1906003132015

Doğal Dil İşleme

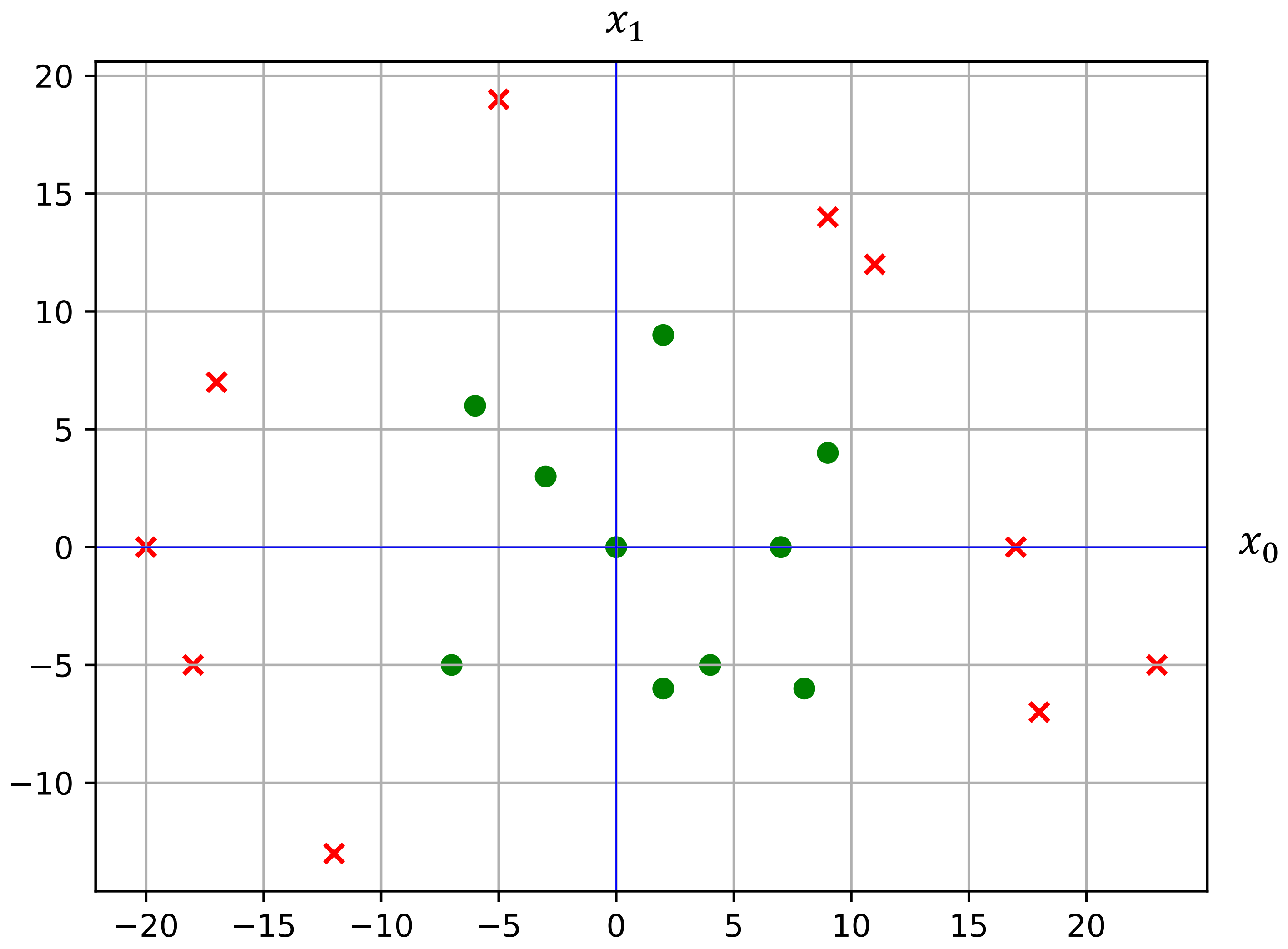
BAİBÜ Bilgisayar Müh.

