Anomaly detection in time series data using autoencoders

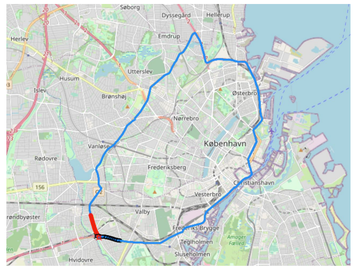
Sebastian Tønne Kotewitz, s174270

Introduction

The goal:
Using a sequence of Z acceleration and a sequence of speed, predict if the mean International Roughness Index value of a given stretch is below or above 4.5.

Data collected from Green Mobility vehicles driving a specific route of copenhagen

The speed sequences are significantly shorter than the acceleration sequences



I will attempt to reach the goal by fitting an autoencoder on the smooth roads, then classifying the two types of road based on reconstruction error

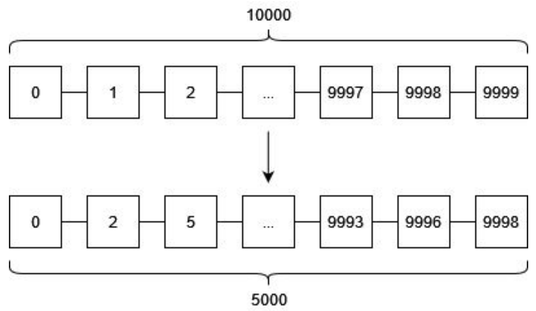
Preparing the data

Data fit between -1 and 1, based on min/max of training sequences

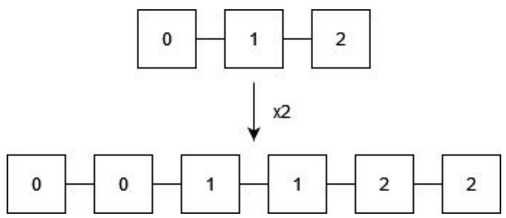
Acceleration sequences are downsampled if their length is above 90th percentile

Speed is upsampled, then downsampled to match the length of their respective acceleration sequences

Downsampling



Upsampling



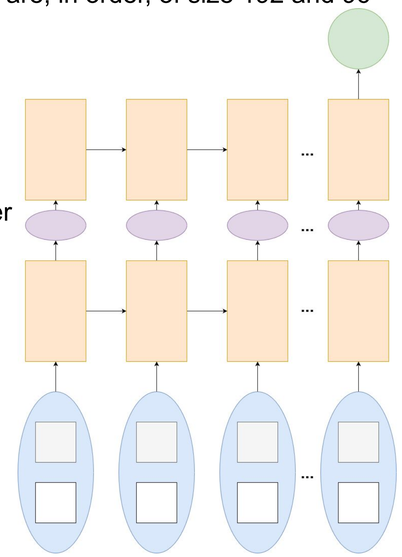
The encoder

The encoder takes the two sequences (now of equal length) and runs them through two RNN layers

The RNN layers use LSTM cells to produce a compressed representation of the sequence data

The hidden layers are, in order, of size 192 and 96

The encoded representation of the acceleration sequence is the final hidden state of the second layer



The decoder

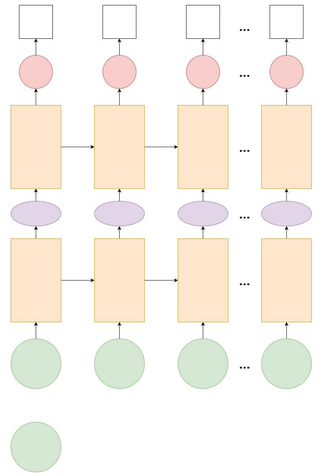
The decoder takes the encoded representation and repeats it to match the sequence length. This is run through two RNN layers using LSTM cells. Finally, the hidden states of the last RNN layer are run through a dense layer with no activation function.

The decoder attempts to have the output of the final dense layer match the original acceleration sequence as close as possible

The hidden RNN layer sizes are, in order, of size 96 and 192

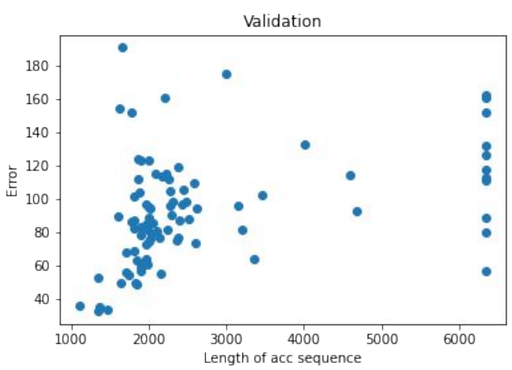
The dense layer takes 192 input features and outputs 1

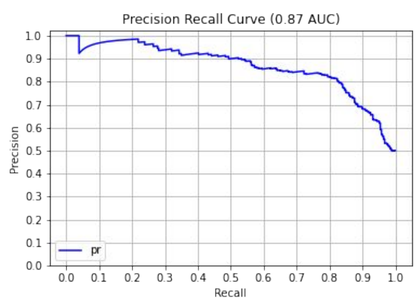
The autoencoder was trained using the Adam optimizer and the L1Loss cost function with reduction=sum



Results

Training size: 2446
Validation size: 306
Test size: 306 conforming + 306 anomalies





F1 score = 0.82
Error threshold = 114.99