Ensemble Learning

7주차
담당: 14기 박상준

1. **Boosting Models** LGBM Catboost
2. **Coding Session**

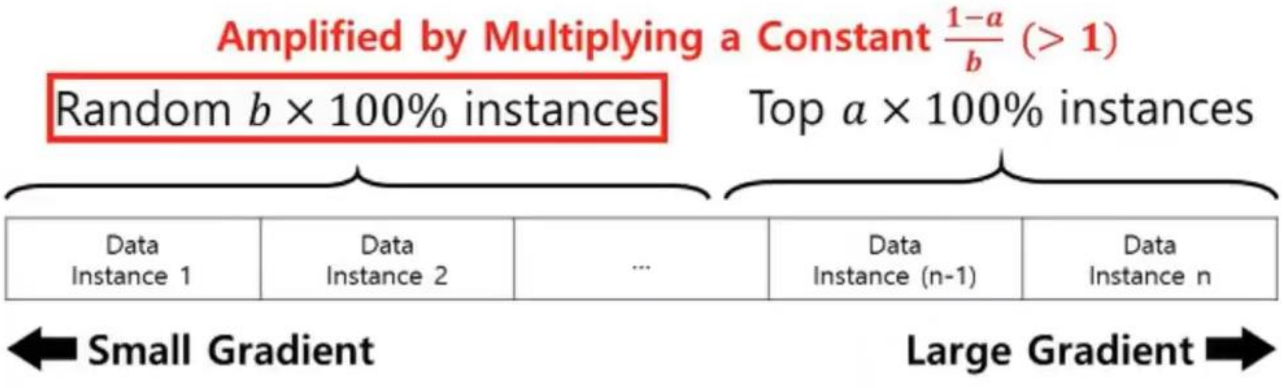
Light GBM

Gradient based One-Side Sampling (GOSS)

Exclusive Feature Bundling (EFB)

Light GBM

Gradient based One-Side Sampling (GOSS)



Light GBM

Exclusive Feature Bundling (EFB)

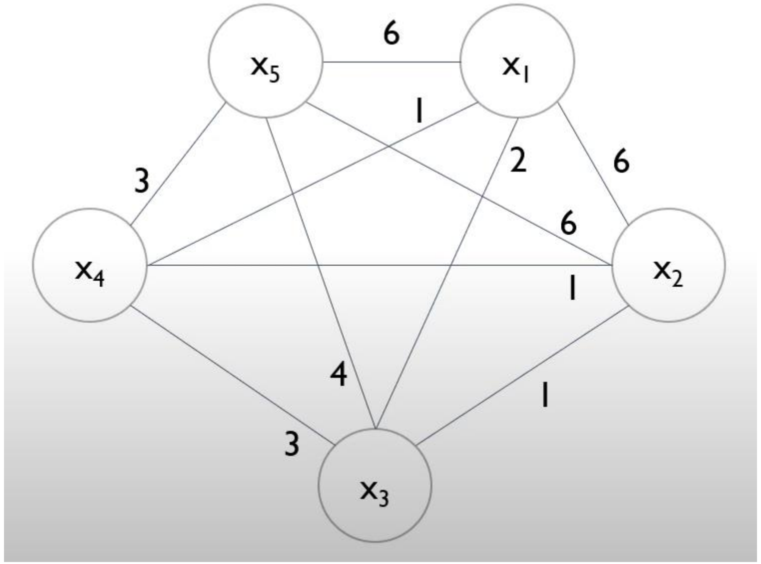
	x_1	x_2	x_3	x_4	x_5
I_1	1	1	0	0	1
I_2	0	0	1	1	1
I_3	1	2	0	0	2
I_4	0	0	2	3	1
I_5	2	1	0	0	3
I_6	3	3	0	0	1
I_7	0	0	3	0	2
I_8	1	2	3	4	3
I_9	1	0	1	0	0
I_{10}	2	3	0	0	2

	x_1	x_2	x_3	x_4	x_5
x_1	-	6	2	1	6
x_2	6	-	1	1	6
x_3	2	1	-	3	4
x_4	1	1	3	-	3
x_5	6	6	4	3	-

	x_5	x_1	x_2	x_3	x_4
d	19	15	14	10	8

Light GBM

Exclusive Feature Bundling (EFB)



Light GBM

Exclusive Feature Bundling (EFB)

	x_1	x_2	x_3	x_4	x_5
I_1	1	1	0	0	1
I_2	0	0	1	1	1
I_3	1	2	0	0	2
I_4	0	0	2	3	1
I_5	2	1	0	0	3
I_6	3	3	0	0	1
I_7	0	0	3	0	2
I_8	1	2	3	4	3
I_9	1	0	1	0	0
I_{10}	2	3	0	0	2

	x_5	x_1	x_4	x_2	x_3
I_1	1	1	0	1	0
I_2	1	0	1	0	1
I_3	2	1	0	2	0
I_4	1	0	3	0	2
I_5	3	2	0	1	0
I_6	1	3	0	3	0
I_7	2	0	0	0	3
I_8	3	1	4	2	3
I_9	0	1	0	0	1
I_{10}	2	2	0	3	0

