N / A / N / O / D / E / G / R / E / E

지도학습 알고리즘

13. Random Forest

Ensemble learning

- Ensemble learning
  - 한 전문가의 의견보다 여러 사람의 종합된 의견이 더 나은 경우가 많음
  - 결과가 중요한 의사 결정에서는 여러 전문가의 의견을 구하게 됨
  - 하나의 좋은 예측기보다 보통 예측기 집단의 예측이 더 나음
- Ensemble learning의 목표
  - 여러 다양한 의견을 고려
  - 논리적 과정을 통해 결합
  - 결정에 대한 신뢰성을 높임

Ensemble learning

- Ensemble을 통한 변동성의 축소
  - 모델들의 오류는 각 샘플에 대해 다른 오류를 발생시키지만 옳은 분류에 대해서는 공통적으로 일치한다고 가정
  - Ensemble model에서 출력을 averaging 하는 것이 오류 요소들에 대해 averaging out하게 만들어 전체 모델의 오류를 감소

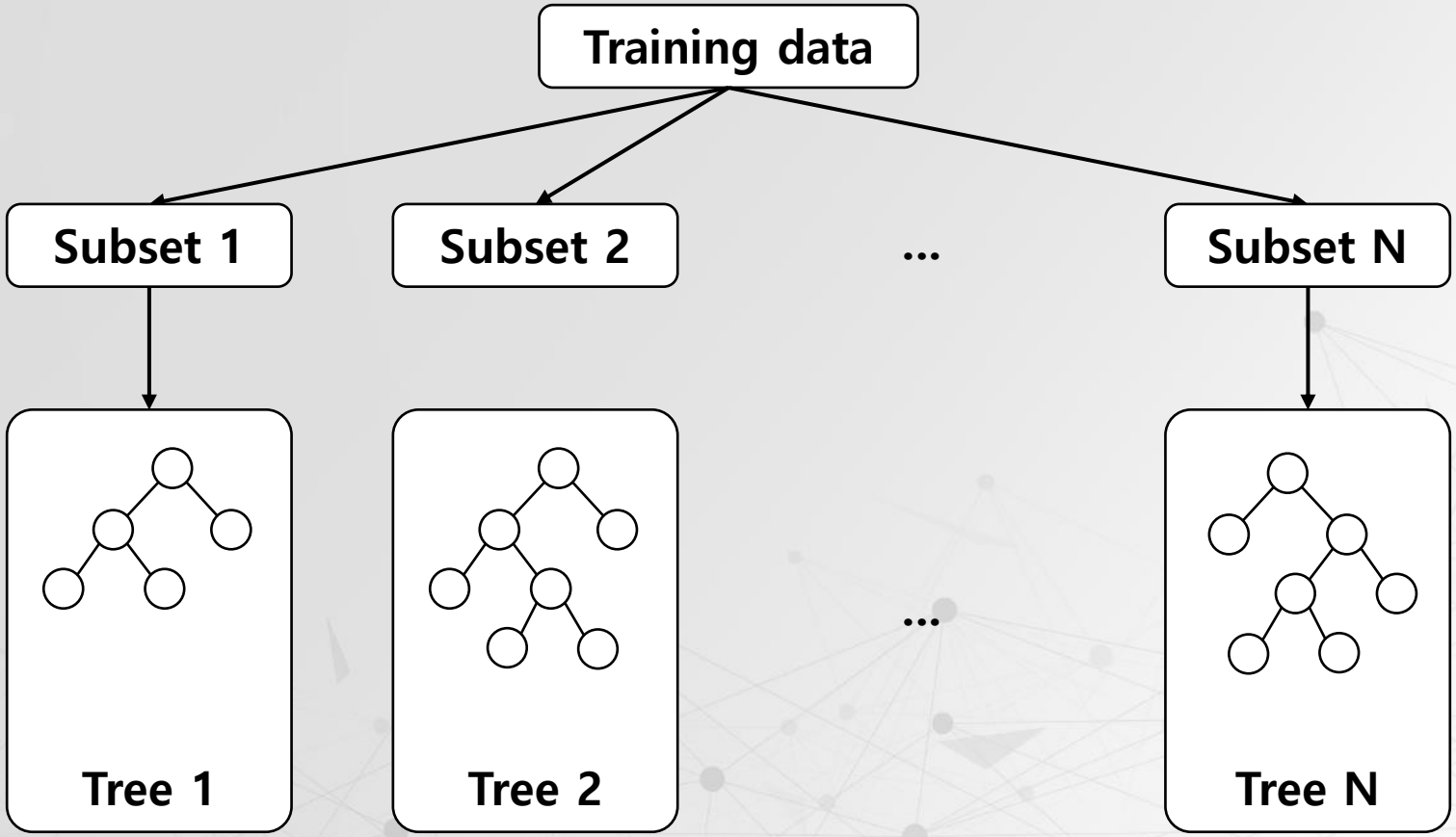
- Decision tree와 ensemble learning
  - Decision tree는 base model로써의 활용도가 높음
  - Low computational complexity: 데이터의 크기가 방대해도 빠른 구축이 가능
  - Nonparametric: 데이터의 분포에 대한 전제가 필요하지 않음

