




M5 Forecasting

현예성 김민석 최성웅
문구영 이노아 강호석



INDEX

1. 주제 소개

2. LSTM

3. Light GBM



Featured Prediction Competition

M5 Forecasting - Accuracy

Estimate the unit sales of Walmart retail goods

\$50,000

Prize Money



University of Nicosia · 4,055 teams · a month to go (a month to go until merger deadline)

Files

- `calendar.csv` - Contains information about the dates on which the products are sold.
- `sales_train_validation.csv` - Contains the historical daily unit sales data per product and store [`d_1` - `d_1913`]
- `sample_submission.csv` - The correct format for submissions. Reference the [Evaluation](#) tab for more info.
- `sell_prices.csv` - Contains information about the price of the products sold per store and date.
- `sales_train_evaluation.csv` - Available once month before competition deadline. Will include sales [`d_1` - `d_1941`]

Data - Item Information

	id	item_id	dept_id	cat_id	store_id	state_id	d_1	d_2	d_3	...	d_1908	d_1909	d_1910	d_1911	d_1912	d_1913
1	ID information						Sales information									
2																
3																
...																
30489																
30490																

Data - Day Information

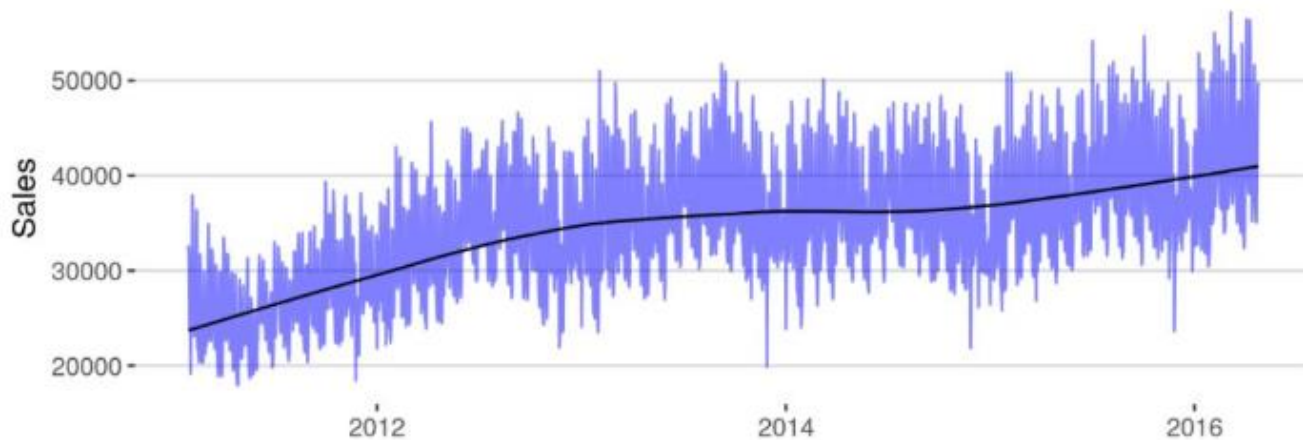
	date	wm_yr_wk	weekday	wday	month	year	d	event_name_1	event_type_1	event_name_2	event_type_2	snap_CA	snap_TX	snap_WI
0	2011-01-29	11101	Saturday	1	1	2011	d_1	NaN	NaN	NaN	NaN	0	0	0
1	2011-01-30	11101	Sunday	2	1	2011	d_2	NaN	NaN	NaN	NaN	0	0	0
2	2011-01-31	11101	Monday	3	1	2011	d_3	NaN	NaN	NaN	NaN	0	0	0
3	2011-02-01	11101	Tuesday	4	2	2011	d_4	NaN	NaN	NaN	NaN	1	1	0
4	2011-02-02	11101	Wednesday	5	2	2011	d_5	NaN	NaN	NaN	NaN	1	0	1