Dr. Öğr. Üyesi İsmail Hakkı Parlak

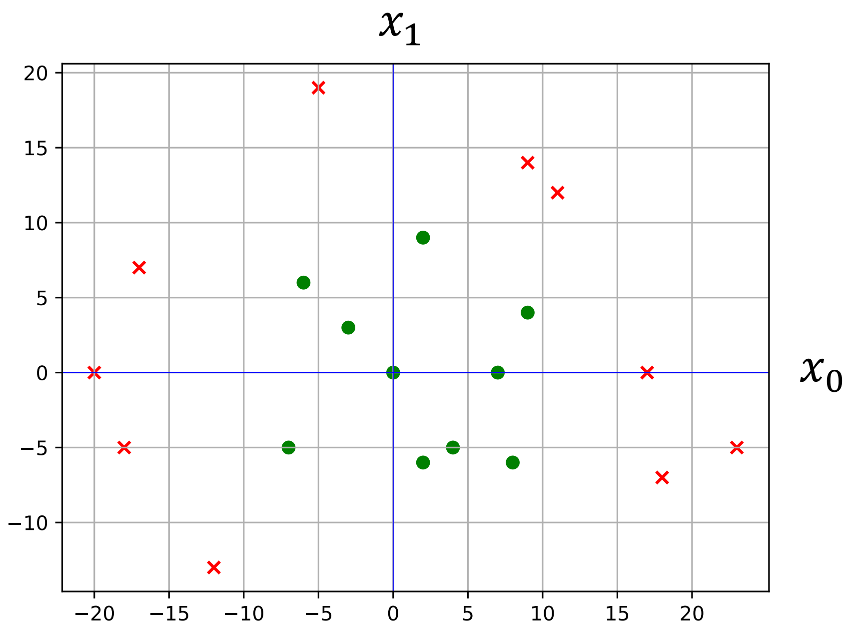
ismail.parlak@ibu.edu.tr

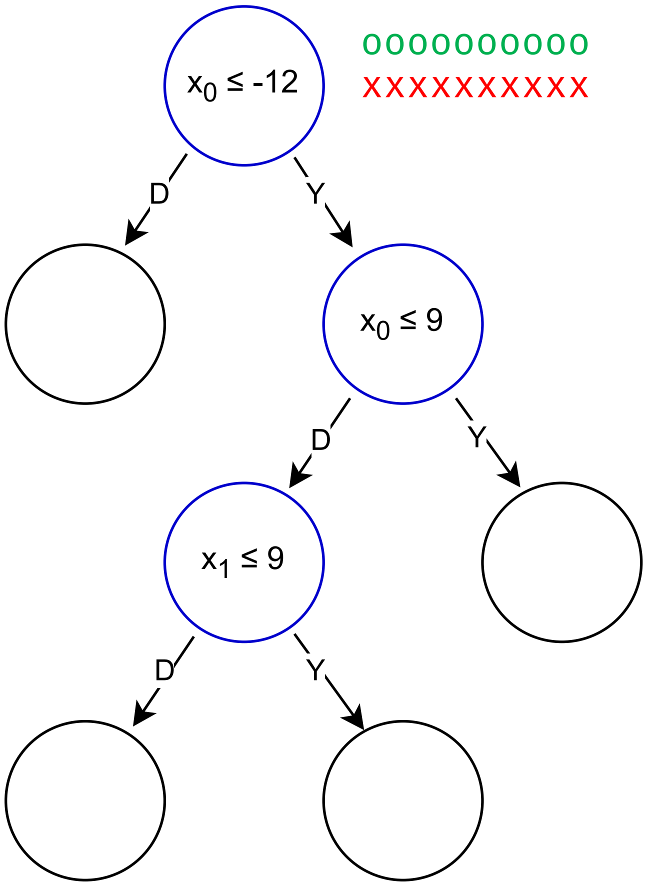
Oda: 335

Karar Ağacı (Decision Tree)

