1 Forecast analysis

1.1 Residuals

Suppose that a forecasting model $\mathbf{f}$ was built, producing forecasts

$$\hat{\mathbf{y}}_i = [\hat{s}(t_i), \dots, \hat{s}(t_i - \Delta t_\mathrm{r})] = \mathbf{f}(\mathbf{x}_i, \mathbf{w}).$$

In traditional step-by-step forecasting scheme $(k+1)$-th component of $\mathbf{y}_i$ depends on the forecasts of the previous k components.

$$\hat{s}(t_i - k) = f(\hat{\mathbf{x}}_i^{(k)}, \mathbf{w}), \quad k = 0, \dots, \Delta t_\mathrm{r} - 1,$$

where $\hat{\mathbf{x}}_i^{(k)}$ includes forecasts of k components of $\mathbf{y}_i$:

$$\hat{\mathbf{x}}_i^{(0)} = \mathbf{x}_i = [s(t_i - \Delta t_\mathrm{r} - 1), \dots, s(t_i - \Delta t_\mathrm{r} - \Delta t_\mathrm{p})],$$

$$\hat{\mathbf{x}}^{(1)} = [\hat{s}(t_i - \Delta t_\mathrm{r}), s(t_i - \Delta t_\mathrm{r} - 1), \dots, s(t_i - \Delta t_\mathrm{r} - \Delta t_\mathrm{p} + 1)],$$

$$\dots$$

$$\hat{\mathbf{x}}_i^{(k)} = [\underbrace{\hat{s}(t_i - \Delta t_\mathrm{r} + k - 1), \dots, \hat{s}(t_i - \Delta t_\mathrm{r})}_{k}, s(t_i - \Delta t_\mathrm{r} - 1), \dots, s(t_i - \Delta t_\mathrm{r} - \Delta t_\mathrm{p} + 1)],$$

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

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

Let $\mathcal{B}$ and $\mathcal{B}_0$ denote the train set and the test set of indices i. Since 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

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

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

Option 1. In our forecasting scheme all k components are obtained at one step, due to the way the matrix $\mathbf{X}^*$ is designed. Furthermore, to get rid of logical dependencies between components of subsequent vectors $\mathbf{y}_i,\ \mathbf{y}_{i+1}$ we consider the following testing procedure. The available history of the time series $\mathbf{s}$ is split into several training blocks $\mathcal{B}_k$, $|\mathcal{B}| = (\Delta t_\mathrm{r} - 1)\Delta t_\mathrm{p} + \Delta t_\mathrm{r}$, separated from each other by Δt_p time points, which comprise target objects $(\mathbf{x}_k, \mathbf{y}_k)$ for the test set set. The complete procedure is given be the algorithm 1:

1.2 Ensuring forecast model validity

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

1.3 Forecasting quality

Data: $\mathcal{D} = \{\mathbf{s}, \mathbf{s}', \dots\},\ \mathbf{s} = \{s(T), \dots, s(1)\}$. Parameters: number of testing procedures N, local history size Δt_{p}, requested forecast length Δt_{r}.
Result: Train-test matrix.
Result: Forecasting quality: root-mean-squared error.
$n = 1$;
while $n < N$: **do**

$\mathbf{X}_{\text{train}}, \mathbf{X}_{\text{test}} = \mathit{TrainTestSplit}(\{\mathbf{s}, \mathbf{s}', \dots\}, \Delta t_{\text{p}}, \Delta t_{\text{r}})$;
train forecasting model $\mathbf{f}(\mathbf{x}, \hat{\mathbf{w}})$, using $\mathbf{X}_{\text{train}}$;
obtain vector of residuals $\boldsymbol{\varepsilon} = [\varepsilon_T, \dots, \varepsilon_{T-K\Delta t-\Delta t_{\text{r}}}]$ with respect to $\mathbf{X}_{\text{test}}$:

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \mathbf{y}_0 - \mathbf{f}(\mathbf{x}_0, \hat{\mathbf{w}}) \\ \dots \\ \mathbf{y}_K - \mathbf{f}(\mathbf{x}_K, \hat{\mathbf{w}}) \end{bmatrix};$$

compute forecasting quality:

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

$n = n + 1$;

end
Average RMSE by data splits.

TrainTestSplit()
Data: $\mathcal{D} = \{\mathbf{s}, \mathbf{s}', \dots\},\ \mathbf{s} = \{s(T), \dots, s(1)\}$. Parameters: local history size Δt_{p}, requested forecast length Δt_{r}. (For simplicity, let $T = M\Delta t_{\text{p}} + \Delta t_{\text{r}}$.)
Result: Train-test matrices $\mathbf{X}^*_{\text{train}}$, $\mathbf{X}^*_{\text{test}}$.
Denote $\Delta T = (\Delta t_{\text{r}} \Delta t_{\text{p}} - \Delta t_{\text{p}} + \Delta t_{\text{r}})$;
$k = 0,\ K = \lfloor M/(\Delta t_{\text{r}} - 1) \rfloor - 1$;
while $k < K$: **do**

add $T - k\Delta T$ to $\mathcal{B}_0$:
$\mathbf{y}_k = [s(T - k\Delta T), \dots, s(T - k\Delta T)]$;
$\mathbf{x}_k = [s(T - k\Delta T - \Delta t_{\text{r}} - 1), \dots, s(T - k\Delta T - \Delta t_{\text{r}} - \Delta t_{\text{p}})]$;
from $\mathcal{B}_k = \{T - k\Delta T - 1, \dots, T - (k+1)\Delta T\}$ make matrix $\mathbf{X}^*_k$;
k = k + 1;

end
Join matrices $\mathbf{X}^*_k$ vertically into a training matrix $\mathbf{X}^*_{\text{train}}$;
Join $(\mathbf{x}_k, \mathbf{y}_k)$, $k = 0, K-1$ into $\mathbf{X}^*_{\text{test}}$.

Algorithm 1: Train-test split.