A Multiplication Layer for Sequence Data

◎Joonho Lee, Hideaki Hayashi, Seiichi Uchida

Kyushu University

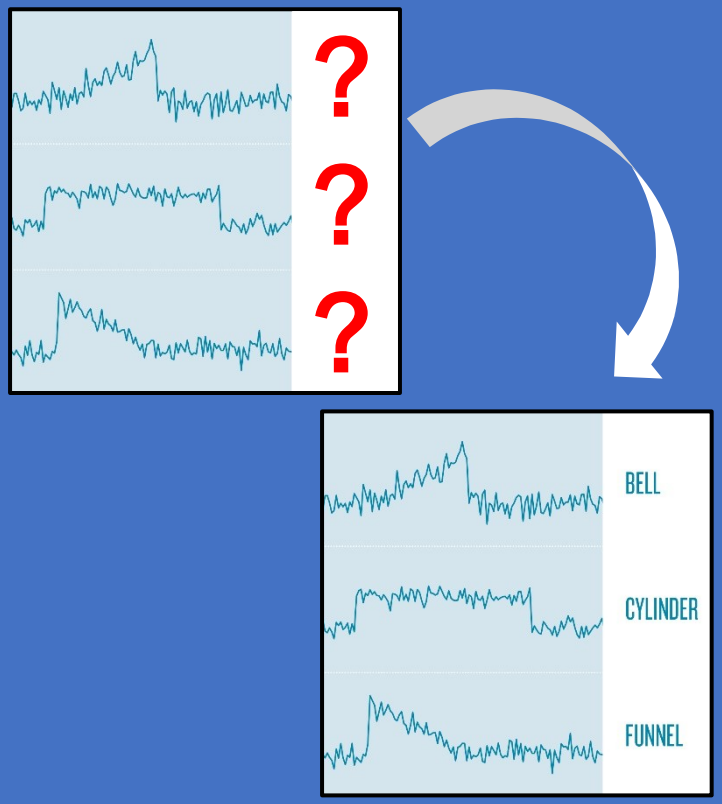
START

P2

What is time series classification ?

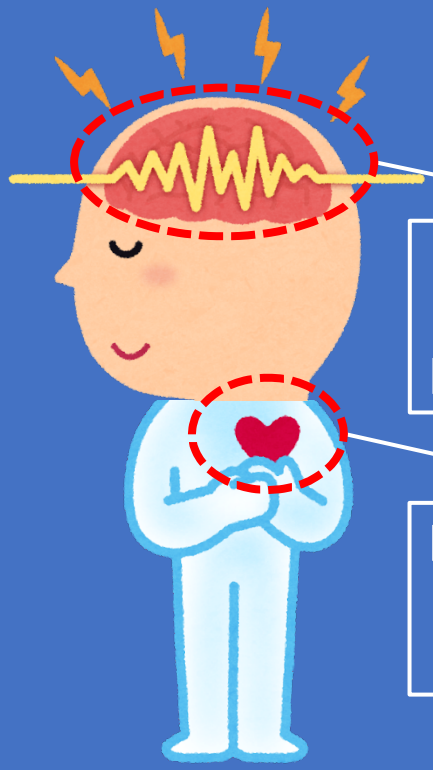
Assigning labels to each observed time series