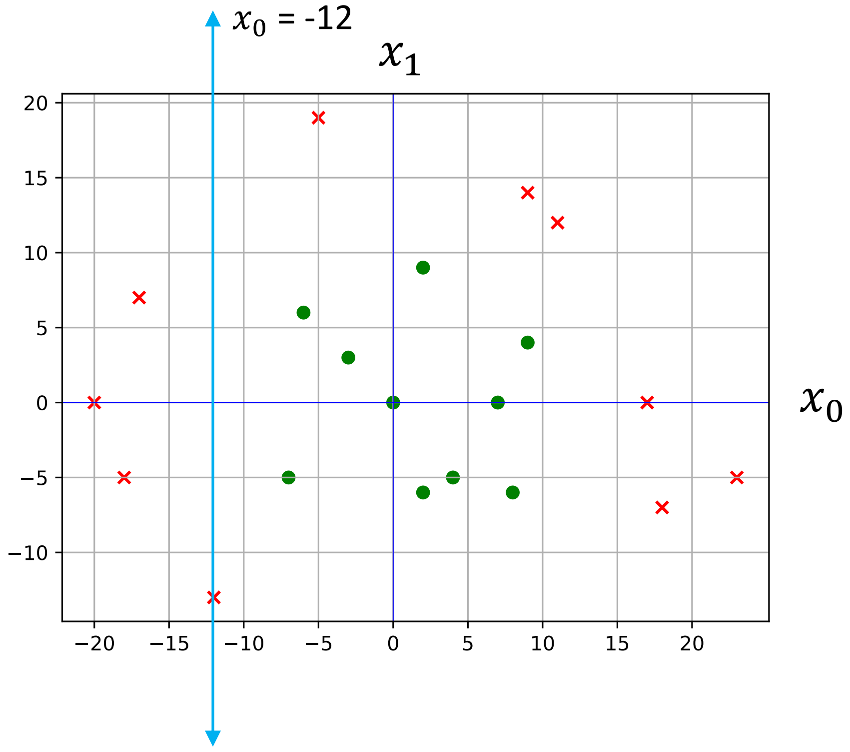


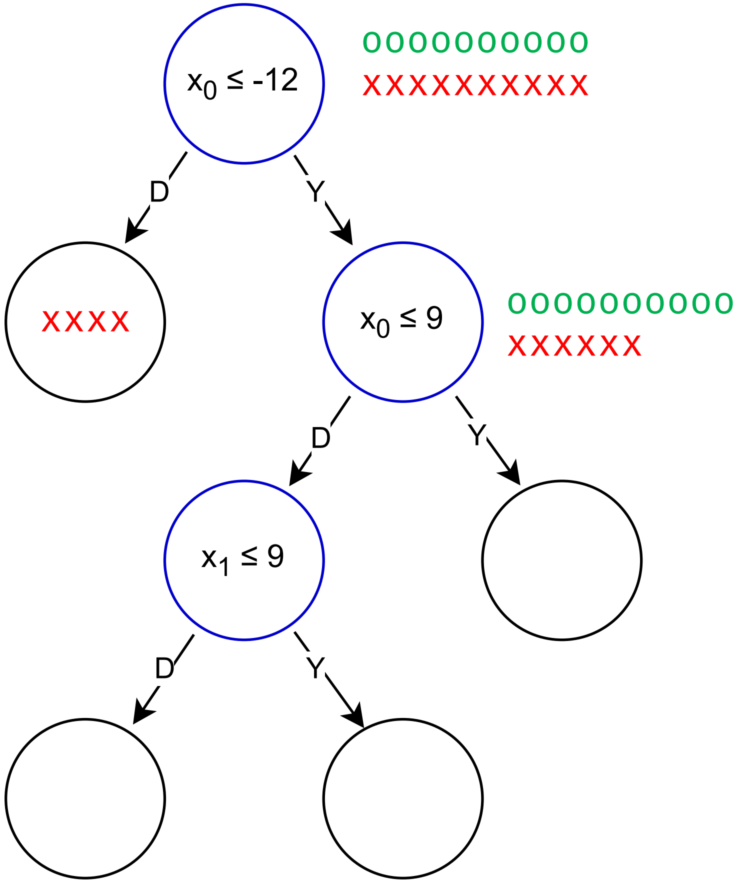
Karar Ağacı (Decision Tree)