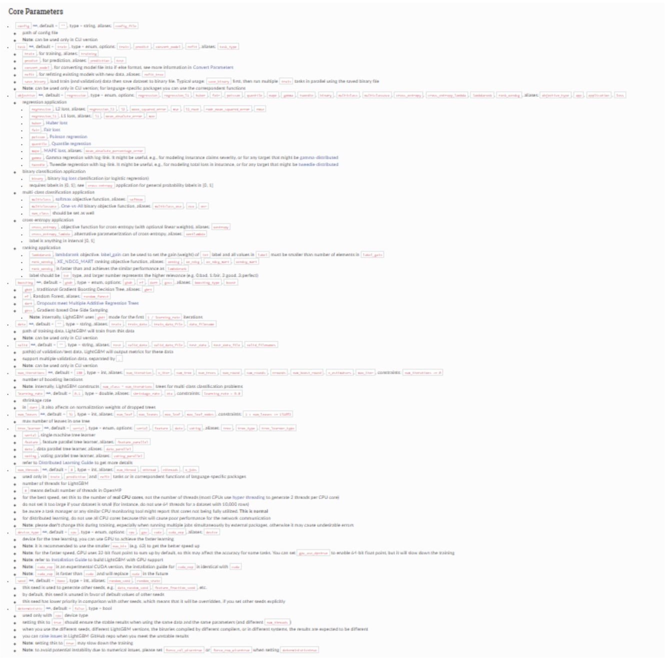
Light GBM

Exclusive Feature Bundling (EFB)

	x_5	x_1	x_4	x_2	x_3
I_1	1	1	0	1	0
I_2	1	0	1	0	1
I_3	2	1	0	2	0
I_4	1	0	3	0	2
I_5	3	2	0	1	0
I_6	1	3	0	3	0
I_7	2	0	0	0	3
I_8	3	1	4	2	3
I_9	0	1	0	0	1
I_{10}	2	2	0	3	0

	x_5	x_{14}	x_{23}
I_1	1	1	1
I_2	1	4	4
I_3	2	1	2
I_4	1	6	5
I_5	3	2	1
I_6	1	3	3
I_7	2	0	6
I_8	3	1	2
I_9	0	1	4
I_{10}	2	2	3

Coding Session


Core Parameters

- boosting_type(gbdt, rf, dart, goss): 기본 설정은 gbdt(GBM), goss로 바꾸면 GOSS 적용 가능
- top_rate(default = 0.2): retain ratio of large gradient data
- low_rate(default = 0.1): retain ratio of small gradient data
- enable_bundle(default = True): EFB 실행 여부

Catboost

Target Leakage

→ Ordered TS(Target Statistics)

Prediction Shift

→ Ordered Boosting

Catboost

Ordered TS

	…	x^i	…
I_1	…	A	…
I_2	…	B	…
I_3	…	C	…
I_4	…	A	…
I_5	…	B	…
I_6	…	C	…
I_7	…	B	…
I_8	…	C	…
I_9	…	C	…
I_{10}	…	C	…

	…	$x^i(A)$	$x^i(B)$	$x^i(C)$	…
I_1	…	1	0	0	…
I_2	…	0	1	0	…
I_3	…	0	0	1	…
I_4	…	1	0	0	…
I_5	…	0	1	0	…
I_6	…	0	0	1	…
I_7	…	0	1	0	…
I_8	…	0	0	1	…
I_9	…	0	0	1	…
I_{10}	…	0	0	1	…

Catboost

Ordered TS

Index	...	X1	...	Y
L1	...	A	...	1
L2	...	B	...	1
L3	...	C	...	1
L4	...	A	...	0
L5	...	B	...	1
L6	...	C	...	1
L7		B	...	0
L8		C	...	1
L9		C	...	0

Index	...	X1(TS)	...	Y
L1	...	0.5	...	1
L2	...	0.67	...	1
L3	...	0.75	...	1
L4	...	0.5	...	0
L5	...	0.67	...	1
L6	...	0.75	...	1
L7		0.67	...	0
L8		0.75	...	1
L9		0.75	...	0

Catboost

Ordered TS

✓ Another popular method: to group categories by target statistics (TS)

