

Comparison Long Short-Term Memory Network and Convolution Neural Network with Missing Value Imputation

LSTM 네트워크, CNN 네트워크 구현 및
결측 값 대체

Geonil Yun,
rjsdlf45@gmail.com
2021. 06. 08.



Missing Value

Deletion

- Likewise deletion

dataset

index	x	y
0	10	7
1	3	5
2	6	missing
3	missing	10
4	3	2

remove

- Remove index 2, 3 from the dataset

- Dropping variable

dataset

index	x	y	z
0	10	7	missing
1	3	5	missing
2	6	missing	missing
3	missing	10	missing
4	3	2	7

remove

- Remove 'z' column from the dataset

Missing Value Imputation

Mean-Mode imputation

- fill missing data with mean or mode from data set (or median)

dataset

index	x	y
0	10	7
1	1	7
2	6	missing
3	missing	10
4	3	7

fill 5 (mean)

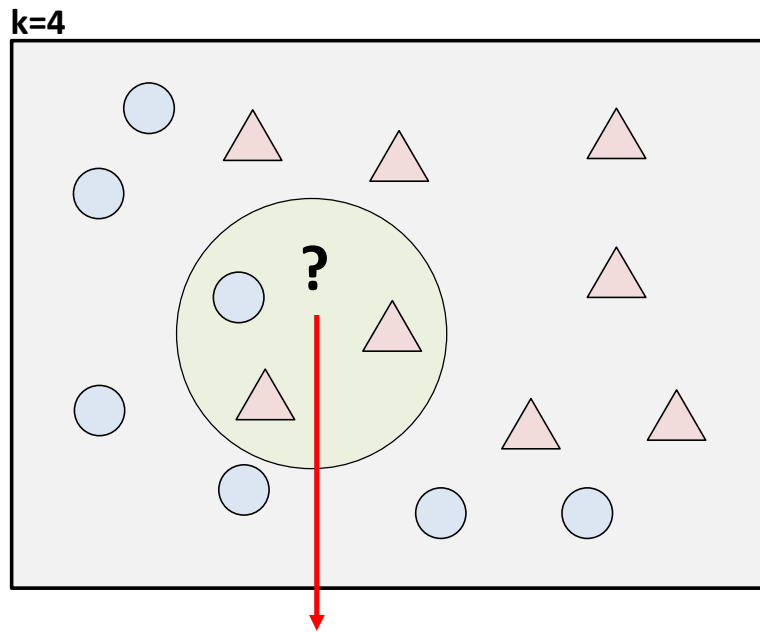
fill 7 (mode)

- using representative value : **fast**
- not consider the **correlation** with other features
- low **accuracy**
- mean reduce **variance**
- mode create **bias**

Missing Value Imputation

KNN imputation

- fill missing value by using KNN(K-Nearest Neighbor)

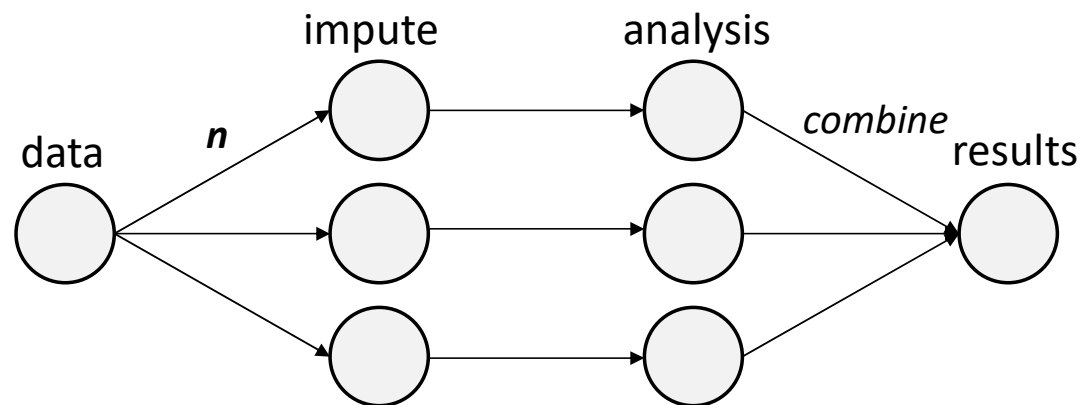


- More **accurate** than M-M impute
- Sensitive to **feature scale**
- Too many **calculations**

