1 Feature generation procedures

Denote the generated feature vector as $\boldsymbol{\phi}$. This vector consists of concatenated row-vectors $\boldsymbol{\phi} = [\boldsymbol{\phi}', \boldsymbol{\phi}'', \dots]$, which corresponds to time series local histories $\mathbf{s} = [\mathbf{s}', \mathbf{s}'', \dots]$, modified with set of transformations $\mathfrak{G}$. The elements $g : \mathbf{s} \to \boldsymbol{\phi}$ of this set are listed below.

1.1 Transformations of local history

The table 1 lists the time series transformation functions. There are non-parametric and parametric procedures to generate features. For the parametric functions $g = g(\mathbf{b}, s)$ the default values of the parameters $\mathbf{b}$ are assigned empirically.

The parametric procedure request two optimization problem statements of the model parameters $\mathbf{w}$ and the primitive function parameters $\mathbf{b}$. The first one fixes the vector $\hat{\mathbf{b}}$, collected over all the primitive functions $\{g\}$, which generate features $\boldsymbol{\phi}$:

$$\hat{\mathbf{w}} = \arg\min S\big(\mathbf{w}|\mathbf{f}(\mathbf{w}, \mathbf{x}), \mathbf{y}\big), \quad \text{where} \quad [\mathbf{y}, \mathbf{x}] = \boldsymbol{\phi}(\hat{\mathbf{b}}, \mathbf{s}).$$

The second one optimizes the transformation parameters $\hat{\mathbf{b}}$ given obtained model parameters $\mathbf{w}$

$$\hat{\mathbf{b}} = \arg\min S\big(\mathbf{b}|\mathbf{f}(\hat{\mathbf{w}}, \mathbf{x}), \mathbf{y}\big).$$

This procedure repeats two problems until vectors $\hat{\mathbf{w}}, \hat{\mathbf{b}}$ converge. The initial values of vector $\mathbf{b}$ (are shown in table 1). Due to the various origins of the time series and their transformations the residual vector should be normalized:

$$\boldsymbol{\varepsilon}' = \frac{\hat{\mathbf{y}}' - \mathbf{y}'}{|\mathbf{y}'| \cdot \|\mathbf{y}'\|_2^{\frac{1}{2}}}.$$

It does not change the number elements in the vectors, $|\boldsymbol{\phi}'| = |\mathbf{s}'|$.

1.2 Convolutions, statistics and parameters of local history

The listed feature generation functions convolves time series, so they reduce the dimentionality $|\boldsymbol{\phi}' = \mathbf{g}(\mathbf{s}')| < |\mathbf{s}'|$.

1.3 Parameters of local history forecast

For the time series $\mathbf{s}'$ construct the Hankel matrix with a period k and shift p, so that for $\mathbf{s} = [s_1, \dots, s_T]$ the matrix

$$\mathbf{H}^* = \left[\begin{array}{c|cc} s_T & \dots & s_{T-k+1} \\ \vdots & \ddots & \vdots \\ s_{k+p} & \dots & s_{1+p} \\ s_k & \dots & s_1 \end{array}\right], \text{ where } 1 \geqslant p \geqslant k.$$

Reconstruct the regression to the first column of the matrix $\mathbf{H}^* = [\mathbf{h}, \mathbf{H}]$ and denote its least square parameters as the feature vector

$$\boldsymbol{\phi}' = \arg\min \|\mathbf{h} - \mathbf{H}\boldsymbol{\phi}\|_2^2.$$

For the time series $[\mathbf{s}_i', \mathbf{s}_i'', \dots]$ use the parameters $[\boldsymbol{\phi}_i', \boldsymbol{\phi}_i'']$ as the features.

1.4 Distances to centroids of local clusters

This procedure applies the kernel trick to the time series. For given local history time series $\mathbf{s}_1', \dots, \mathbf{s}_m'$ compute k-means centroids $\mathbf{c}_1', \dots, \mathbf{c}_P'$. With the selected k-means distance function ρ construct the feature vector

$$\boldsymbol{\phi}_i' = [\rho(\mathbf{c}_1', \mathbf{s}_i'), \dots, \rho(\mathbf{c}_P', \mathbf{s}_i')] \in \mathbb{R}_+^P.$$

This k-means of another clustering procedure may use internal parameters, so that there are no parameters to be included to the feature vector or to the forecasting model.

Table 1: List of primitive functions.

Function name	Formula	Output dimension	# of arguments	# of parameters
Plus	$x_1 + x_2$	1	2	0
Minus	$x_1 - x_2$	1	2	0
Product	$x_1 \cdot x_2$	1	2	0
Division	$\frac{x_1}{x_2}$	1	2	0
Quadratic	$w_2x^2 + w_1x + w_0$	1	1	3
Cubic	$w_3x^3 + w_2x^2 + w_1x + w_0$	1	1	4
Linear	$w_1x + w_0$	1	1	2
Add constant	$x + w$	1	1	1
Nonparametric sin	$\sin(x)$	1	1	0
Sin	$\sin(w_0 + w_1x)$	1	1	2
Square root	$\sqrt{x}$	1	1	0
Arctangent	$\arctan x$	1	1	0
Nonparametric log-sigmoid	$1/(1 + \exp(-x))$	1	1	0
Logarithmic	$\ln x$	1	1	0
Logarithmic sigmoid	$1/(w_0 + \exp(-w_1x))$	1	1	2
Exponent	$\exp x$	1	1	0
Normal	$\frac{1}{w_1\sqrt{2\pi}} \exp\left(\frac{(x-w_2)^2}{2w_1^2}\right)$	1	1	2
Hyperbolic tangent	$\tanh(x)$	1	1	0
Hyperbolic tangent sigmoid	$\frac{2}{1-\exp(-wx)} - 1$	1	1	1
Multiply by constant	$x \cdot w$	1	1	1
Monomial	$w_1x^{w_2}$	1	1	2
Weibull-2	$w_1w_2x^{w_2-1} \exp -w_1x^{w_2}$	1	1	2
Weibull-3	$w_1w_2x^{w_2-1} \exp -w_1(x - w_3)^{w_2}$	1	1	3

Table 2: Growth rate of monotone functions.

Function name	Formula	Output dimension	# of arguments	# of parameters	Constraints
Exponential rate	$\exp(w_1 x + w_0)$	1	1	2	$w_1 > 0$
Polynomial rate	$\exp(w_1 \ln x + w_0)$	1	1	2	$w_1 > 1$
Sublinear polynomial rate	$\exp(w_1 \ln x + w_0)$	1	1	2	$0 < w_1 < 1$
Logarithmic rate	$w_1 \ln x + w_0$	1	1	2	$w_1 > 0$
Slow convergence	$w_0 + w_1/x$	1	1	2	$w_1 \neq 0$
Fast convergence	$w_0 + w_1 \cdot \exp(-x)$	1	1	2	$w_1 \neq 0$
Sigmoid	$1/(w_0 + \exp(-w_1 x))$	1	1	2	$w_1 > 0$

Table 3: Variable number of parameters.

Quasi-linear	$\sum_{j=1}^{n} a_j g_j(x)$	1	1	n
Sum of Gaussians	$\sum_{j=1}^{n} a_j \exp(-\frac{(x-b_j)^2}{c_j})$	1	1	$3n$
Polynomial	$\sum_{j=0}^{n} a_j x^j$	1	1	n
Rational polynomial	$\frac{\sum_{j=0}^{n} a_j x^j}{x^m + \sum_{j=0}^{m-1} b_j x^j}$	1	1	$n+m+1$