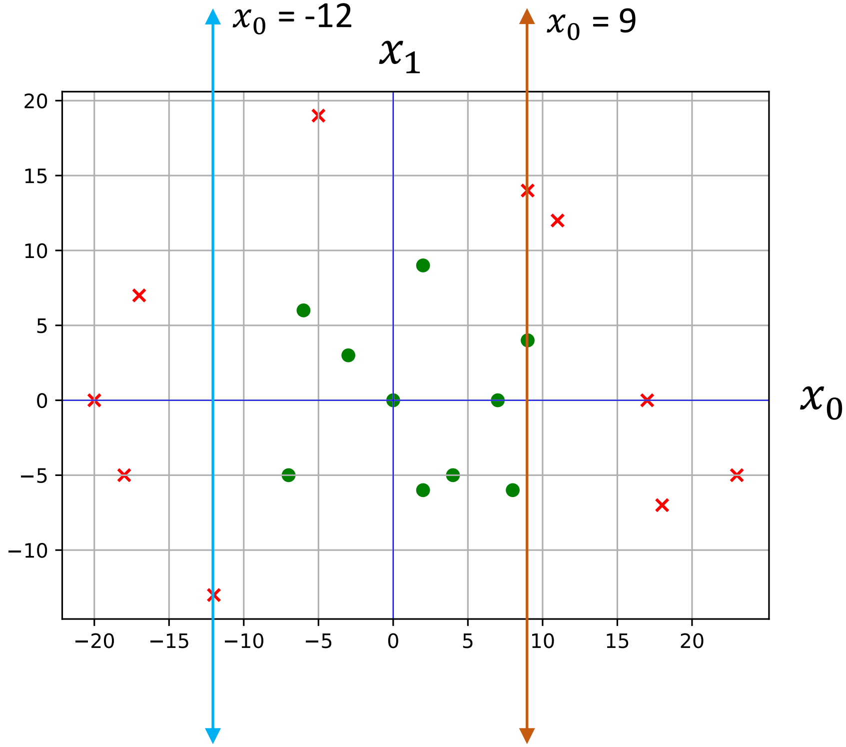


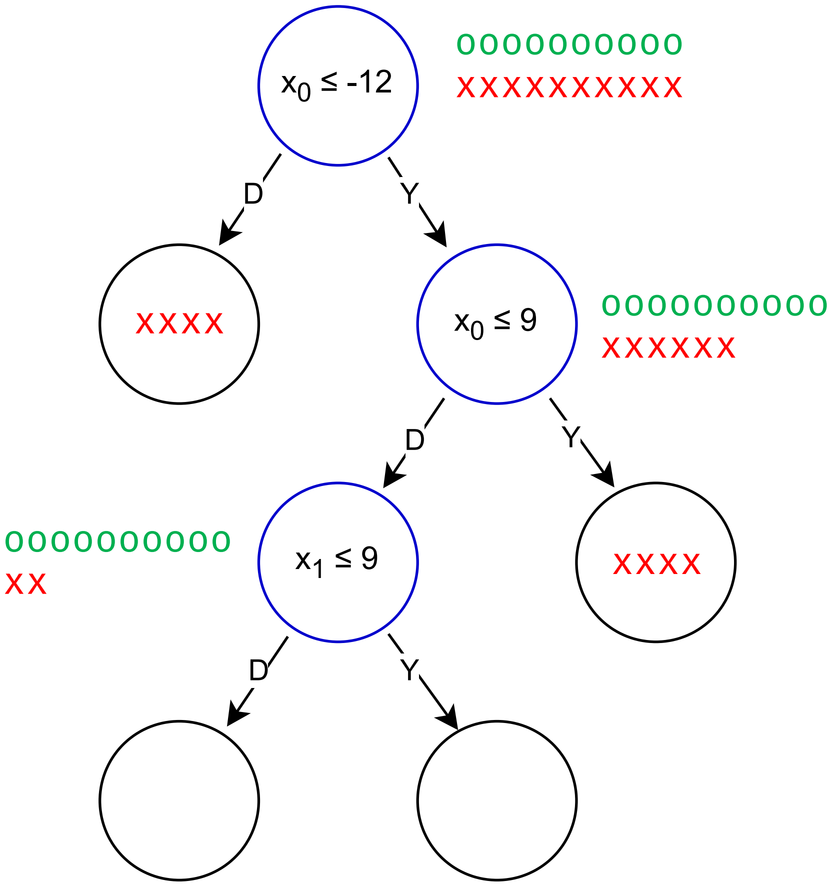
Karar Ağacı (Decision Tree)