Objective



http://www.zib.de/features/similarity-search-and-insect-classification

Application



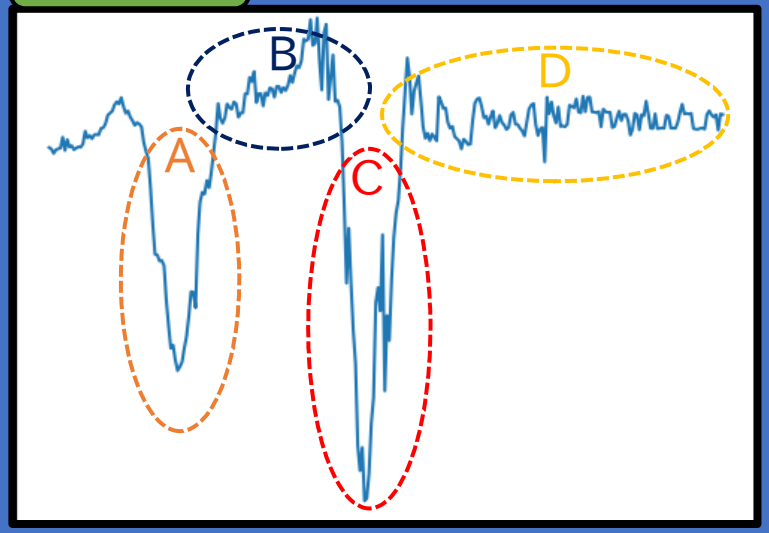
EEG analysis
↓
Emotional analysis

Heart rate analysis
↓
Disease detection

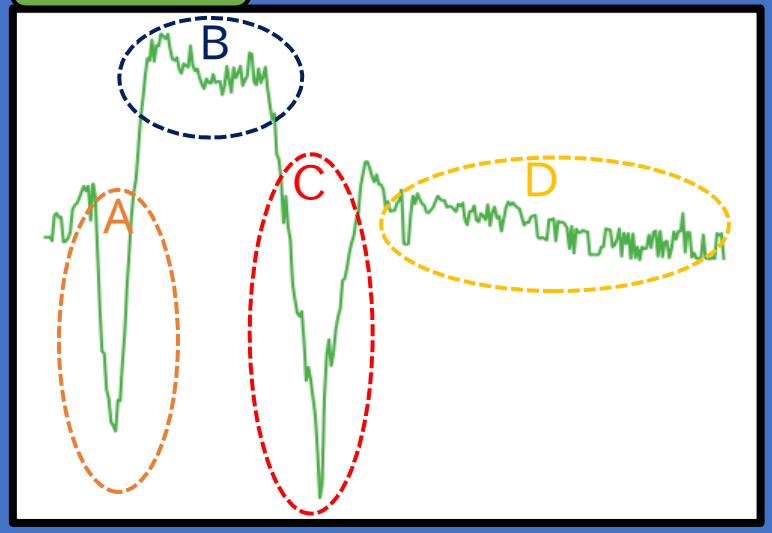
P3

Where is the best feature to classify ?

Class A



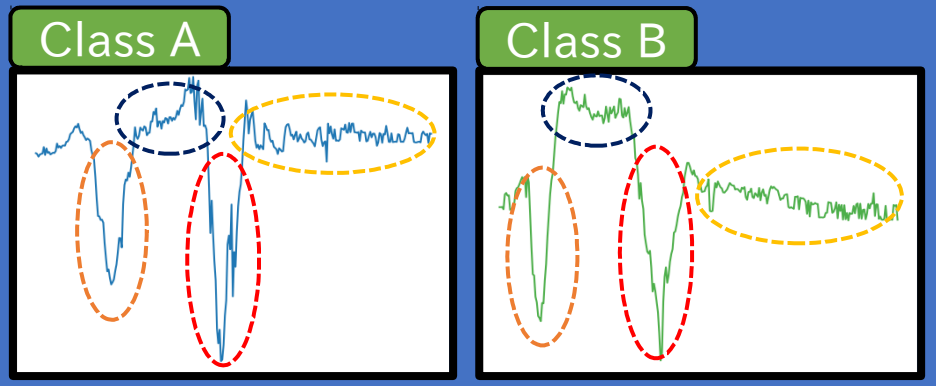
Class B

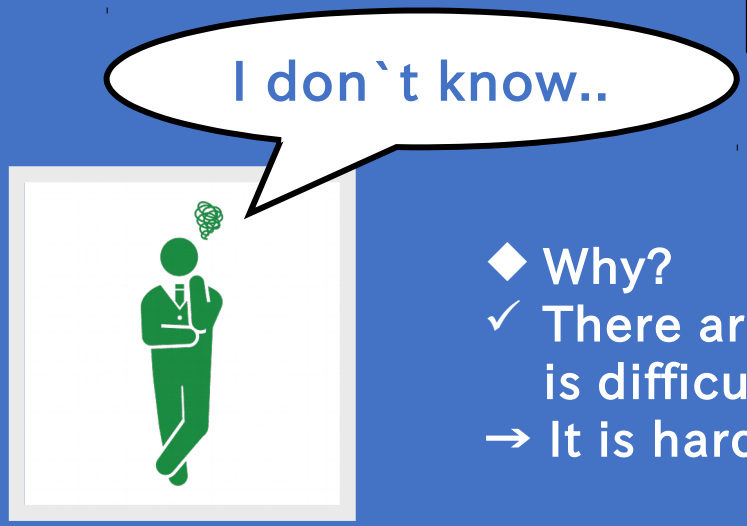


Did you notice these places, right ?

Problem

P4 The answer is ..!





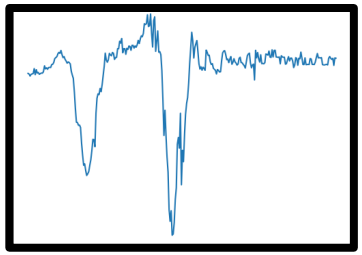
- Why?
  - ✓ There are **multiple** patterns and pattern that is difficult to notice easily.
  - → It is hard to judge where to place weights.

Many studies have been conducted to classify the labels

P5

How to classify ?