- Greedy TS with smoothing

$$\hat{x}_k^i = \frac{\sum_{j=1}^{n} \mathbb{1}_{\{x_j^i = x_k^i\}} \cdot y_j + ap}{\sum_{j=1}^{n} \mathbb{1}_{\{x_j^i = x_k^i\}} + a}$$

- a > 0 is a parameter
- A common setting for p is the average target value in the dataset
- Used to remove the negative effect of low-frequency noisy categories

Catboost

Ordered TS

$$\hat{x}_k^i = \frac{\sum_{j=1}^{n} \mathbb{1}_{\{x_j^i = x_k^i\}} \cdot y_j + ap}{\sum_{j=1}^{n} \mathbb{1}_{\{x_j^i = x_k^i\}} + a}$$

	Y=1	Y=0	TS
A	10	10	0.5
B	40	10	0.8
C	10	40	0.2
D	25	25	0.5
E	1	0	1

Catboost

Ordered TS

Index	...	X	...	TS	Y
L1	...	A	...	0.000	1
L2	...	B	...	1.000	1
L3	...	C	...		1
L4	...	A	...		0
L5	...	B	...		1
L6	...	C	...		1
L7	...	B	...		0
L8	...	C	...		1
L9	...	C	...		0
L10	...	C	...		1

$$\hat{x}_k^i = \frac{\sum_{\mathbf{x}_j \in \mathcal{D}_k} \mathbb{1}_{\{x_j^i = x_k^i\}} \cdot y_j + ap}{\sum_{\mathbf{x}_j \in \mathcal{D}_k} \mathbb{1}_{\{x_j^i = x_k^i\}} + a}$$

$$= \frac{0 + 0.1 \times 0}{0 + 0.1} = 0$$

$$= \frac{0 + 0.1 \times 1.0}{0 + 0.1} = 1.0$$

Catboost

Ordered TS

Index	...	X	...	TS	Y
L1	...	A	...	0.000	1
L2	...	B	...	1.000	1
L3	...	C	...	1.000	1
L4	...	A	...	1.000	0
L5	...	B	...	0.977	1
L6	...	C	...		1
L7	...	B	...		0
L8	...	C	...		1
L9	...	C	...		0
L10	...	C	...		1

$$= \frac{1 + 0.1 \times 0.75}{1 + 0.1} = 0.977$$

Catboost

Ordered TS

Index	...	X	...	TS	Y
L1	...	A	...	0.000	1
L2	...	B	...	1.000	1
L3	...	C	...	1.000	1
L4	...	A	...	1.000	0
L5	...	B	...	0.977	1
L6	...	C	...	0.982	1
L7	...	B	...	0.922	0
L8	...	C	...		1
L9	...	C	...		0
L10	...	C	...		1