Data - Price and Submission

	store_id	item_id	wm_yr_wk	sell_price
0	CA_1	HOBBIES_1_001	11325	9.58
1	CA_1	HOBBIES_1_001	11326	9.58
2	CA_1	HOBBIES_1_001	11327	8.26
3	CA_1	HOBBIES_1_001	11328	8.26
4	CA_1	HOBBIES_1_001	11329	8.26

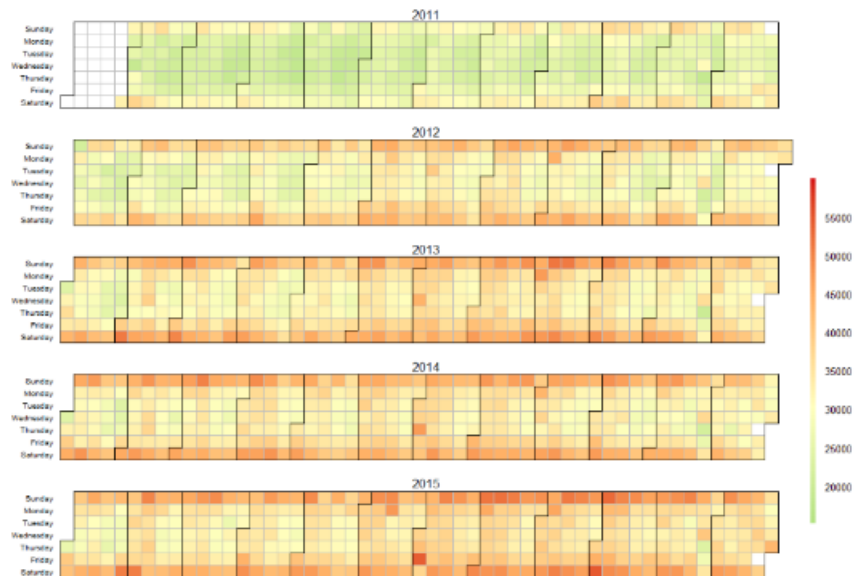
id	F1	F2	F3	F4	F5	F6	F7
HOBBIES_1_001_CA_1_evaluation	0	0	0	0	0	0	0
HOBBIES_1_002_CA_1_evaluation	0	0	0	0	0	0	0
HOBBIES_1_003_CA_1_evaluation	0	0	0	0	0	0	0
HOBBIES_1_004_CA_1_evaluation	0	0	0	0	0	0	0
HOBBIES_1_005_CA_1_evaluation	0	0	0	0	0	0	0
...
FOODS_3_823_WI_3_evaluation	0	0	0	0	0	0	0
FOODS_3_824_WI_3_evaluation	0	0	0	0	0	0	0
FOODS_3_825_WI_3_evaluation	0	0	0	0	0	0	0
FOODS_3_826_WI_3_evaluation	0	0	0	0	0	0	0
FOODS_3_827_WI_3_evaluation	0	0	0	0	0	0	0