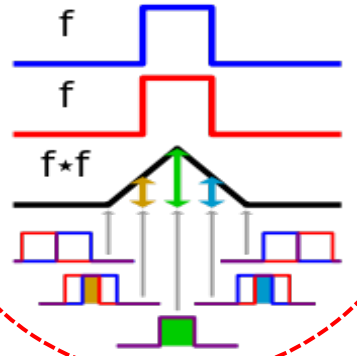
Hand crafted



Input data

Feature extraction

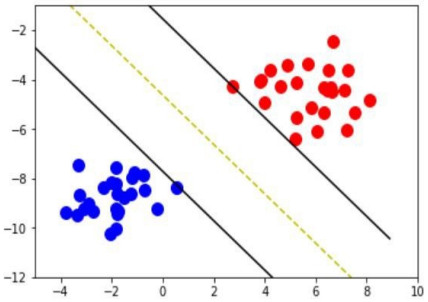
Autocorrelation



www.ibric.org/scicafe/read.php?Board=scicafe000692&id=15891

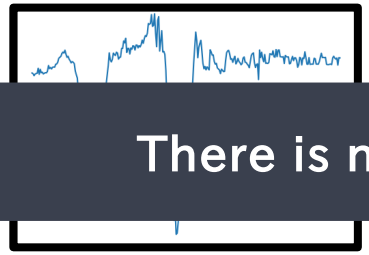
Classifier

SVM



https://medium.com/deep-math-machine-learning-ai/chapter-3

Deep learning

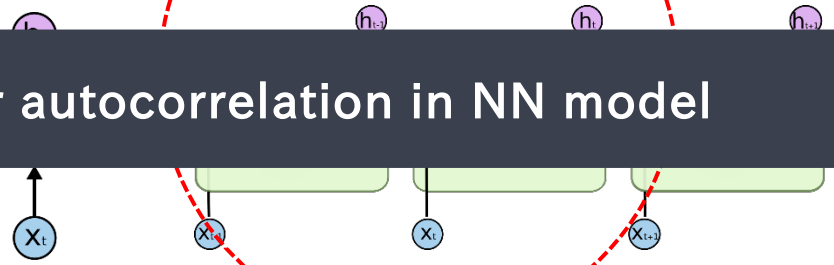


Input data

Neural Networks model

RNN

LSTM

