## Denoising Autoencoder (encode-decode)

Inputs are stochastically corrupted during training, and the model learns to reconstruct clean windows. The bottleneck captures noise-invariant structure, and reconstruction metrics quantify how well essential patterns are preserved.

This example demonstrates how to use a denoising autoencoder to encode and reconstruct time-series windows, enabling evaluation of reconstruction quality under noise.

Prerequisites - Python with PyTorch accessible via reticulate - R packages: daltoolbox, tspredit, daltoolboxdp, ggplot2

Quick notes - Noise is applied to the input during training; at inference, the reconstruction tends to be smoother. - Per-column metrics (R2 and MAPE) help assess robustness per step.

# Denoising Autoencoder transformation (encode-decode)  
  
# Considering a dataset with $p$ numerical attributes.   
  
# The goal of the autoencoder is to reduce the dimension of $p$ to $k$, such that these $k$ attributes are enough to recompose the original $p$ attributes. However, from the $k$ dimensions the data is returned back to $p$ dimensions. The higher the autoencoder quality, the more similar the output is to the input.   
  
# Installing packages  
#install.packages("tspredit")  
#install.packages("daltoolboxdp")

# Loading packages  
library(daltoolbox)  
library(tspredit)  
library(daltoolboxdp)  
library(ggplot2)

# Example dataset (series -> windows)   
data(tsd)  
  
sw\_size <- 5  
ts <- ts\_data(tsd$y, sw\_size)  
  
ts\_head(ts)

## t4 t3 t2 t1 t0  
## [1,] 0.0000000 0.2474040 0.4794255 0.6816388 0.8414710  
## [2,] 0.2474040 0.4794255 0.6816388 0.8414710 0.9489846  
## [3,] 0.4794255 0.6816388 0.8414710 0.9489846 0.9974950  
## [4,] 0.6816388 0.8414710 0.9489846 0.9974950 0.9839859  
## [5,] 0.8414710 0.9489846 0.9974950 0.9839859 0.9092974  
## [6,] 0.9489846 0.9974950 0.9839859 0.9092974 0.7780732

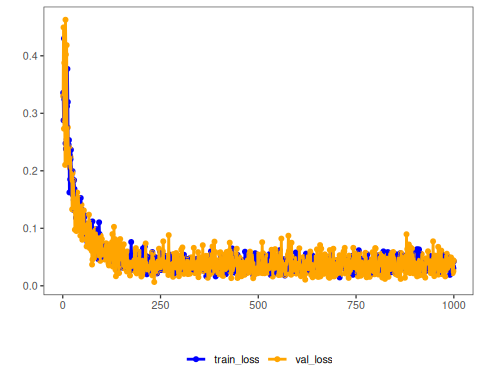
# Normalization (min-max by group)  
preproc <- ts\_norm\_gminmax()  
preproc <- fit(preproc, ts)  
ts <- transform(preproc, ts)  
  
ts\_head(ts)

## t4 t3 t2 t1 t0  
## [1,] 0.5004502 0.6243512 0.7405486 0.8418178 0.9218625  
## [2,] 0.6243512 0.7405486 0.8418178 0.9218625 0.9757058  
## [3,] 0.7405486 0.8418178 0.9218625 0.9757058 1.0000000  
## [4,] 0.8418178 0.9218625 0.9757058 1.0000000 0.9932346  
## [5,] 0.9218625 0.9757058 1.0000000 0.9932346 0.9558303  
## [6,] 0.9757058 1.0000000 0.9932346 0.9558303 0.8901126

# Train/test split  
samp <- ts\_sample(ts, test\_size = 10)  
train <- as.data.frame(samp$train)  
test <- as.data.frame(samp$test)

# Training autoencoder (reduce 5 -> 3)  
auto <- autoenc\_denoise\_ed(5, 3)  
auto <- fit(auto, train)

fit\_loss <- data.frame(x=1:length(auto$train\_loss), train\_loss=auto$train\_loss,val\_loss=auto$val\_loss)  
  
grf <- plot\_series(fit\_loss, colors=c('Blue','Orange'))  
plot(grf)



# Testing the autoencoder  
# Show test samples and display reconstruction  
print(head(test))

## t4 t3 t2 t1 t0  
## 1 0.7258342 0.8294719 0.9126527 0.9702046 0.9985496  
## 2 0.8294719 0.9126527 0.9702046 0.9985496 0.9959251  
## 3 0.9126527 0.9702046 0.9985496 0.9959251 0.9624944  
## 4 0.9702046 0.9985496 0.9959251 0.9624944 0.9003360  
## 5 0.9985496 0.9959251 0.9624944 0.9003360 0.8133146  
## 6 0.9959251 0.9624944 0.9003360 0.8133146 0.7068409

result <- transform(auto, test)  
print(head(result))

## [,1] [,2] [,3] [,4] [,5]  
## [1,] 0.6947847 0.8615562 0.9326051 0.9514428 1.0000805  
## [2,] 0.8000759 0.9425692 0.9938369 0.9771729 1.0066901  
## [3,] 0.8871007 0.9984631 1.0295606 0.9688694 0.9756524  
## [4,] 0.9551534 1.0276260 1.0374366 0.9286628 0.9070619  
## [5,] 0.9861873 1.0247359 1.0047252 0.8576791 0.8191626  
## [6,] 0.9831920 0.9905756 0.9398021 0.7665498 0.7109544

# Reconstruction metrics per column: R2 and MAPE  
result <- as.data.frame(result)  
names(result) <- names(test)  
r2 <- c()  
mape <- c()  
for (col in names(test)){  
 r2\_col <- cor(test[col], result[col])^2  
 r2 <- append(r2, r2\_col)  
 mape\_col <- mean((abs((result[col] - test[col]))/test[col])[[col]])  
 mape <- append(mape, mape\_col)  
 print(paste(col, 'R2 test:', r2\_col, 'MAPE:', mape\_col))  
}

## [1] "t4 R2 test: 0.997345746840775 MAPE: 0.0248847228954877"  
## [1] "t3 R2 test: 0.999171817922396 MAPE: 0.0272133506388344"  
## [1] "t2 R2 test: 0.997479769746117 MAPE: 0.0412884059573127"  
## [1] "t1 R2 test: 0.998726267305548 MAPE: 0.0494961708050386"  
## [1] "t0 R2 test: 0.999711212156747 MAPE: 0.0124775902404824"

print(paste('Means R2 test:', mean(r2), 'MAPE:', mean(mape)))

## [1] "Means R2 test: 0.998486962794317 MAPE: 0.0310720481074312"

# Note: the noise level impacts reconstruction capacity.

References - Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008). Extracting and composing robust features with denoising autoencoders. ICML.