EDA - All items



전체 데이터의 Year Seasonality 확인

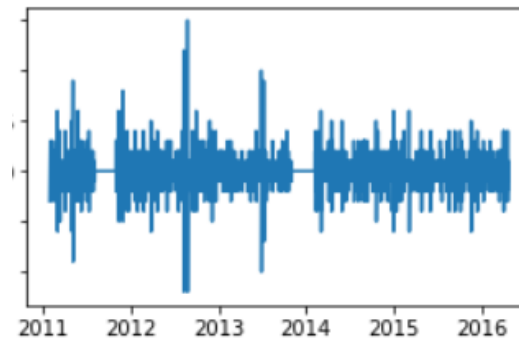
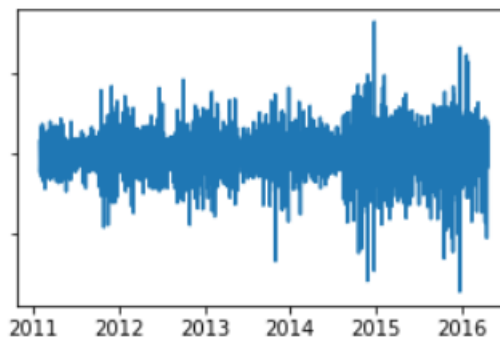
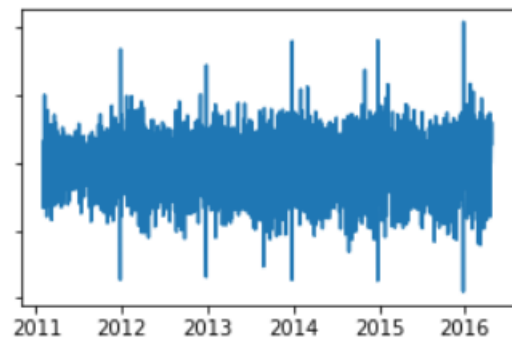
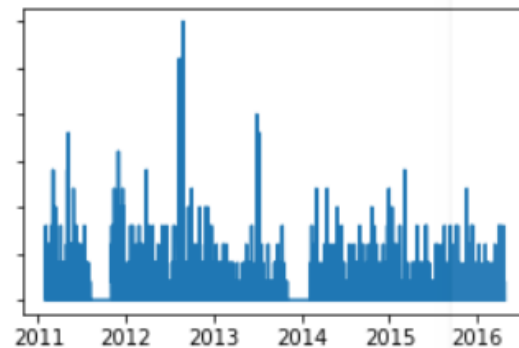
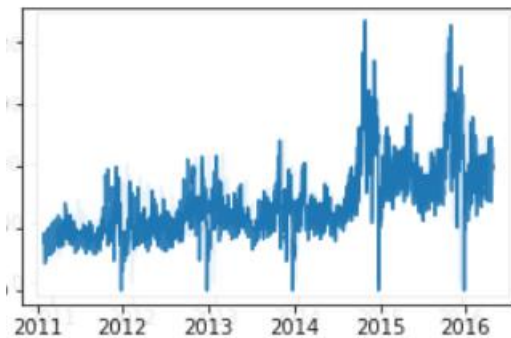
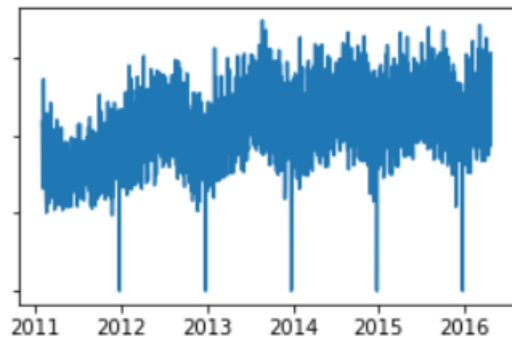
EDA - Week