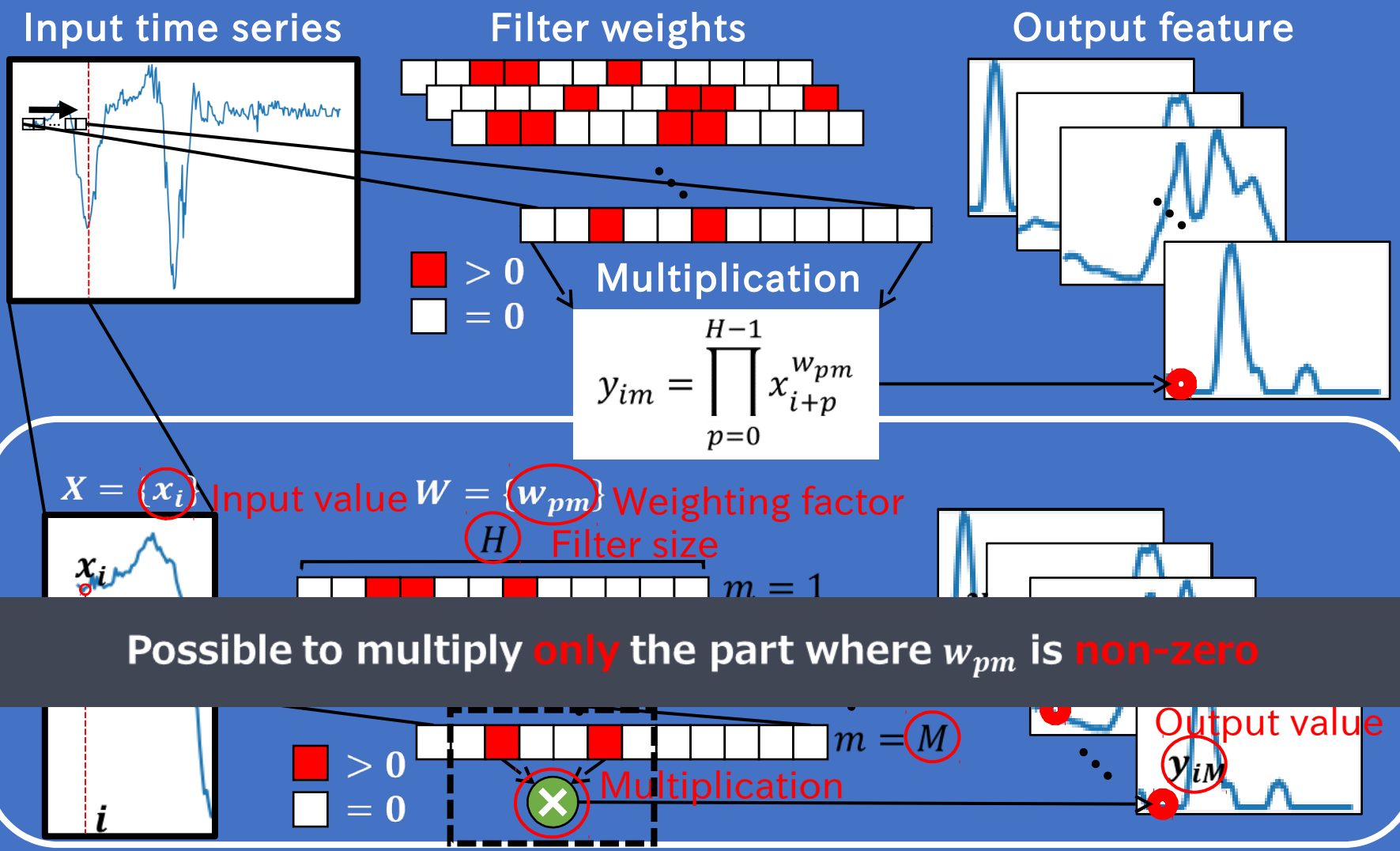


https://qiita.com/KojiOhki/items/89cd7b69a8a6239d67ca

There is no layer for autocorrelation in NN model

P6

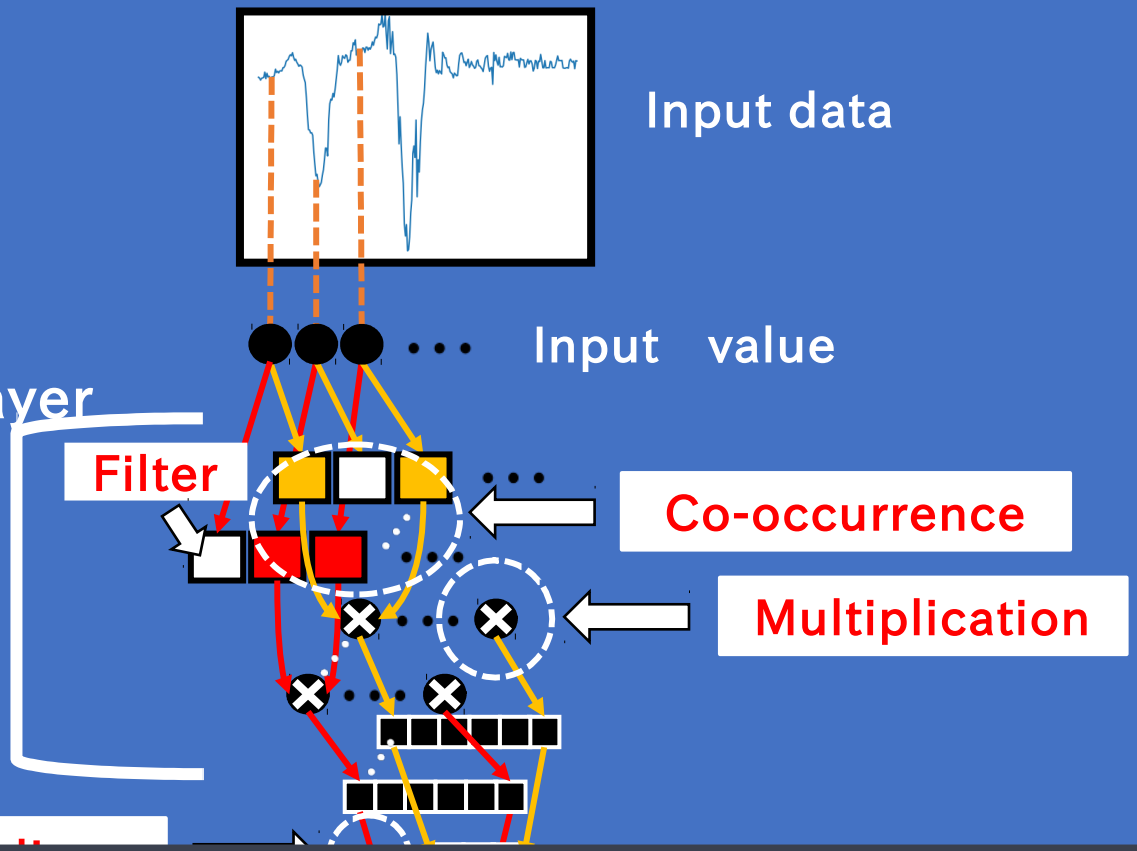
Let me propose multiplication layer !



Our research

P7

How to use ?



Possible to discriminatively extract high-order autocorrelation features

NN model

Our research

high-order autocorrelation | discriminatively

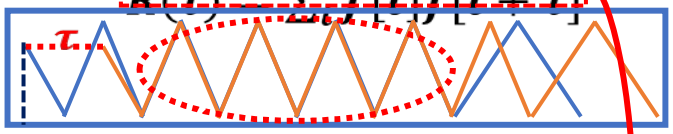
P8

What it means ?