Ensemble learning: diversity의 확보

- 다음 조건을 만족할 때, ensemble model이 base model보다 우수
  - Base model이 서로 독립적이고
  - Base model이 무작위로 예측할 때보다 성능이 뛰어난 경우
- Ensemble model의 성능을 확보하기 위한 핵심: 다양성과 무작위성의 확보
  - 훈련 데이터의 서로 다른 부분집합을 사용하여 학습: bootstrap, bagging
  - 사용 가능한 feature의 서로 다른 부분집합을 사용: random subspace

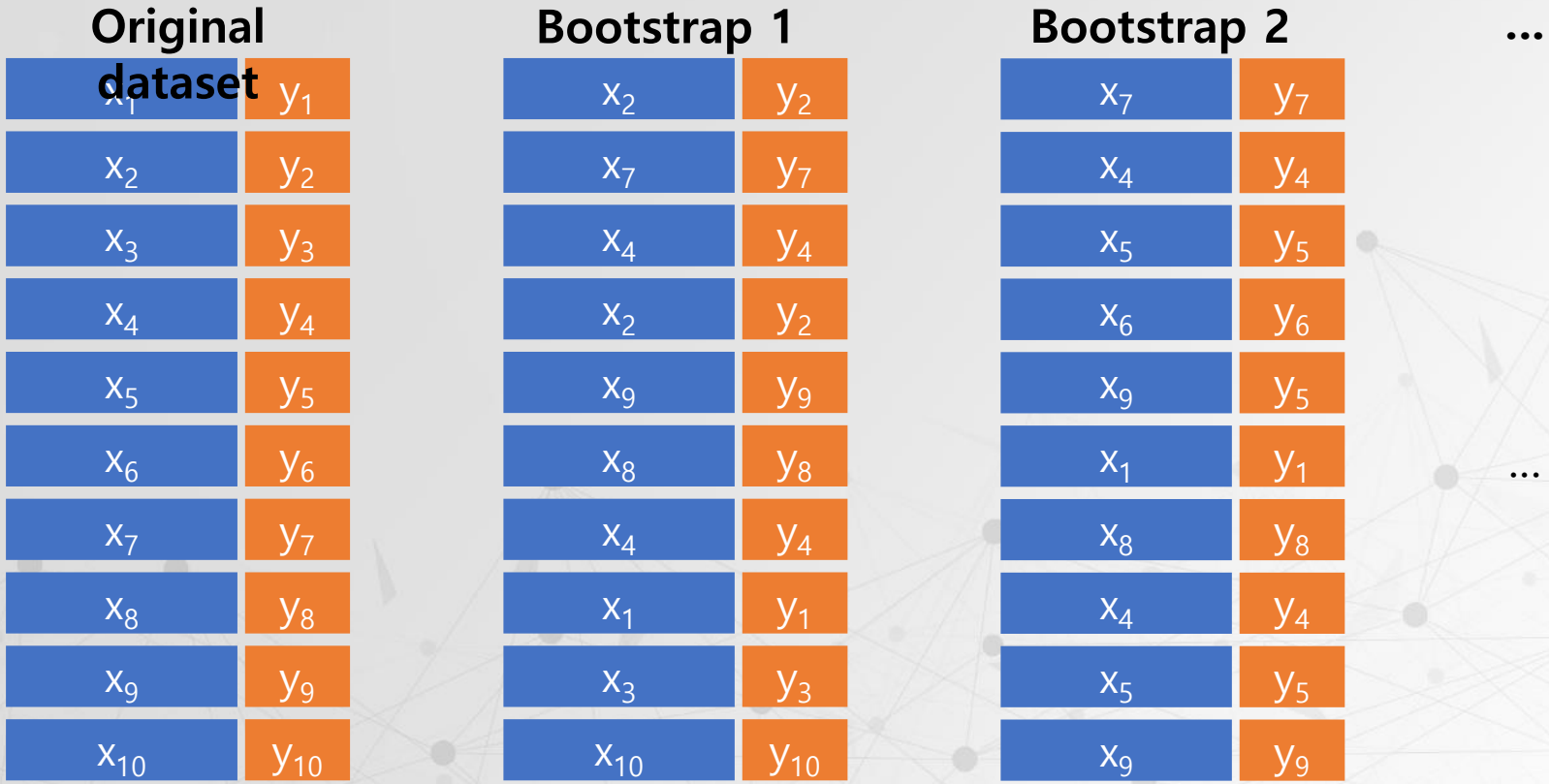
Bagging (bootstrap aggregating)