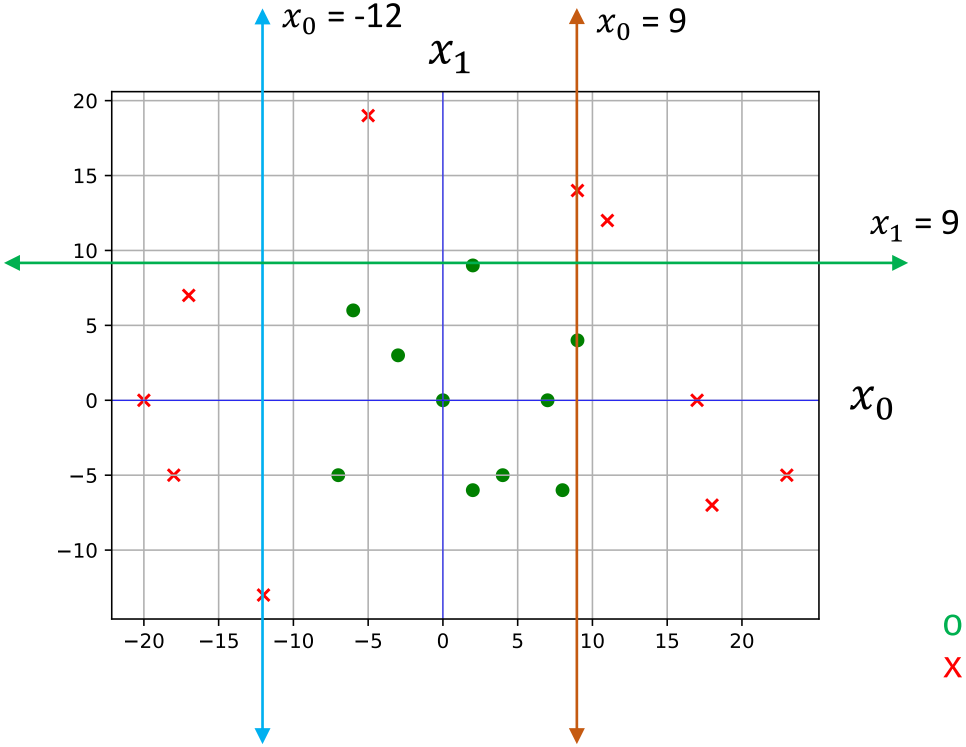


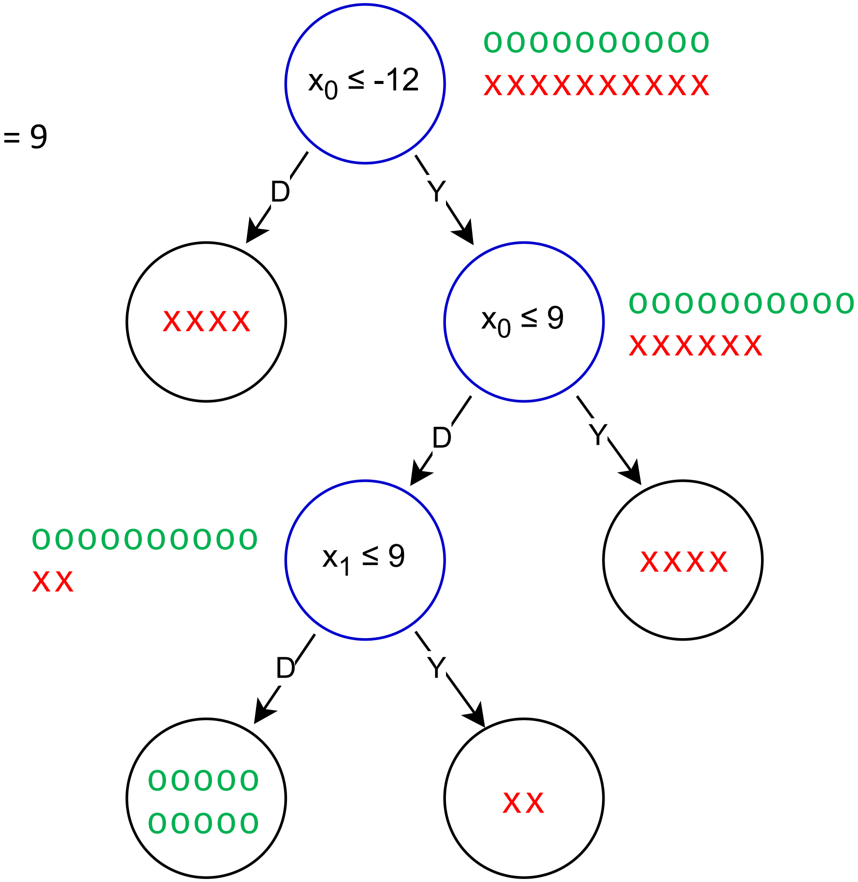