Autocorrelation

$R(\tau) = \sum_t f[t] f[t+\tau]$

Ex)



co-occurrence takes place when pattern is same

High-order autocorrelation

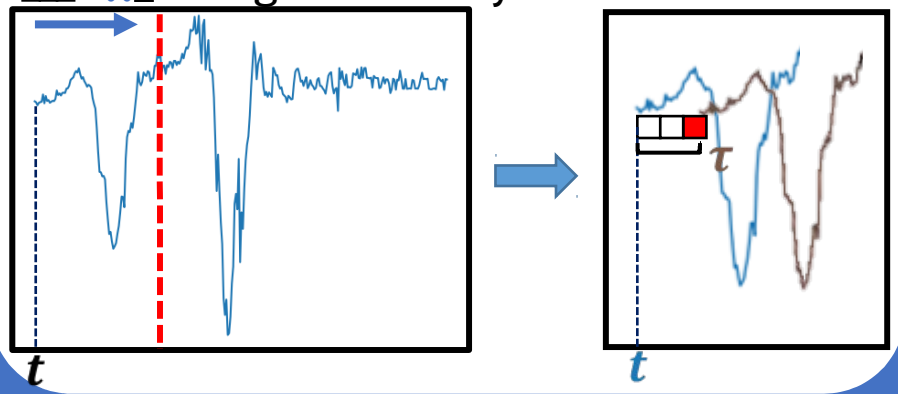
Lth-order autocorrelation

$$R(\tau_1, \tau_2, \ldots, \tau_L) = \sum_D f[t] f[t+\tau_1] f[t+\tau_2] \cdots f[t+\tau_L]$$

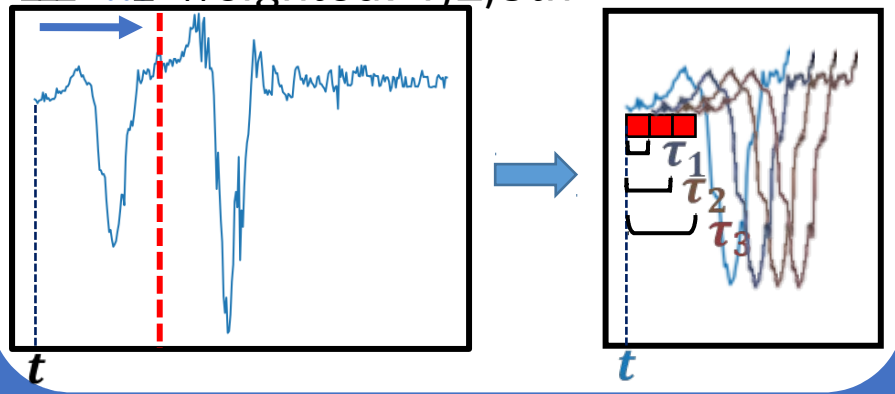
$$\mathbf{D} = \{t | t + \tau_l \in D, \forall l \in \{1, 2, \ldots, L\}\}$$

Key point : we can extract autocorrelation features in any situation.

weighted: Only 3th



weighted: 1,2,3th



Our research

high-order autocorrelation | discriminatively

P9

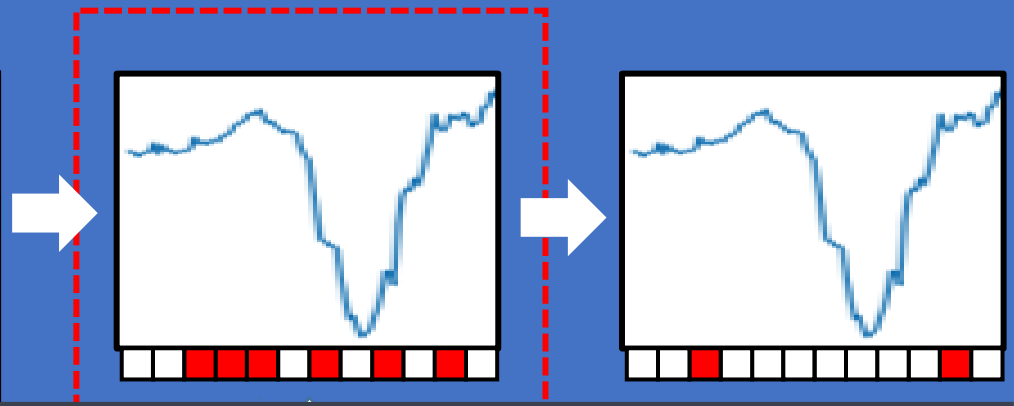
What it means ?

Constraint 1

✓ Let w be zero if possible

Constraint 2

✓ If all w are added, it becomes



Key point : The features can be extracted discriminatively.