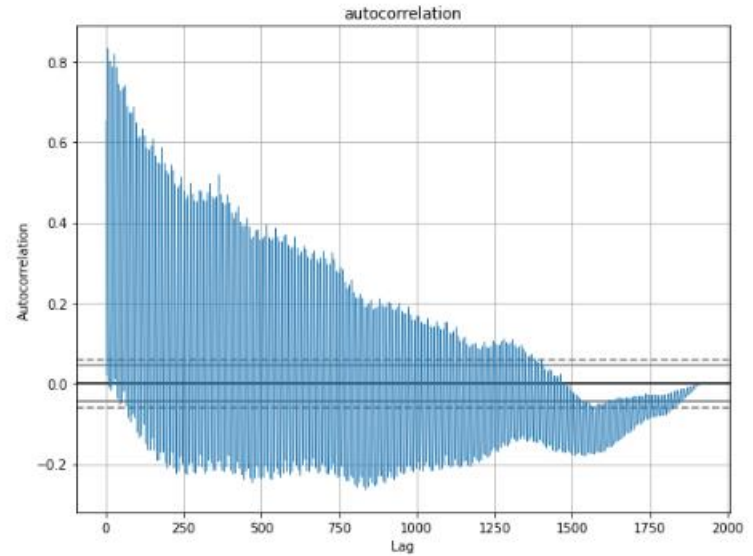
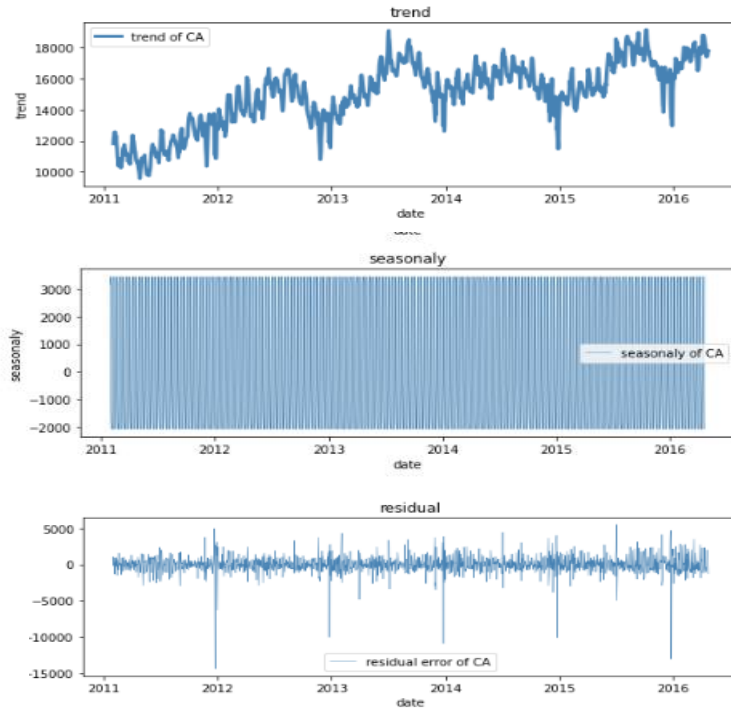
- Weekly Seasonality 확인