Missing Value Imputation

Multiple Imputation

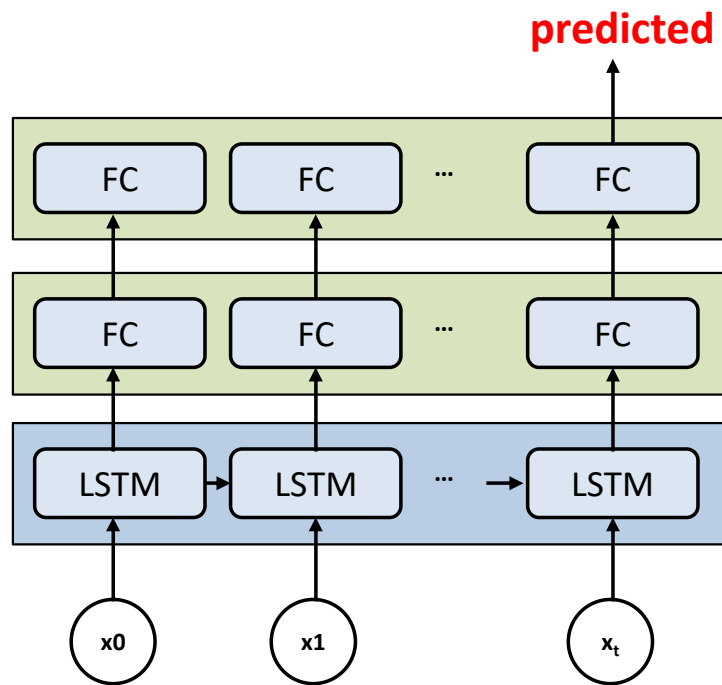
- fill missing value by using Multiple Imputation
 - ① **Imputation** : impute n datasets.
 - ② **Analysis** : analysis result of n datasets.
 - ③ **Combine** : combine result of n datasets.



Experimental setting

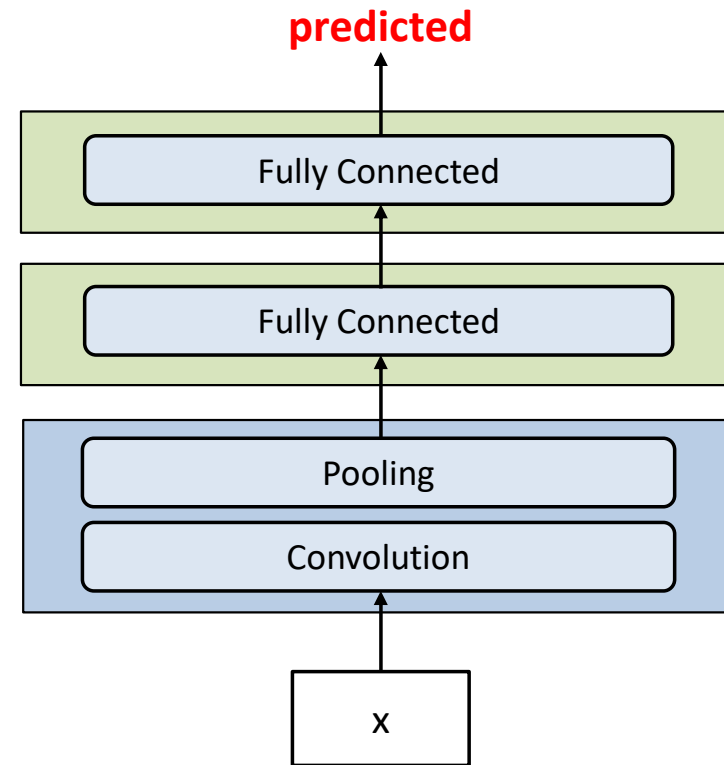
Structure (LSTM)

$$\text{MSE for Loss} = \frac{1}{2} \sum (y - t)^2$$



- input data shape : (batch, timestep, features)