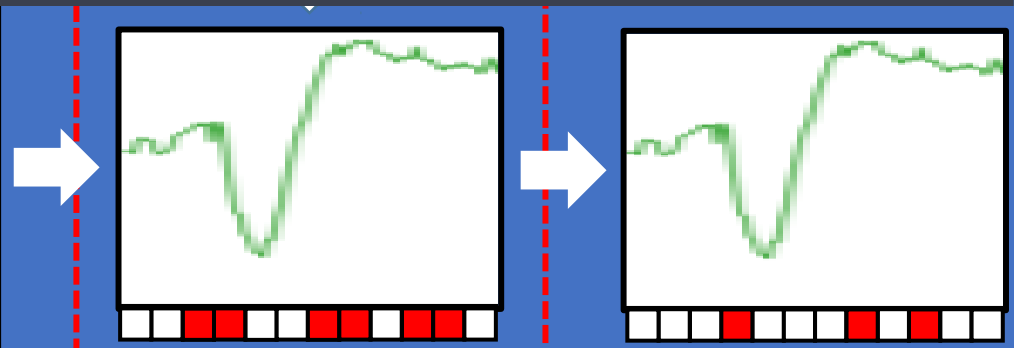
Constraint 3

✓ Let w have values between zero and one

+

Comparison

✓ Compare with each other class extracted features

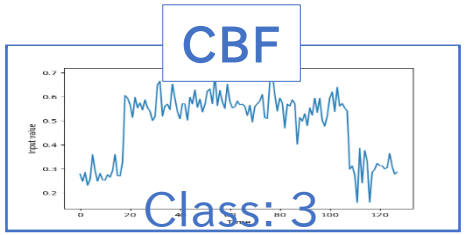


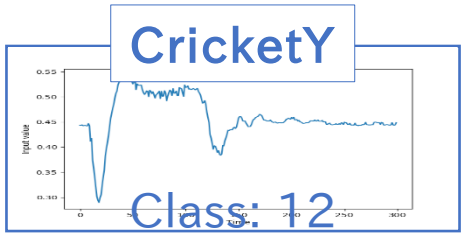
Learning by constraints

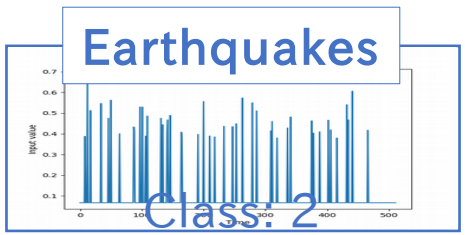
P10 Let's check how effective the proposed layer is !

- ✓ Combining the Multi. with the **head** of NN model
- ✓ Comparing when Multi. is **absent** and **Conv.** is combined instead of Multi.

Dataset







Input

Model

Multi. + NN model

Conv. + NN model

NN model

Multi. Conv.

Filter size: 96
Number of filter: 30
Activation: relu

+

NN model

RNNs
LSTMs
GRUs
Deep CNN
LENET

P11

Comparison of classification accuracies

Model	CBF	CricketY	Earthquakes
RNNs	0.331	0.07	0.748
LSTMs	0.939	0.144	0.784
GRUs	0.522	0.151	0.763
RNNs + Conv.	0.997	0.164	0.712
LSTMs + Conv.	0.982	0.09	0.683
GRUs + Conv.	0.983	**0.09**	0.698
LENET + Multi.	0.976	0.[illegible]6	0.748
Deep CNN + Multi.		0.[illegible]1	0.748
RNNs + Multi.		0.[illegible]1	0.748
LSTMs + Multi.	0.970	0.[illegible]0	0.748
GRUs + Multi. (1)	0.971	**0.636**	0.741
GRUs + Multi. (2)	0.932	0.07	0.799

Big difference

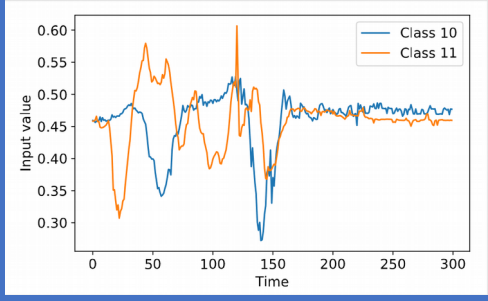
Differences in parameters

Discussion (CricketY)

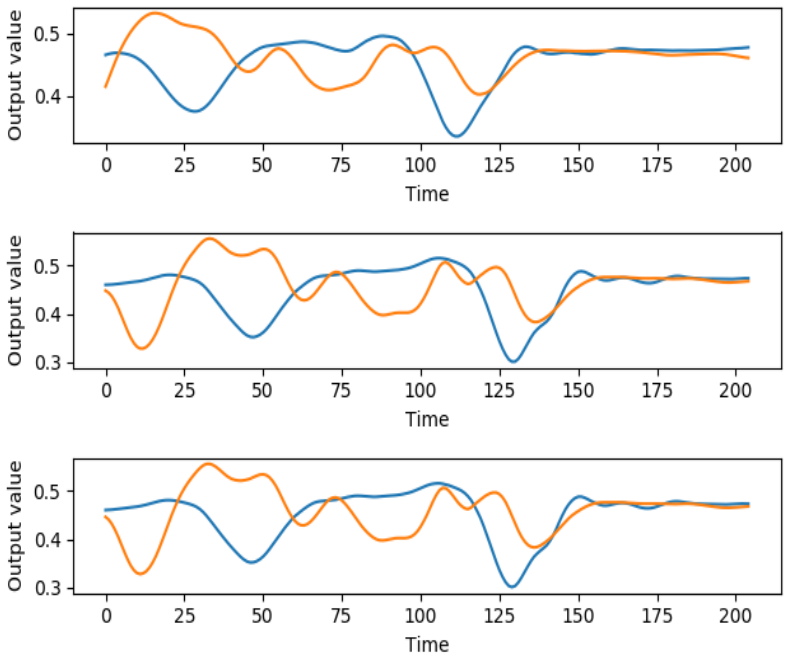
Comparing extracted features of Multi. with those of Conv.

