Random Forest regression: Random Forest is an ensemble of decision trees trained on bootstrapped samples with random feature selection at each split. For time-series regression in a sliding-window setup, each row of the input matrix represents recent lags, and the target is the next value. The forest reduces variance via bagging and captures nonlinear relationships without strong parametric assumptions. Key hyperparameters include the number of trees (ntree) and terminal node size (nodesize).

Objective: Train and evaluate a Random Forest model for time-series forecasting with sliding windows, including normalization, fitting, and results visualization.

# Time Series Regression - Random Forest  
  
# Installing the package (if needed)  
#install.packages("tspredit")

# Loading the packages  
library(daltoolbox)  
library(tspredit)

# Series for study and sliding windows  
  
data(tsd)  
ts <- ts\_data(tsd$y, 10)  
ts\_head(ts, 3)

## t9 t8 t7 t6 t5 t4 t3 t2 t1  
## [1,] 0.0000000 0.2474040 0.4794255 0.6816388 0.8414710 0.9489846 0.9974950 0.9839859 0.9092974  
## [2,] 0.2474040 0.4794255 0.6816388 0.8414710 0.9489846 0.9974950 0.9839859 0.9092974 0.7780732  
## [3,] 0.4794255 0.6816388 0.8414710 0.9489846 0.9974950 0.9839859 0.9092974 0.7780732 0.5984721  
## t0  
## [1,] 0.7780732  
## [2,] 0.5984721  
## [3,] 0.3816610

# Series visualization  
library(ggplot2)  
plot\_ts(x=tsd$x, y=tsd$y) + theme(text = element\_text(size=16))



# Train-test split and projection (X, y)  
  
samp <- ts\_sample(ts, test\_size = 5)  
io\_train <- ts\_projection(samp$train)  
io\_test <- ts\_projection(samp$test)

# Preprocessing (global min-max normalization)  
  
preproc <- ts\_norm\_gminmax()

# Training the Random Forest model  
  
model <- ts\_rf(ts\_norm\_gminmax(), input\_size=8, nodesize=1, ntree=20)  
model <- fit(model, x=io\_train$input, y=io\_train$output)

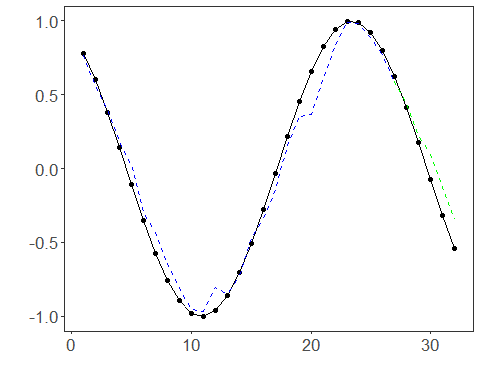
# Fit evaluation (train)  
  
adjust <- predict(model, io\_train$input)  
adjust <- as.vector(adjust)  
output <- as.vector(io\_train$output)  
ev\_adjust <- evaluate(model, output, adjust)  
ev\_adjust$mse

## [1] 0.01000336

# Forecast on test set (5 steps ahead)  
  
prediction <- predict(model, x=io\_test$input[1,], steps\_ahead=5)  
prediction <- as.vector(prediction)  
output <- as.vector(io\_test$output)  
ev\_test <- evaluate(model, output, prediction)  
ev\_test

## $values  
## [1] 0.41211849 0.17388949 -0.07515112 -0.31951919 -0.54402111  
##   
## $prediction  
## [1] 0.4373025 0.2236842 0.1007814 -0.1205018 -0.3432312  
##   
## $smape  
## [1] 0.7333964  
##   
## $mse  
## [1] 0.02279811  
##   
## $R2  
## [1] 0.8030916  
##   
## $metrics  
## mse smape R2  
## 1 0.02279811 0.7333964 0.8030916

# Plot comparing actual vs fit (train) and forecast (test)  
  
yvalues <- c(io\_train$output, io\_test$output)  
plot\_ts\_pred(y=yvalues, yadj=adjust, ypre=prediction) + theme(text = element\_text(size=16))



References - L. Breiman (2001). Random Forests. Machine Learning, 45(1), 5–32.