Structure (CNN)

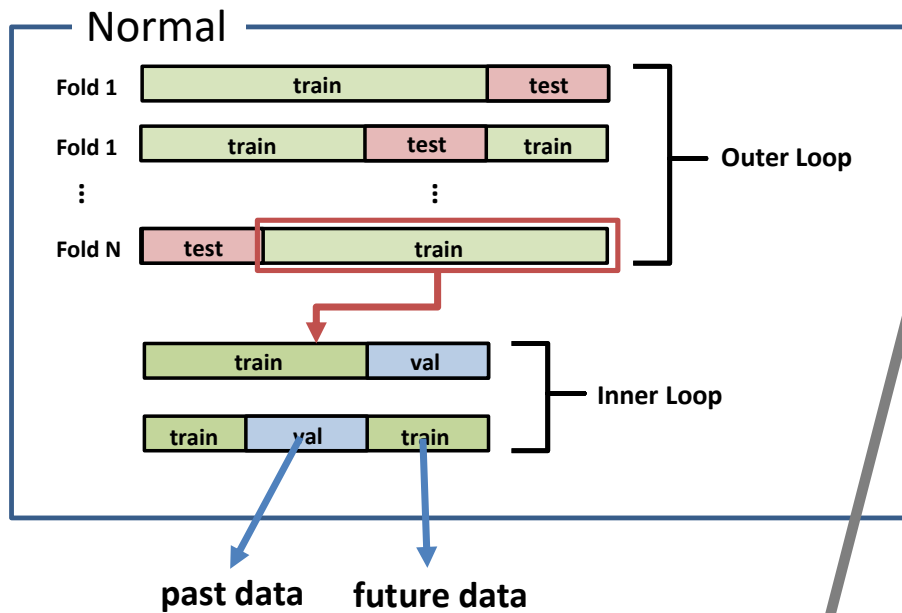


- input data shape : (batch, 1, timestep, features)

Experimental setting

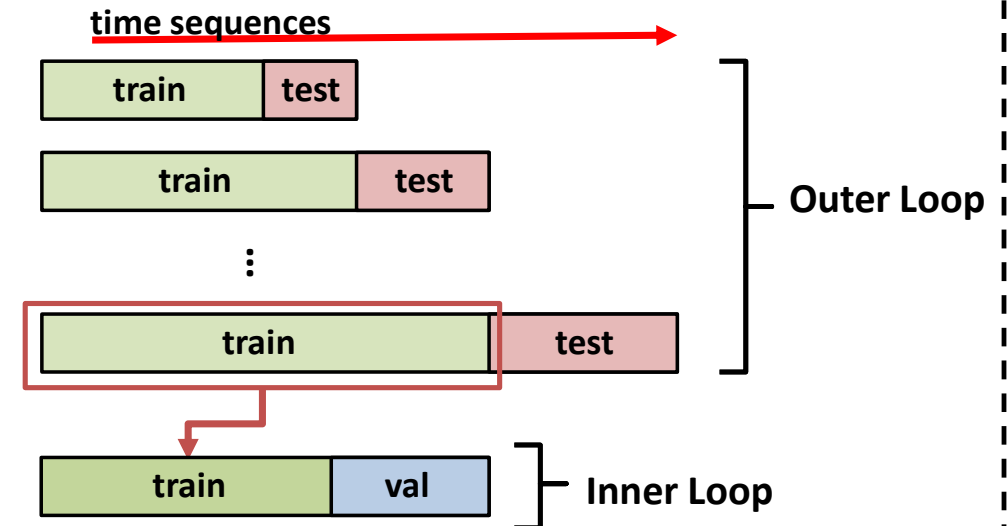
Nested Cross Validation

- Problem in **Time series data**



- train **future** data & predict **past** data

Nested Cross Validation for time series



- train **past** data & predict **future** data
- handling time series data in Nested CV