Karar Ağacı (Decision Tree)





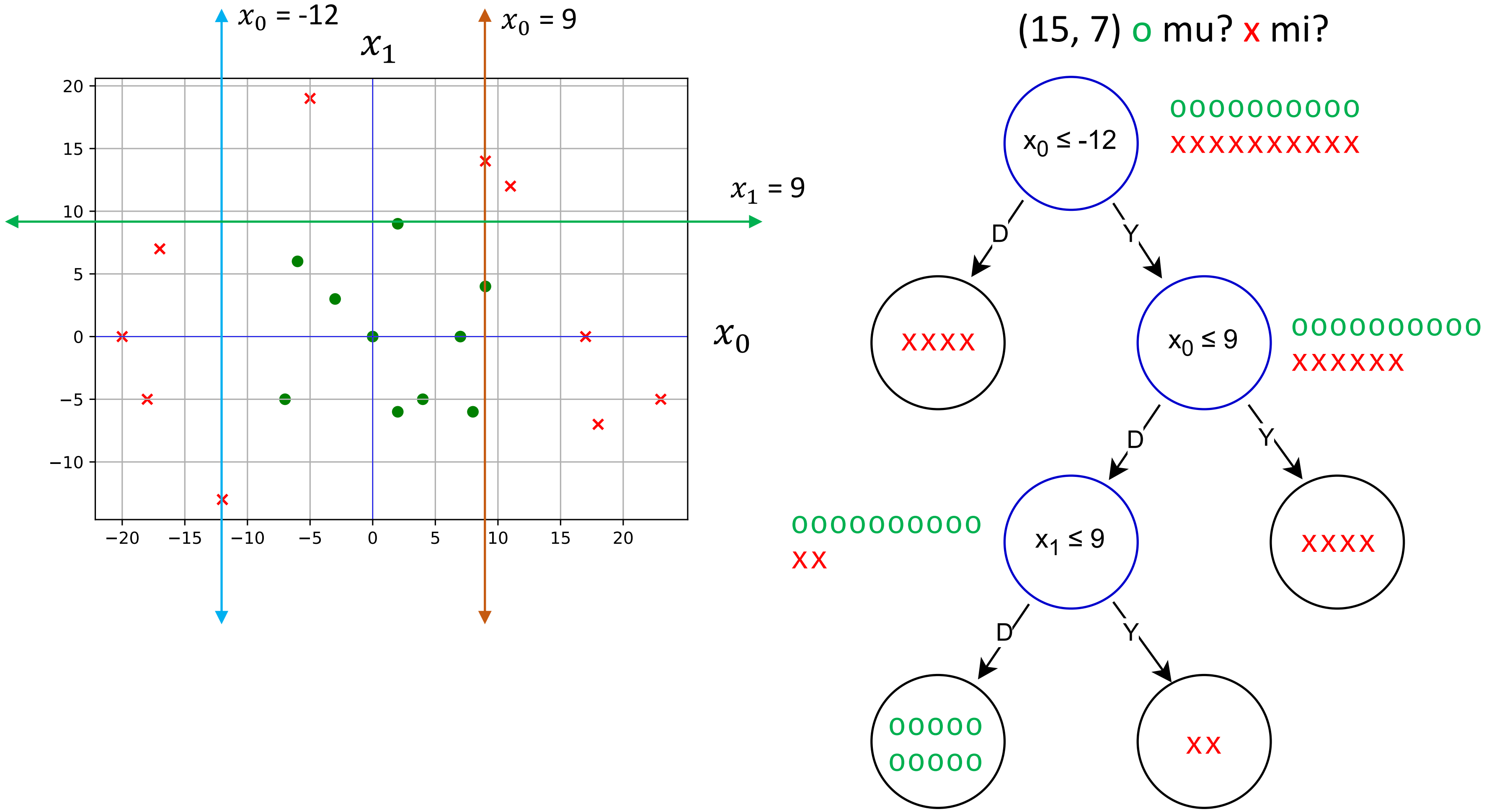
Karar Ağacı (Decision Tree)



