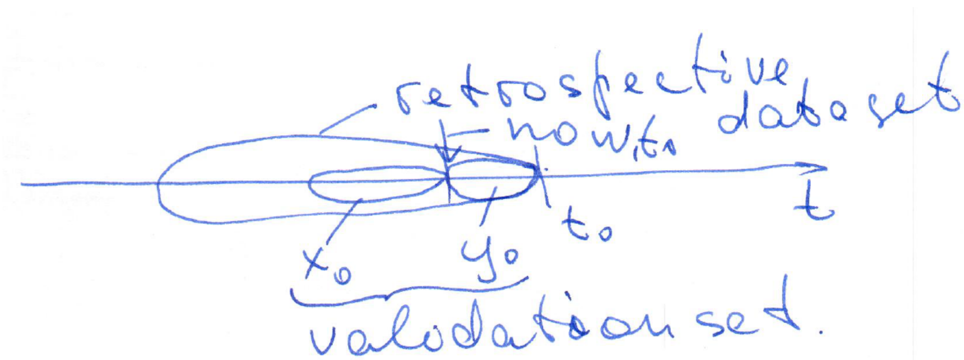

Figure 2: Retrospective forecast includes most recent samples in data set.

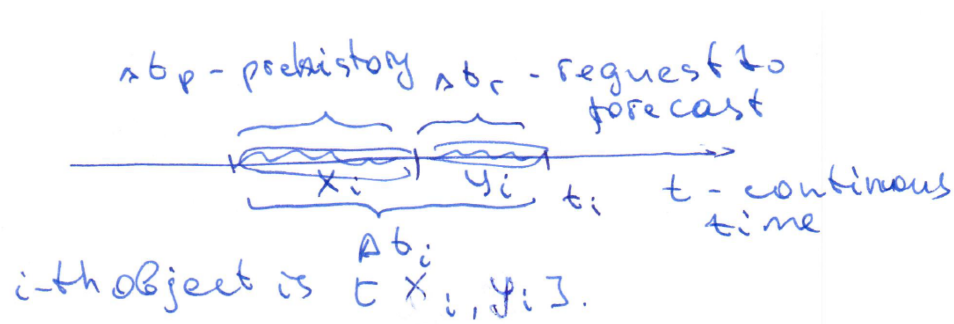

Figure 3: Draw an object from time series history.

2 Forecast analysis

We consider the forecast testing procedure, given by the algorithm 1. Our ultimate goal is to construct a forecasting model be able to obtain forecasts at any given time t **??**. To construct such model we have to imitate this setting, using the so-called retrospective forecast 1 (rolling forecasts, walking ahead predictions). Here we conceal most recent historical samples and make predictions as if they were unknown. Then the quality of

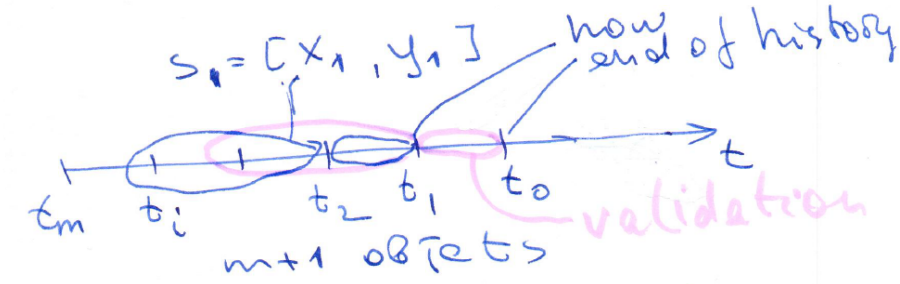

Figure 4: Design matrix generation.

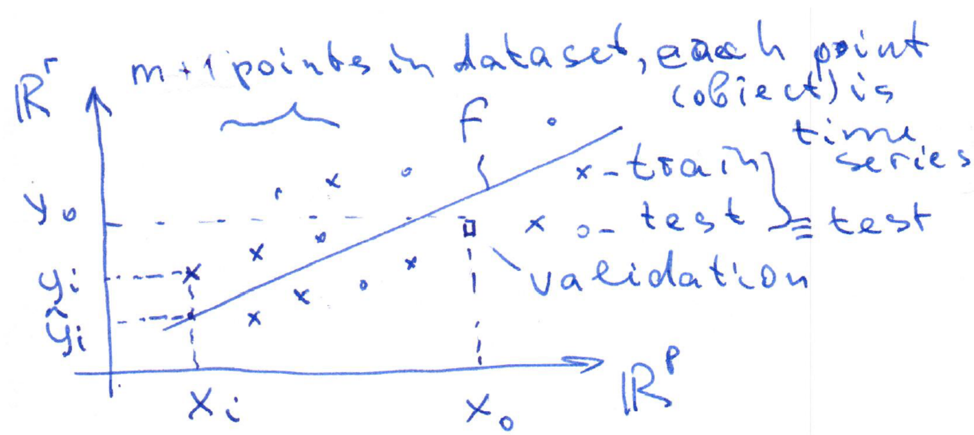

Figure 5: Forecasting as regression problem.

the model is evaluated according to its performance on these concealed samples. In this paper we convert the forecasting problem to regression problem, considering the segments of time series as objects of the design matrix $\mathbf{X}^*$ 1. Sliding 1 the "current" time point t according to the chosen strategy (1–5), we obtain the design matrix $\mathbf{X}$ and further we solve the regression problem instead of forecasting problem **??**.