Experimental setting

Dataset

- Beijing Air pollution (uci-dataset/00381)
- Hungarian Chickenpox (uci-dataset/00580)
- household power consumption(uci-dataset/00235)  **Missing Value (use mean imputation)**

Cross validation

Dataset	Beijing air pollution		Hungarian Chickenpox		household power consumption	
Model	CNN	LSTM	CNN	LSTM	CNN	LSTM
filter size	3, 5		3, 5		3, 5	
filter num	30		30		30	
hidden size	100		100		100	
input length	24		8		30	
output length	1		1		1	
epoch	10	10	30	30	30	30
batch size	64		8, 16		32, 64	
learning rate	0.1, 0.01	0.1, 0.01, 0.001	0.1, 0.01	0.01, 0.001	0.1, 0.01	0.01, 0.001

Comparison results

- RMSE (Root Mean Squared Error)

Data	LSTM	CNN
Beijing Air Pollution	21.473	22.614
Hungary Chickenpox	70.466	69.416
Household Power Consumption	0.490	18.711

Comparison results

- MAE (Mean Absolute Error)

Data	LSTM	CNN
Beijing Air Pollution	13.374	13.742
Hungary Chickenpox	39.216	45.16
Household Power Consumption	0.270	18.128

Comparison results

- Nested Cross Validation(RMSE)

fold	Beijing air pollution		Hungarian chinckenpox		household power consumption	
	LSTM	CNN	LSTM	CNN	LSTM	CNN
1	23.32	26.41	65.83	62.57	0.13	3.68
2	22.56	28.64	74.48	77.56	0.11	34.66
3	21.47	22.61	54.89	74.16	0.16	1.91
4			35.47	34.62	1.05	81.10
5			47.62	47.61	0.39	4.86
6			51.12	61.54	0.08	0.13
7			58.90	89.56	0.28	3.98
8			48.44	54.82	0.24	14.54
9			31.58	85.07	0.19	3.44
10			70.46	69.42	0.49	18.71
Mean	22.45	25.88	53.882	65.693	0.316	16.701

Code

Model	view with code
LSTM	https://colab.research.google.com/drive/1oNW4HzBzmrjyWcWZbjFMtksP1vTV9IO2?usp=sharing
CNN	https://colab.research.google.com/drive/1Ccd7l65bJwiD2HW8XmOKTHO1PEUMuQI3?usp=sharing
LSTM-Beijing air pollution	https://colab.research.google.com/drive/1De0Tn_iD5NQBKAu_eKr3PWZwvcnNTu97?usp=sharing
CNN-Beijing air pollution	https://colab.research.google.com/drive/1HmjdyXbwUASZxzS40v4opc9myAuuhcrF?usp=sharing
LSTM-Hungarian chickenpox	https://colab.research.google.com/drive/1eeLLtCJGzQ2u_V09pXF0dHg2Oz4YHfxn?usp=sharing
CNN-Hungarian chickenpox	https://colab.research.google.com/drive/1E4PywHa-xACYhHTiMiHRwEkCHT9eBz4A?usp=sharing
LSTM-Power consumption	https://colab.research.google.com/drive/1HVLLJMPGJeixJ4H9O3lvzfUhUWp8DAXN?usp=sharing
CNN-Power consumption	https://colab.research.google.com/drive/1d7wu9u7zo_URAXYd2E01uROGcsdoSc87?usp=sharing

Thank you