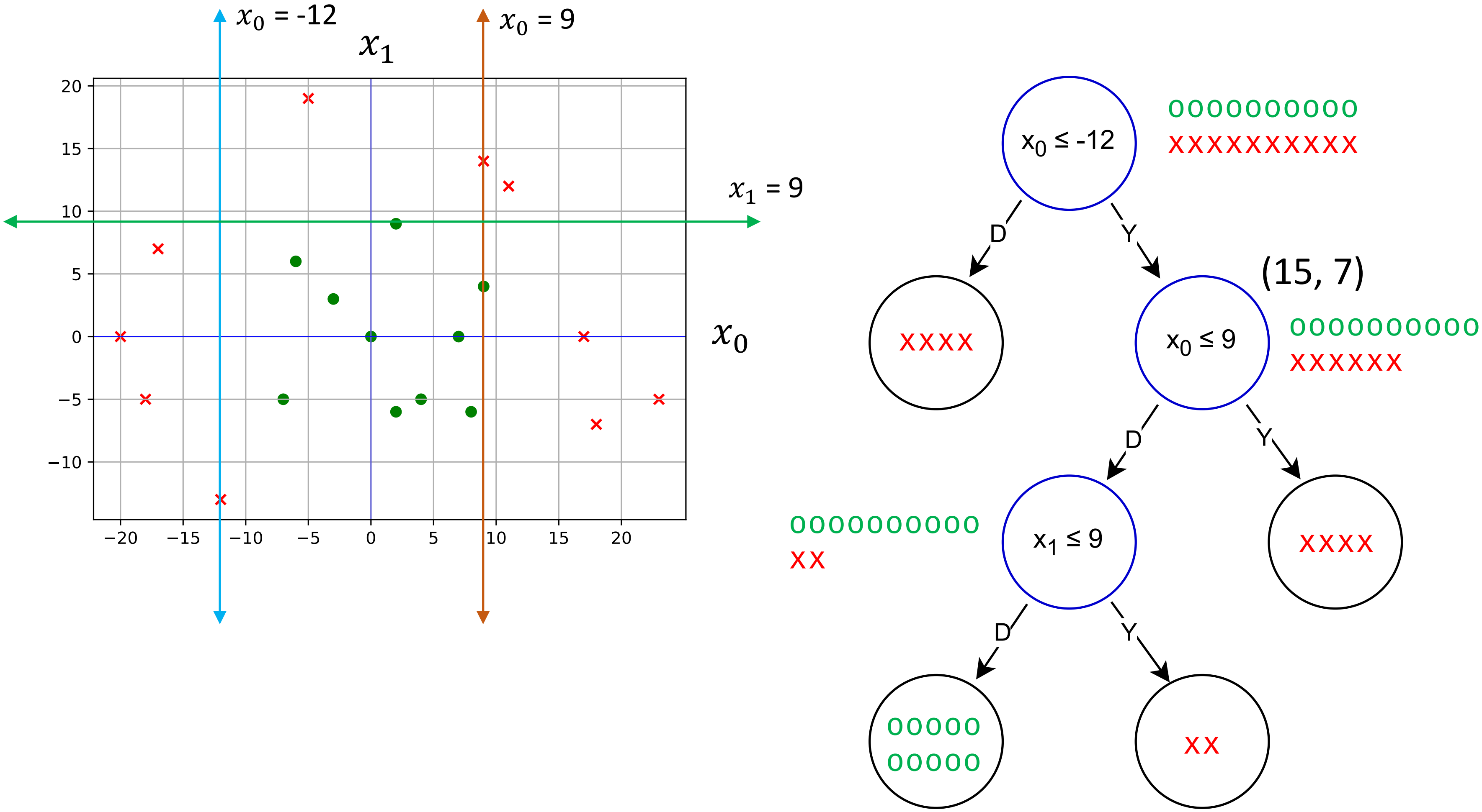


Saf düğümler (pure nodes) elde edilene kadar örnekler ayrıştırılır.

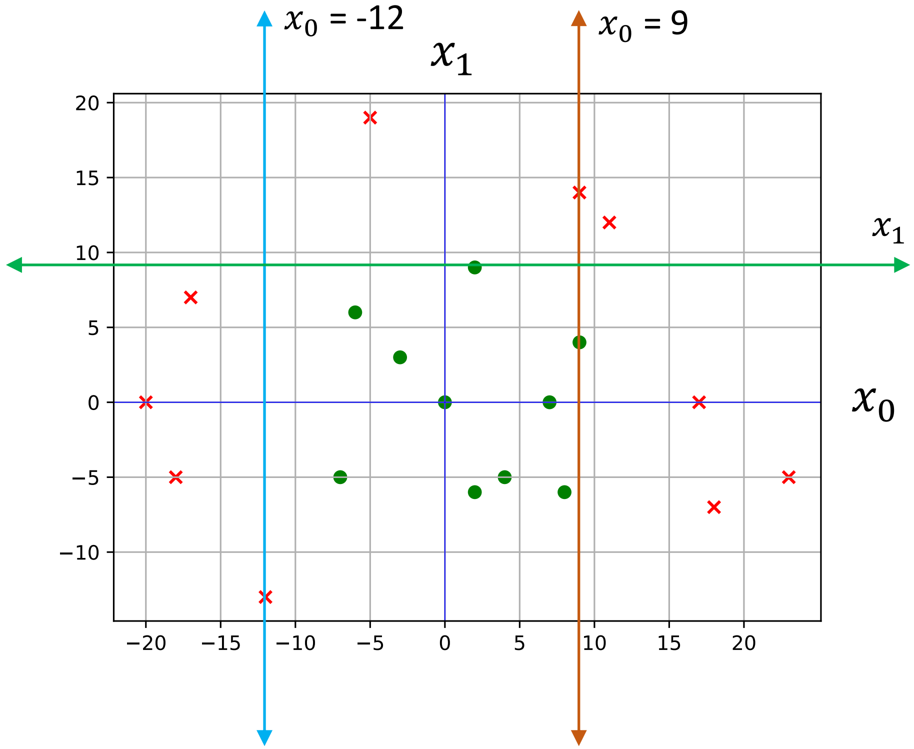
Karar Ağacı (Decision Tree)

