PCA-based regression anomaly detection: Projects multivariate observations onto principal components and reconstructs data; large reconstruction errors indicate anomalies. Data are standardized, PCA is fitted, and reconstruction errors are thresholded.

The PCA-based detector flags large reconstruction errors when projecting multivariate data onto principal components. In this tutorial we:

* Load a multivariate example and create synthetic event labels
* Fit the PCA detector on two dimensions
* Visualize detections over different series and inspect residual magnitudes

# Install Harbinger (if needed)  
#install.packages("harbinger")

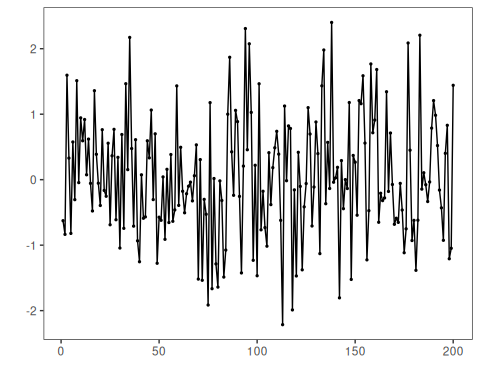
# Load required packages  
library(daltoolbox)  
library(harbinger)   
library(ggplot2)

# Load a multivariate example and define event labels (for demo)  
data("examples\_harbinger")  
dataset <- examples\_harbinger$multidimensional  
dataset$event <- FALSE  
dataset$event[c(101,128,167)] <- TRUE

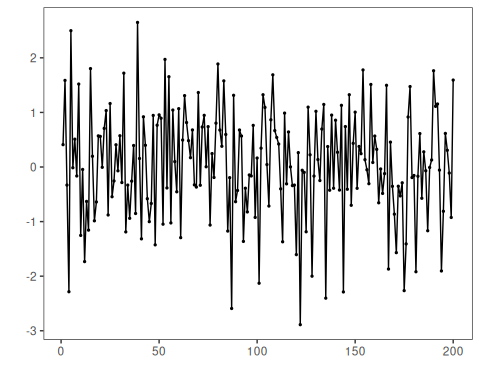
head(dataset)

## serie x event  
## 1 -0.6264538 0.40940184 FALSE  
## 2 -0.8356286 1.58658843 FALSE  
## 3 1.5952808 -0.33090780 FALSE  
## 4 0.3295078 -2.28523554 FALSE  
## 5 -0.8204684 2.49766159 FALSE  
## 6 0.5757814 -0.01339952 FALSE

# Plot the target series  
har\_plot(harbinger(), dataset$serie)



# Plot the second dimension  
har\_plot(harbinger(), dataset$x)

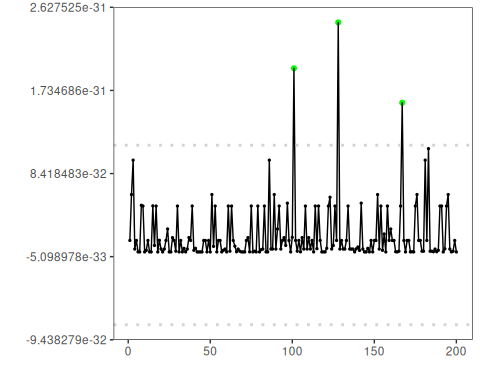


# Fit the PCA detector on the first two columns and run detection  
model <- fit(hmu\_pca(), dataset[,1:2])  
detection <- detect(model, dataset[,1:2])

# Plot detections on the target series  
grf <- har\_plot(model, dataset$serie, detection, dataset$event)  
grf <- grf + ylab("serie")

# Plot detections on the second dimension  
grf <- har\_plot(model, dataset$x, detection, dataset$event)  
grf <- grf + ylab("x")

# Plot residual magnitude and decision thresholds  
har\_plot(model, attr(detection, "res"), detection, dataset$event, yline = attr(detection, "threshold"))



References - Jolliffe, I. T. (2002). Principal Component Analysis. Springer.