- 주말마다 매출이 상승

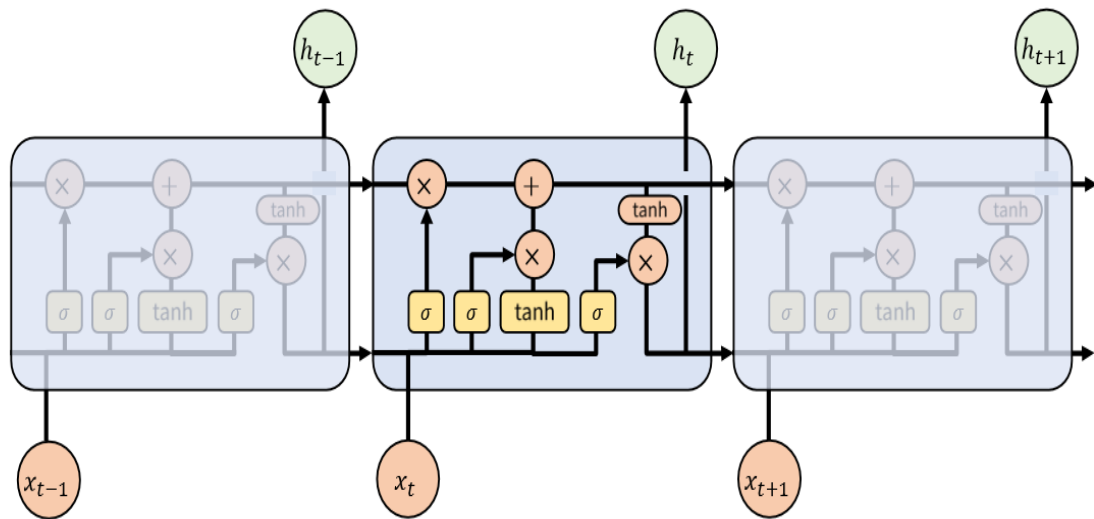
EDA - Items



EDA - Time Series



Long Short-Term Memory



다음 층으로 기억된 값을 넘길 지, 넘기지 않을 지를 관리

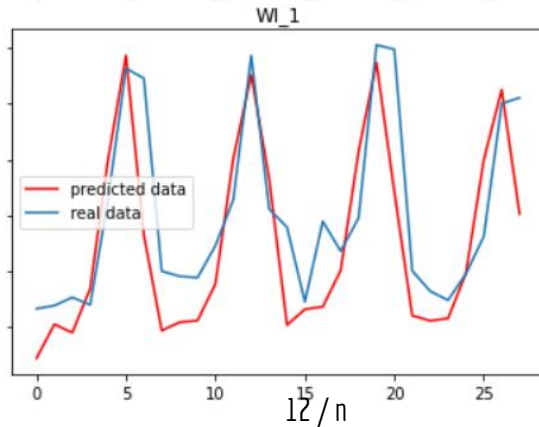
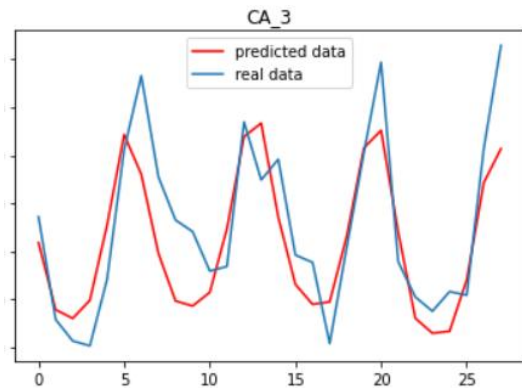
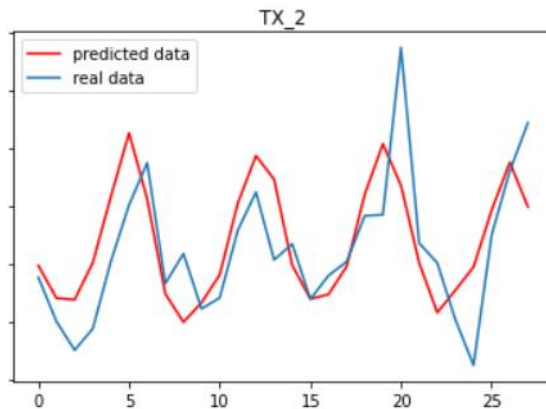
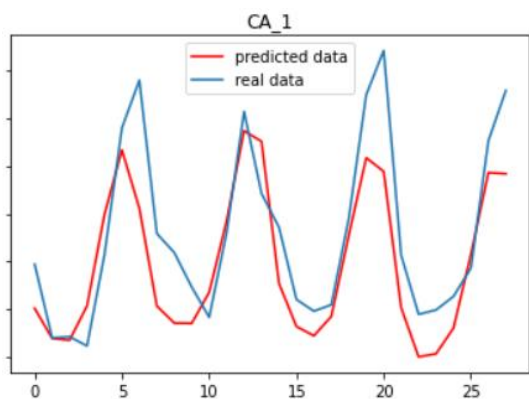
과거의 신호를 전파