- Bootstrap 기법으로 다수의 학습 데이터 생성
- 생성된 데이터로 모델 구축
- 주어진 새 입력에 대해 예측을 종합



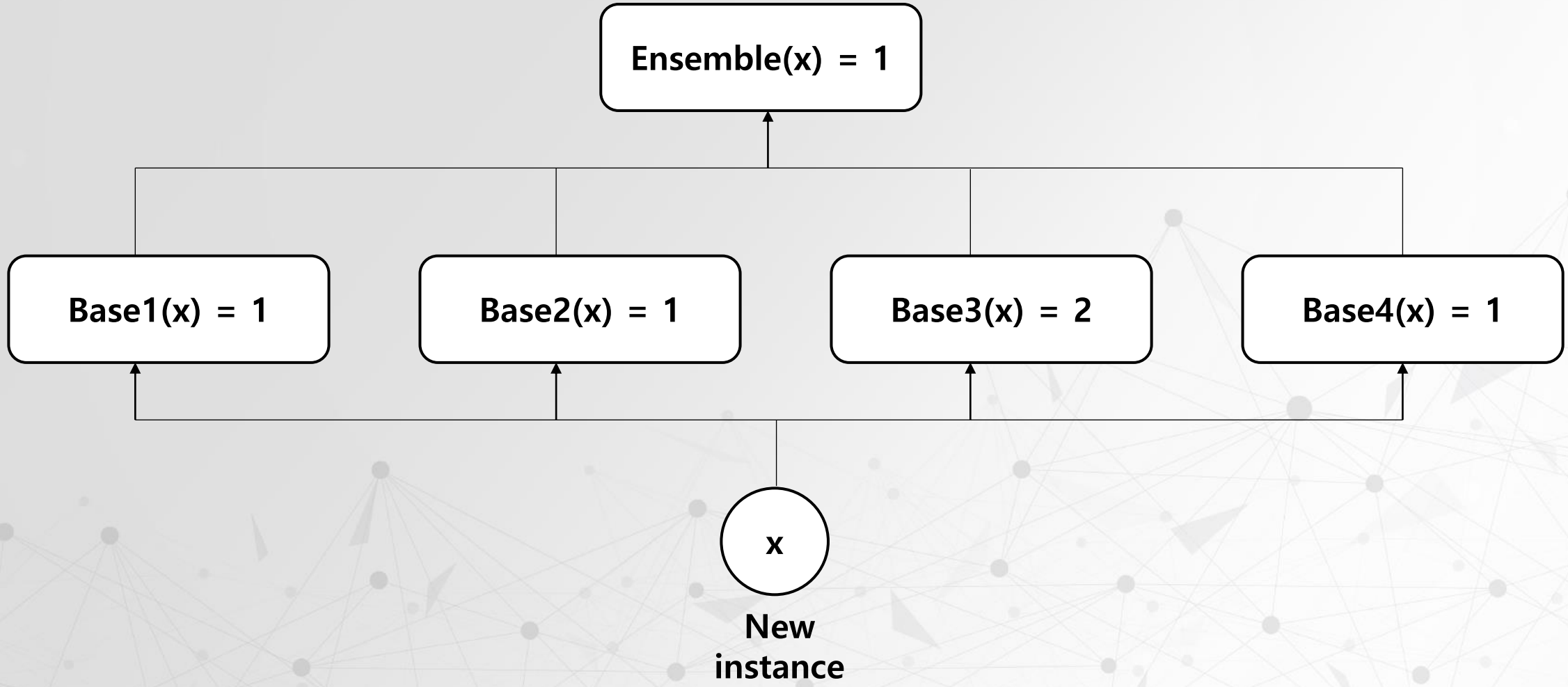
Bagging (bootstrap aggregating): bootstrapping

