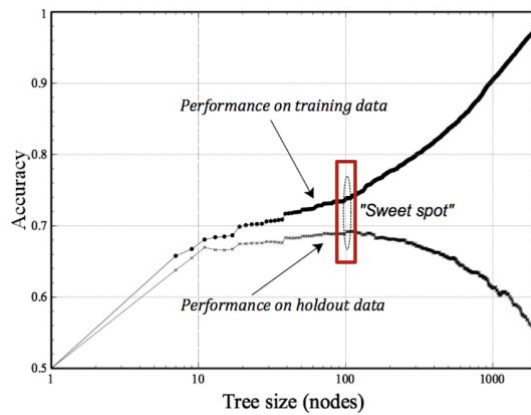


1.(사진은 예시)



이 그림에서 x축, y축에 들어갈 단어를 쓰고(지워진 채로 나옴),  
Sweet spot이 어딘지 그래프 위에 표시하시오

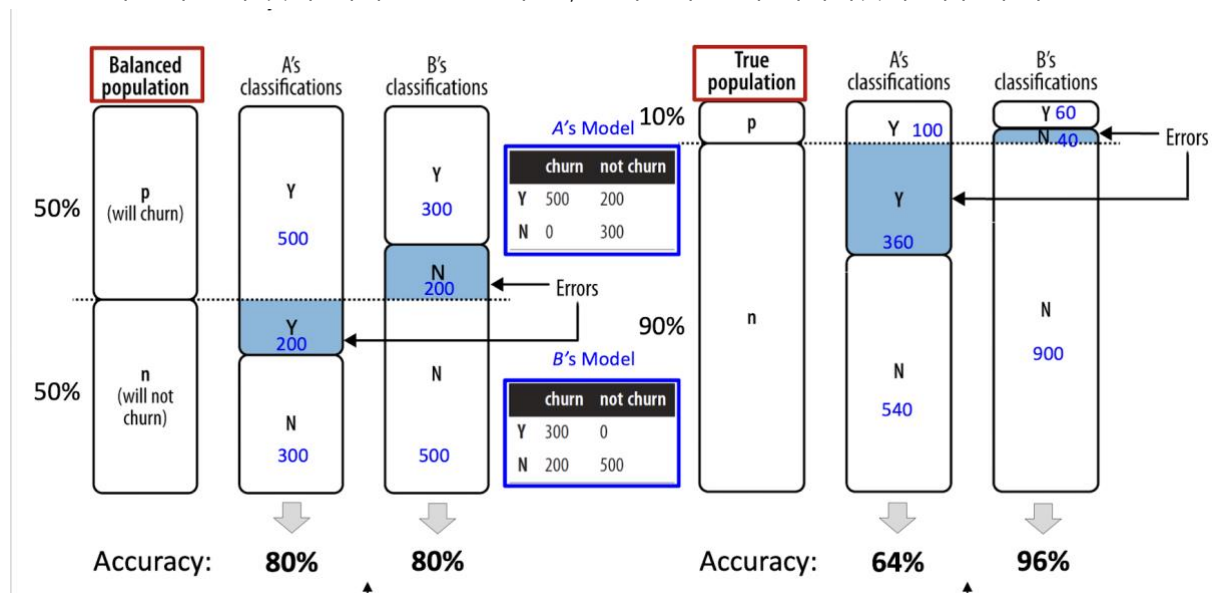
2. (사진은 예시) 단어들로 구성된 두개의 집합을 주고, Jaccard distance 구하도록 함

- $X = \{\text{"I", "am", "a", "boy"}\}$
- $Y = \{\text{"You", "are", "a", "boy"}\}$

3. (사진은 예시)두 점의 x,y,z 좌표 주고 코사인 거리 계산하도록 함

- $A = (7, 3, 2)$
- $B = (2, 3, 0)$

4. (사진은 예시) 데이터셋 하나에 대한 두가지 분류법의 결과 보여주는데 막대그래프에서 파란색으로 칠해진 부분의 숫자 지워놓고 쓰도록 함, 정확도와 전체 데이터셋의 개수 주어짐



5. (사진은 예시) 한 데이터셋에 대한 confusion matrix와 cost-benefit matrix를 주고 expected profit, accuracy, recall, precision을 계산하도록 함

	p	n
Y	56	7
N	5	42

	Actual	
	p	n
Predicted Y	99	-1
N	0	0

Confusion matrix

6. (사진은 예시) 다음과 같은 사진을 주고 총 4개의 threshold에 대한 confusion matrix를 그린 뒤 ROC 그래프를 그려 표시하도록 함

Instance description	True class	Score
.....	p	0.99
.....	p	0.98
.....	n	0.96
.....	n	0.90
.....	p	0.88
.....	n	0.87
.....	p	0.85
.....	p	0.80
.....	n	0.70
.....	p	0.65
.....	.	.
.....	.	.
.....	.	.

7. 전체 케이스의 개수를 주고, 랜덤으로 소팅했을 때의 lift와 75%를 정확하게 맞혔을 때의 lift를 구하도록 함

8. (사진은 예시) 다음과 같은 표를 주고, 한 줄의 flu 여부를 지운 뒤 naive bayes classifier를 이용해 flu가 Y일지 N일지 구하도록 함

Chills	Runny nose	Headache	Fever	Flu?
Y	N	Mild	Y	N
Y	Y	No	N	Y
Y	N	Strong	Y	Y
N	Y	Mild	Y	Y
N	N	No	N	N
N	Y	Strong	Y	Y
N	Y	Strong	N	N
Y	Y	Mild	Y	Y

9. 수업시간에 배웠던 "당뇨병 환자의 기대수명 예측" 선형회귀 사이킷 런 코드가 나와있음, 코드의 결과값을 쓰도록 함(넘파이 슬라이싱 등을 이용해 알아낼 수 있는 코드)

10. knn을 이용한 사이킷 런 코드가 나와 있고, index, pred, dist 변수값이 얼마로 출력되는지 작성하도록 함