기울기 소실 문제를 극복

기억할 것은 오래 기억하고, 잊을 것은 빨리 잊어버리는 능력

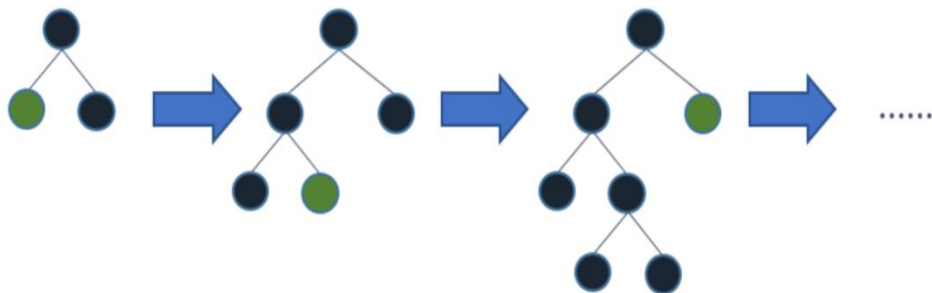
삭제, 입력, 출력 게이트로 구성

LSTM result by store group



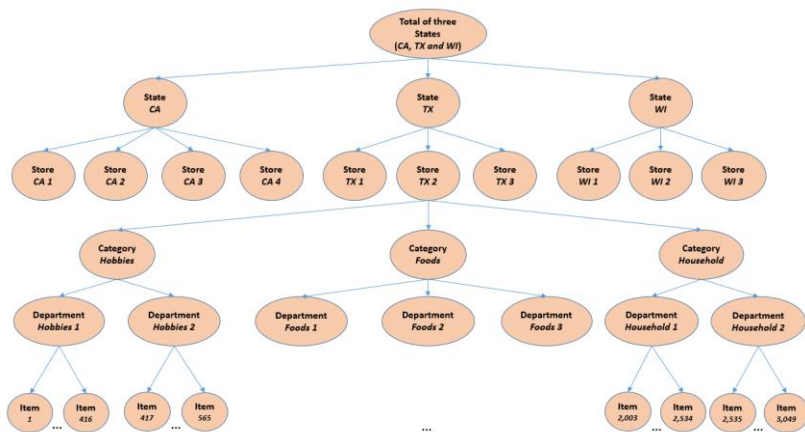
나쁘지 않은 예측률을 확인

Light GBM



- 빠른 학습 속도 (GOSS)
- 적은 메모리 사용
- 대용량 데이터 처리 가능
- Overfitting 문제 발생 가능

Light GBM - Features

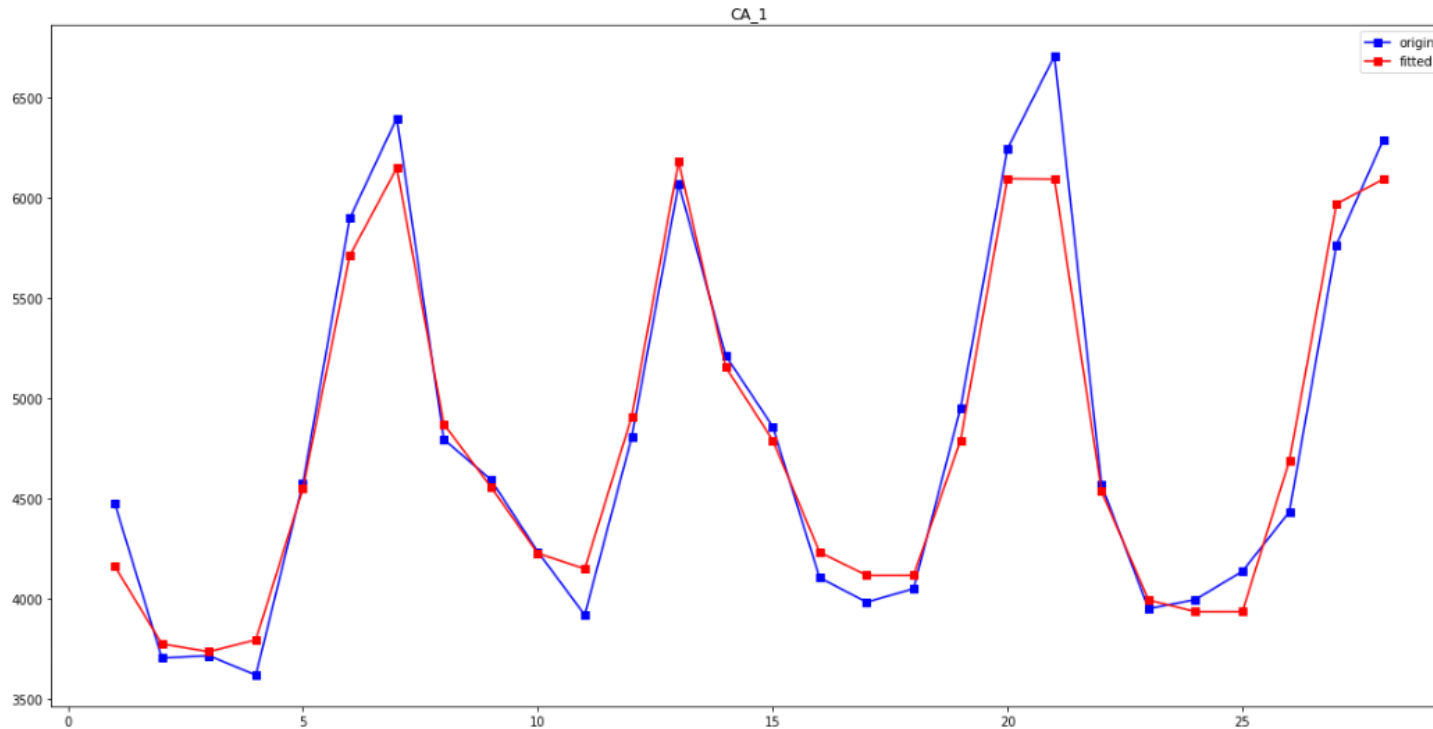


lag_7	lag_28	rmean_7_7	rmean_28_7	rmean_7_28	rmean_28_28	week	quarter	mday
0.0	0.0	0.000000	0.142857	0.178571	0.285714	48	4	29
0.0	0.0	0.000000	0.000000	0.142857	0.285714	48	4	30
0.0	1.0	0.000000	0.142857	0.142857	0.321429	48	4	1
0.0	0.0	0.000000	0.142857	0.142857	0.321429	48	4	2
1.0	1.0	1.428571	1.428571	1.607143	1.785714	48	4	29

Light GBM - Hyperparametric Tuning

```
params = {  
    "objective" : "poisson", #poisson regression  
    "metric" : "rmse",  
    "force_row_wise" : True, #learning direction (relatively small columns)  
    "learning_rate" : 0.075,  
    "sub_row" : 0.75,  
    "bagging_freq" : 1, #0 means disable bagging; k means perform bagging at every k iteration  
    "lambda_l2" : 0.1, #L2 regularization  
    "metric": ["rmse"],  
    'verbosity': 1, # Info  
    'num_iterations' : 1200,  
    'num_leaves': 128,  
    "min_data_in_leaf": 100,  
}
```