P12

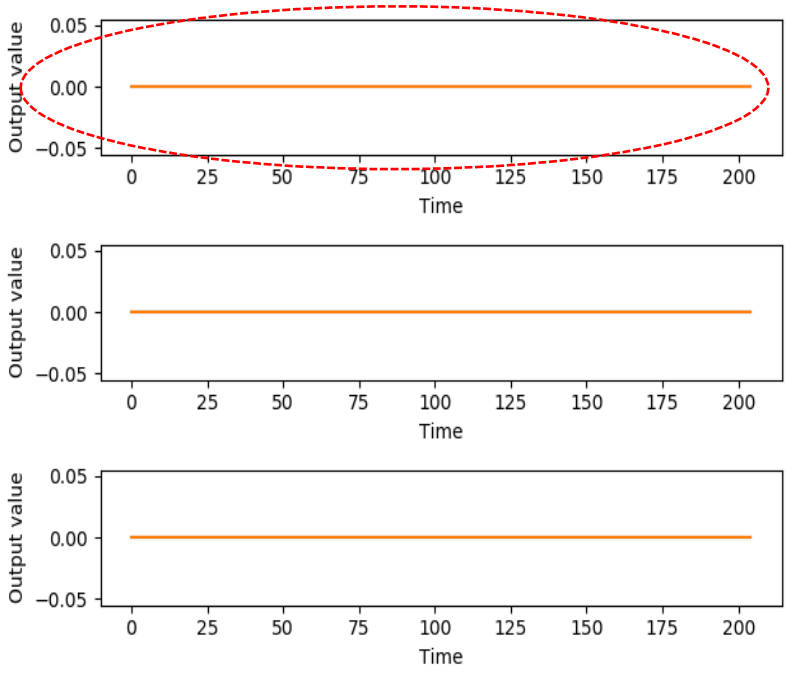
Visualization of extracted features



Multiplication layer



Convolutional layer



Big difference

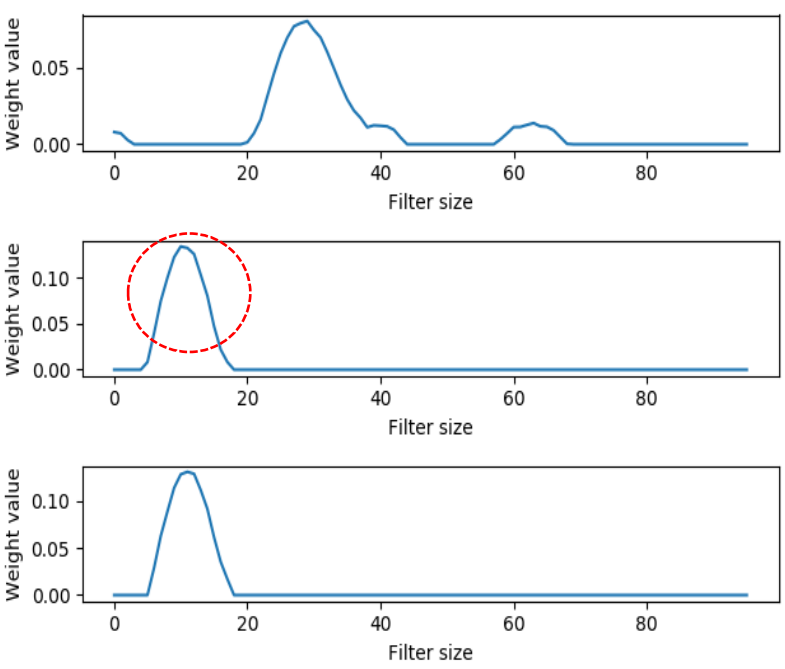
Class 10 Class 11

Comparing filter weights of Multi. with those of Conv.