$$= \frac{2 + 0.1 \times 0.833}{2 + 0.1} = 0.992$$

3. Cat Boost의 직관적 이해

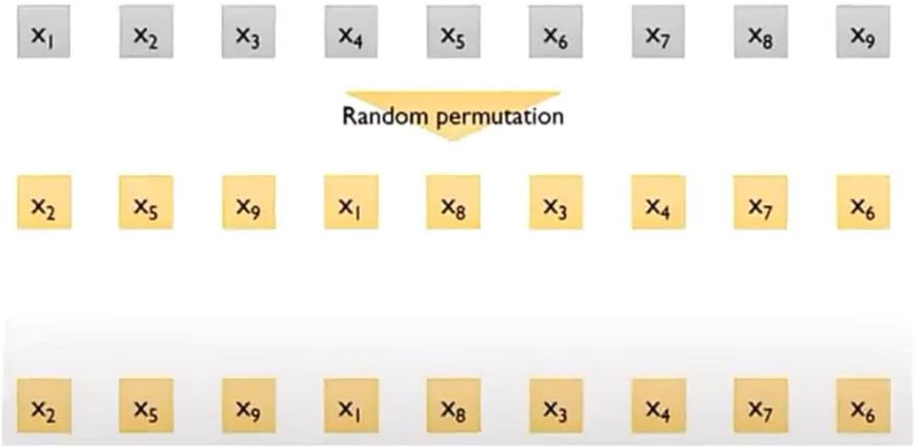
Ordered Boosting

	F1	F2	F3	F4	F5	F6	F7	Y
X1								
X2								
X3								
X4								
X5								
X6								
X7								
X8								

Catboost

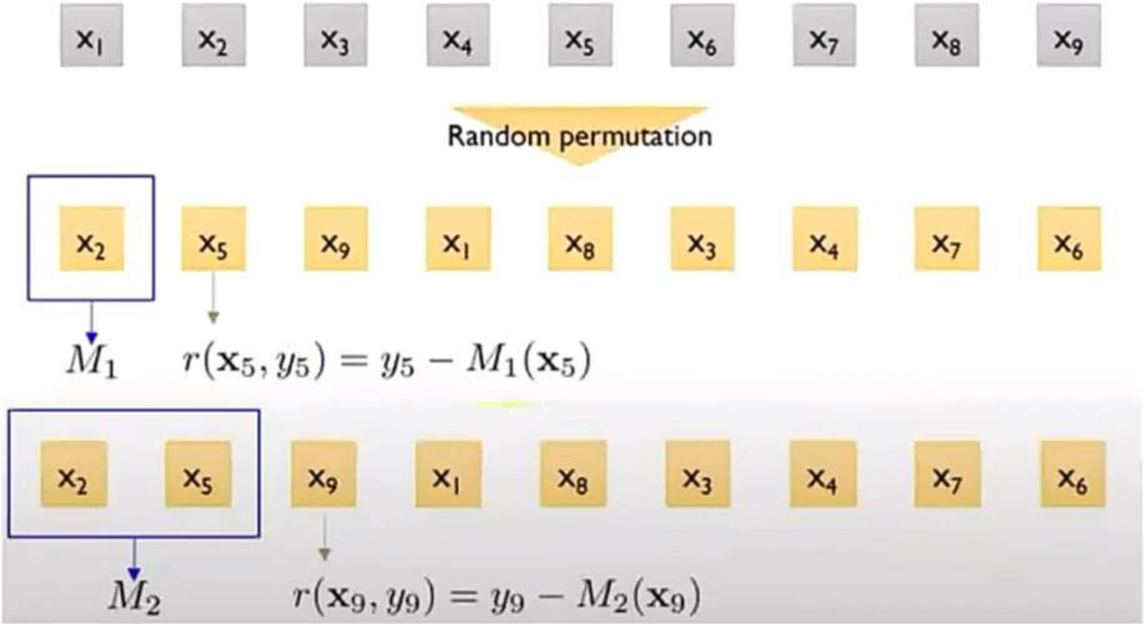
Ordered Boosting

- Ordered Boosting
  - ✓ A boosting algorithm not suffering from the prediction shift problem



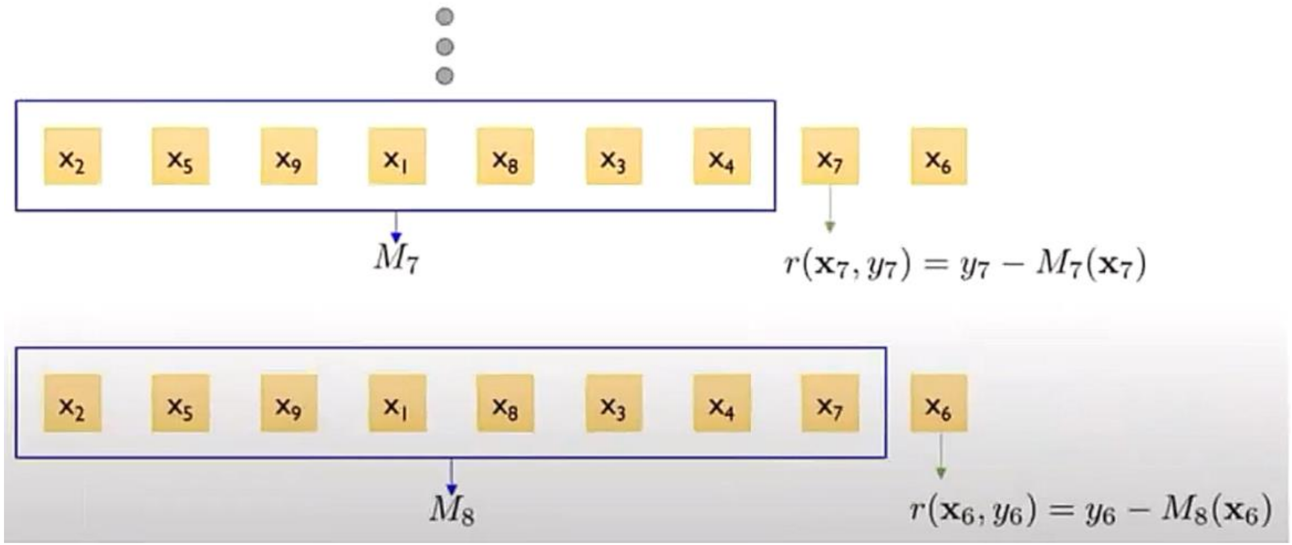
Catboost

Ordered Boosting



Catboost

Ordered Boosting



Coding Session

- cat_features: 범주형 feature 입력
- boosting_type(ordered, plain): “ordered” 설정 시 ordered boosting
- has_time : 시간에 따라 변화하는 데이터 유무
- sum_models: 범주 특성 별 모델을 각자 생성하고 후에 결합