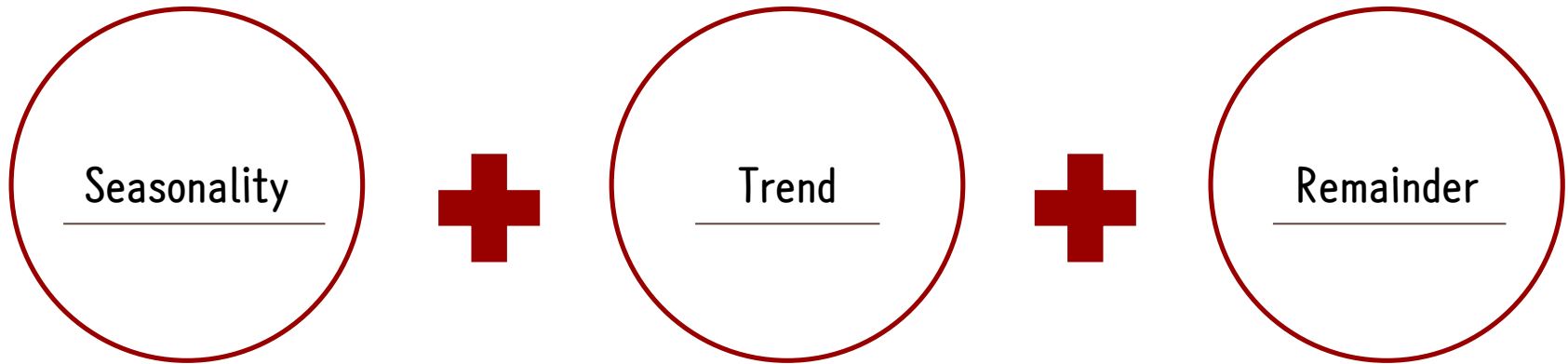
Light GBM - Fit



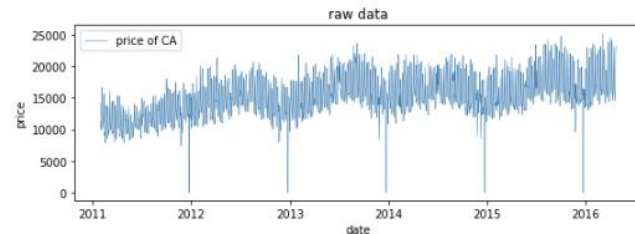
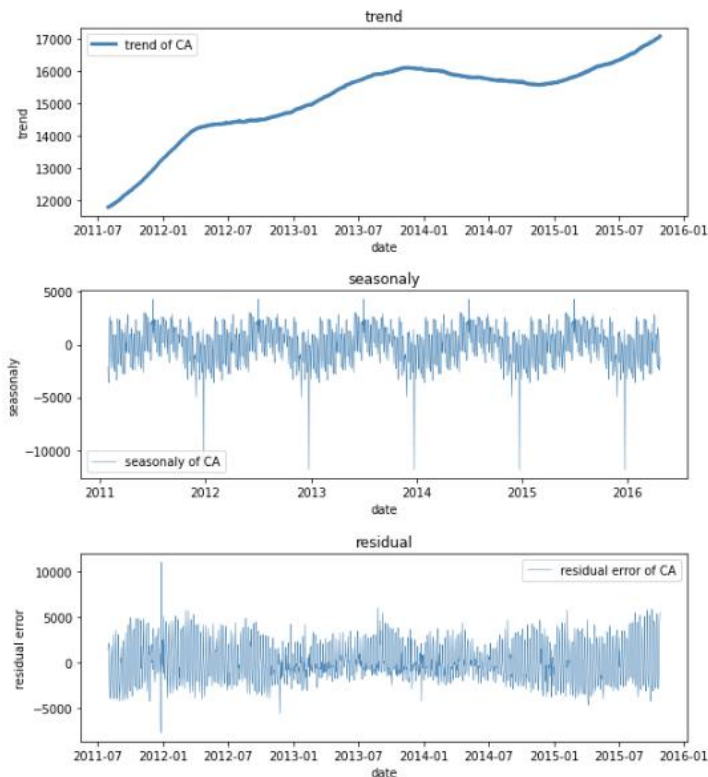
0.49670

Light GBM - Seasonal Decomposition

$$y_t = S_t + T_t + R_t$$



Light GBM - Seasonal Decomposition



$$\hat{S}_t + \hat{T}_t + \hat{R}_t \Rightarrow \hat{y}_t$$

Light GBM - Seasonal Decomposition

- CA_1 에 대해서만 분해
- 모든 상품에 대해서 확장 ?

```
mean_squared_error(eval_p, submission_p)
```

4.699109448285803

```
mean_squared_error(eval_p, decom_p)
```

4.116426899749475

Thank You!