

< Tree ensembles - Sampling with replacement >

- Sampling with replacement : a technique for building tree ensemble

(복원추출: 한번 사용한 결과를 다시 얻을 수 있도록 모집단에 다시 포함시켜 사용하는 추출방법)

ex) ①, ②, ③, ④ 공을 주머니에 넣고 하나씩 꺼내기

i)  공 하나 꺼내기

ii)  꺼낸 공 다시 넣기

... 4번 반복 \Rightarrow sampling with replacement :

① 2 1 1 4
② 1 2 4 3
③ 4 4 3 2
⋮

\Downarrow Tree ensemble 에 적용

ex) 5 cats, 5 dogs (1 2 3 4 5 , 1 2 3 4 5)

i) 총 10개의 training examples을 가상의 bag에 넣음

ii) 하나씩 꺼내고 다시 넣기를 반복해서 총 10개를 뽑음

	Ear shape	Face shape	Whiskers	Cat
1	Pointy	Round	Present	1
4	Floppy	Not round	Absent	0
3	Pointy	Round	Absent	1
4	Floppy	Not round	Absent	0
⋮	⋮	⋮	⋮	⋮
1	Pointy	Round	Present	1

repeat sampling with replacement
10 times

10 training examples

\rightarrow 이렇게 sampling with replacement를 통해
얻은 데이터셋이 새로운 training dataset이 됨

\therefore Sampling with replacement \Rightarrow construct new training dataset.

bit similar to but also quite different from original set