2.1 Ensuring forecast model validity

A valid forecast model must meet the following conditions:

- Mean of residuals equals to zero.
- Residuals are stationary.
- Residuals are not autocorrelated.

If the forecast does not meet any of these conditions, then it can be further improved by simply adding a constant (minus residual mean) to the model, balancing variance or including more lags. Additionally, desirable properties are normality and homoscedasticity

ComputeForecastingErrors()
Data: $\mathbf{X}^* \in \mathbb{R}^{M\times(\Delta t_\text{r}+\Delta t_\text{p})}$. Parameters: sample size m, train to test ratio α.
Result: Forecasting quality: root-mean-squared error.
while $n \leq M - m$: **do**

define, $\mathbf{X}^*_n = [\mathbf{x}^*_n, \dots, \mathbf{x}^*_{m+n-1}]^\mathsf{T}$;
$\mathbf{X}_\text{train},\ \mathbf{X}_\text{test},\ \mathbf{X}_\text{val} = TrainTestSplit(\mathbf{X}^*_n, \alpha)$;
train forecasting model $\mathbf{f}(\mathbf{x}, \hat{\mathbf{w}}_n)$, using $\mathbf{X}_\text{train}$ and $\mathbf{X}_\text{test}$;
obtain vector of residuals $\boldsymbol{\varepsilon} = [\varepsilon_T, \dots, \varepsilon_{T-\Delta t_\text{r}+1}]$ with respect to $\mathbf{X}_\text{val}$;
compute forecasting quality:

$$\text{RMSE}(n) = \sqrt{\frac{1}{\Delta t_\text{r}} \sum_{t=0}^{\Delta t_\text{r}} \varepsilon^2_{T-t}};$$

$n = n + 1$;

end
Average RMSE by data splits.

TrainTestSplit()
Data: Object-feature matrix $\mathbf{X}^* \in \mathbb{R}^{m\times(\Delta t_\text{r}+\Delta t_\text{p})}$. Train to test ratio α.
Result: Train, test, validation matrices $\mathbf{X}^*_\text{train}$, $\mathbf{X}^*_\text{test}$, $\mathbf{X}^*_\text{val}$.
Set train set and test set sizes:
$m_\text{train} = \lfloor \alpha(m-1) \rfloor$;
$m_\text{test} = m - 1 - m_\text{train}$;
Decompose matrix $\mathbf{X}^*$ into train, test, validation matrices $\mathbf{X}^*_\text{train}$, $\mathbf{X}^*_\text{test}$, $\mathbf{X}^*_\text{val}$:

$$\mathbf{X}^*_\text{train} = \left[\begin{array}{c} \mathbf{x}^*_\text{val} \in \mathbb{R}^{1\times(\Delta t_\text{r}+\Delta t_\text{p})} \\ \hline \mathbf{X}^*_{m_\text{test}} \in \mathbb{R}^{m_\text{test}\times(\Delta t_\text{r}+\Delta t_\text{p})} \\ \mathbf{X}^*_{m_\text{train}} \in \mathbb{R}^{m_\text{train}\times(\Delta t_\text{r}+\Delta t_\text{p})} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{y}_\text{val} & \mathbf{x}_\text{val} \\ \hline \mathbf{Y}_{m_\text{test}} & \mathbf{X}_{m_\text{test}} \\ \mathbf{Y}_{m_\text{train}} & \mathbf{X}_{m_\text{train}} \end{array}\right]$$

Algorithm 1: Train-test split.

of residuals. These properties are not necessary for an adequate model, but allow to obtain theoretical estimations of the confidence interval.

2.2 Forecasting errors

Hyndman [1] divides forecasting errors into four types:

- scale-dependent metrics, such as mean absolute error

$$MAE = \frac{1}{r} \sum_{i=1}^{r} |\varepsilon_i|,$$

- percentage-error metrics such as the mean absolute percent error

$$MAPE = \frac{1}{r}\sum_{i=1}^{r}\frac{|\varepsilon_i|}{|y_0(i)|},$$

  or symmetric MAPE

$$sMAPE = \frac{1}{r}\sum_{i=1}^{r}\frac{2|\varepsilon_i|}{|\hat{y}_0(i) + y_0(i)|},$$

- relative-error metrics, measure the average ratio of the errors from a designed method to the errors $\boldsymbol{\varepsilon}^*$ of a benchmark method

$$MRAE = \frac{1}{r}\sum_{i=1}^{r}\frac{|\varepsilon_i|}{\varepsilon_i^*},$$

- and scale-free error metrics, which express each error as a ratio to an average error from a baseline method:

$$MASE = \frac{n-1}{r}\frac{\sum_{i=1}^{r}|\varepsilon_i|}{\sum_{j=2}^{n}|x_0(j) - x_0(j-1)|}.$$

References

[1] Rob J. Hyndman. Another look at forecast-accuracy metrics for intermittent demand. *Foresight*, (4):43–46, June 2006.