- 각 모델은 서로 다른 학습 데이터셋을 이용
- 학습 데이터셋은 복원 추출(sampling with replacement)을 통해 원래 데이터의 수만큼의 크기를 갖도록 샘플링
- 개별 데이터셋을 bootstrap set이라 부름



Bagging (bootstrap aggregating): result aggregating

- Hard voting, majority voting



Bagging (bootstrap aggregating): result aggregating

- Hard voting, majority voting

Ensemble population	Training accuracy	P(y=1)	Prediction
Model 1	0.80	0.90	1
Model 2	0.76	0.88	1
Model 3	0.82	0.37	0
Model 4	0.94	0.65	1
Model 5	0.83	0.75	1
Model 6	0.72	0.12	0
Model 7	0.85	0.86	1
Model 8	0.91	0.69	1
Model 9	0.77	0.71	1
Model 10	0.87	0.64	1

Hard voting = 1

Bagging (bootstrap aggregating): result aggregating

- Weighted voting: training accuracy를 weight로
  - P(Ensemble = 0) = (0.82 + 0.72) / (0.80 + 0.76 + 0.82 + … + 0.87)
  - P(Ensemble = 1) = (0.80 + 0.76 + 0.94 + … + 0.87) / (0.80 + 0.76 + … + 0.87)

Ensemble population	Training accuracy	P(y=1)	Prediction
Model 1	0.80	0.90	1
Model 2	0.76	0.88	1
Model 3	0.82	0.37	0
Model 4	0.94	0.65	1
Model 5	0.83	0.75	1
Model 6	0.72	0.12	0
Model 7	0.85	0.86	1
Model 8	0.91	0.69	1
Model 9	0.77	0.71	1
Model 10	0.87	0.64	1

P(Ensemble = 0) = 0.186
P(Ensemble = 1) = 0.814

Weighted voting = 1

Bagging (bootstrap aggregating): result aggregating

- Weighted voting: prediction probability를 weight로
  - P(Ensemble = 0) = (0.37 + 0.12) / (0.90 + 0.88 + 0.37 + … + 0.64)
  - P(Ensemble = 1) = (0.90 + 0.88 + 0.65 + … + 0.64) / (0.90 + 0.88 + … + 0.64)