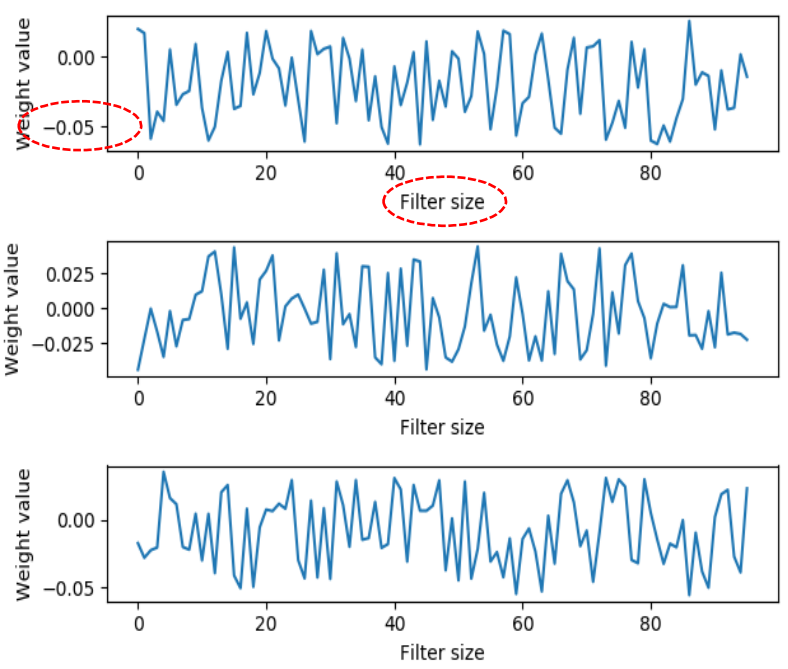
P13

Visualization of filter weights

Multiplication layer



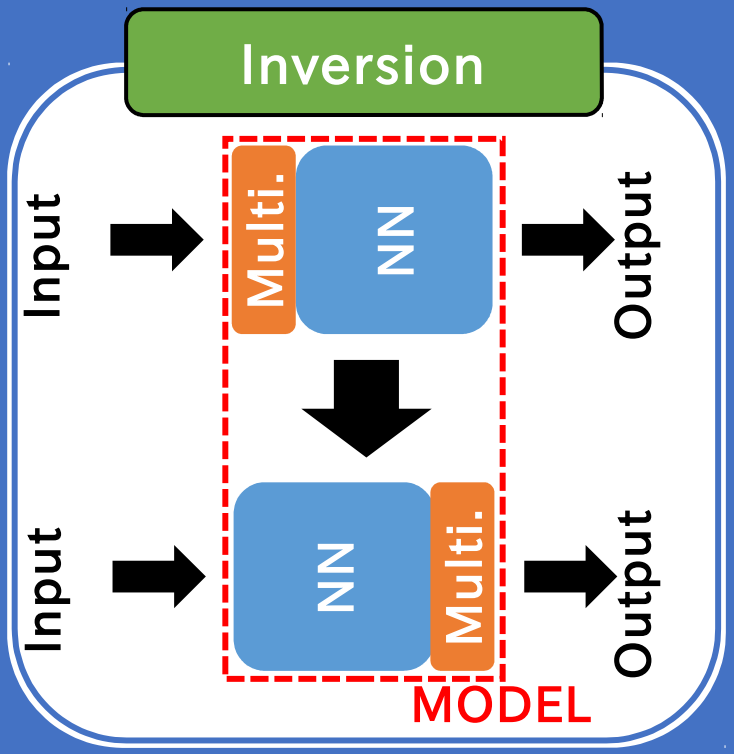
Convolutional layer

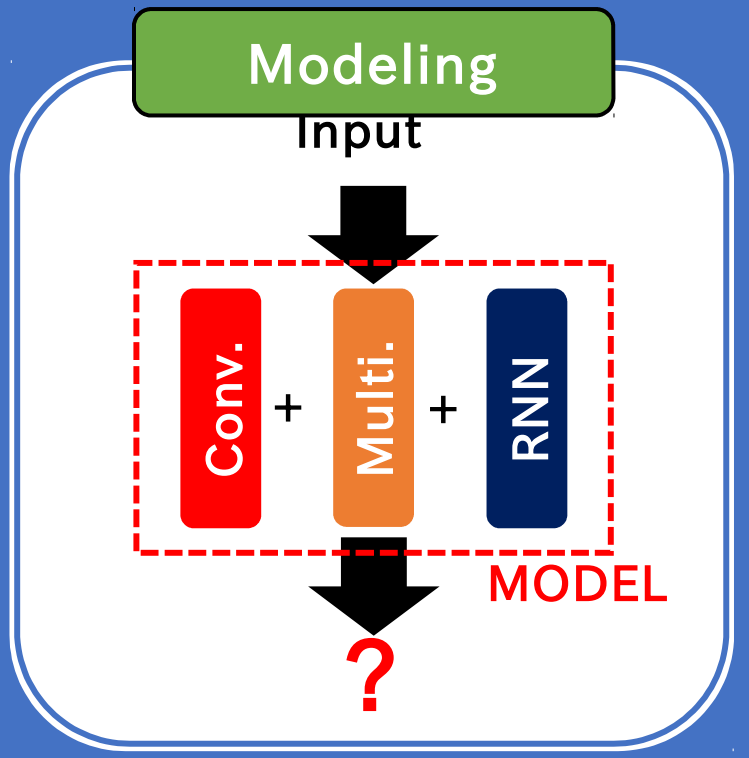


Big difference

P14

I'll do this !





Thank you for your attention !