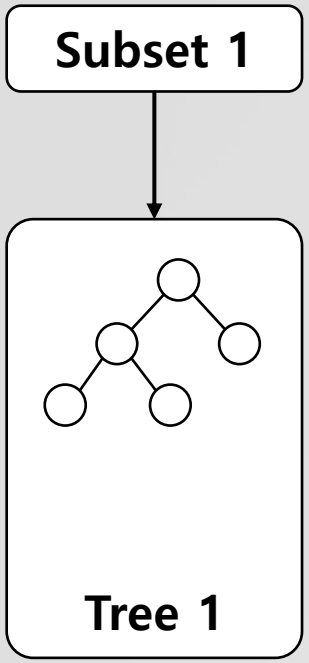
Ensemble population	Training accuracy	P(y=1)	Prediction
Model 1	0.80	0.90	1
Model 2	0.76	0.88	1
Model 3	0.82	0.37	0
Model 4	0.94	0.65	1
Model 5	0.83	0.75	1
Model 6	0.72	0.12	0
Model 7	0.85	0.86	1
Model 8	0.91	0.69	1
Model 9	0.77	0.71	1
Model 10	0.87	0.64	1

P(Ensemble = 0) = 0.075
P(Ensemble = 1) = 0.925

Weighted voting = 1

Random subspace

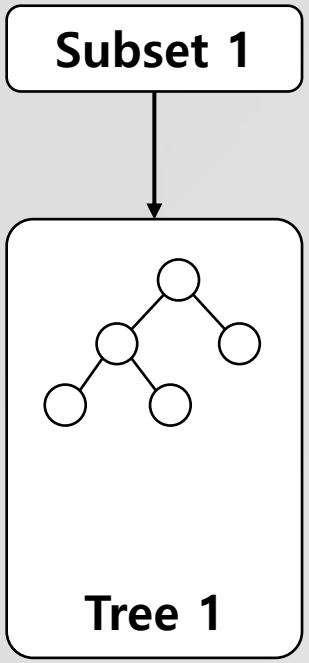
- 원래 변수들 중에서 모델 구축에 쓰일 입력 변수를 무작위로 선택



원래 변수	x1	x2	x3	x4	x5	x6	x7	x8
입력 변수	x1		x3	x4			x7	

Random subspace

- 선택된 입력 변수 중에 분할될 변수를 선택

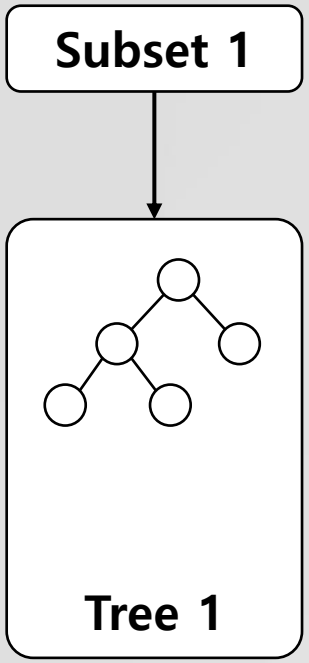


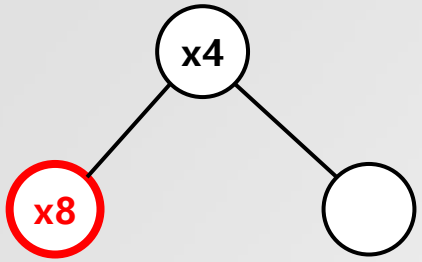
x4

원래 변수	x1	x2	x3	x4	x5	x6	x7	x8
입력 변수	x1		x3	x4			x7	

Random subspace

- 분할된 노드에서 동일한 과정을 반복





원래 변수	x1	**x2**	x3	**x4**	**x5**	x6	x7	**x8**
입력 변수		x2		x4	x5			**x8**

Generalization error

- 각각의 decision tree는 과적합 될 수 있지만
- Random forest는 그 수가 충분히 많을 때
  큰 수의 법칙에 의해 전체 에러는 바운드됨

$$e \leq \frac{\bar{\rho}(1-s^2)}{s^2}$$

  - $\bar{\rho}$: decision tree 사이의 평균 상관관계
  - s : 올바로 예측한 tree의 수와 잘못 예측한 tree의 수 차이의 평균
- 개별 tree의 정확도가 높을수록 s가 증가함
- Bagging과 random subspace로 모델 사이의 상관관계를 감소
- 개별 tree의 정확도가 높고, 각각의 독립성이 높을수록 전체 성능이 증가