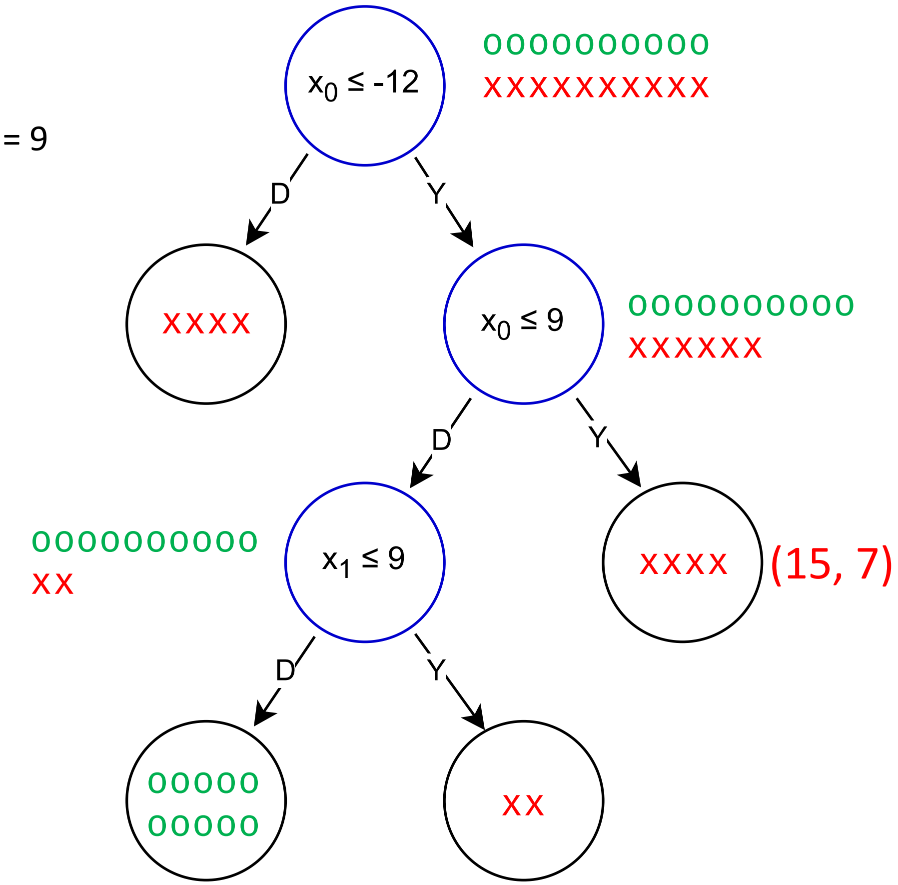


Karar Ağacı (Decision Tree)



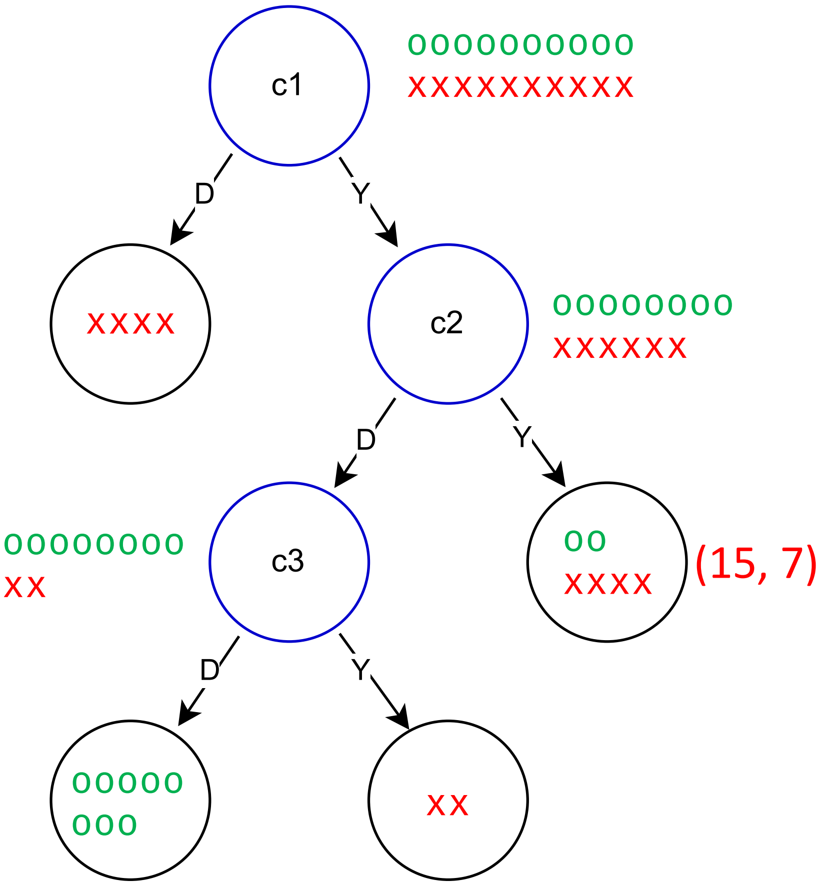
Karar Ağacı (Decision Tree)





Karar Ağacı (Decision Tree)

Eğer düğümler saf değilse çoğunluk oylaması yapılır.



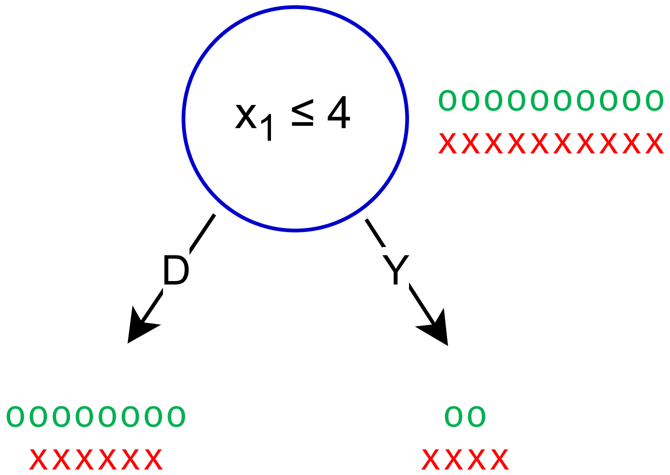
Karar Ağacı (Decision Tree)

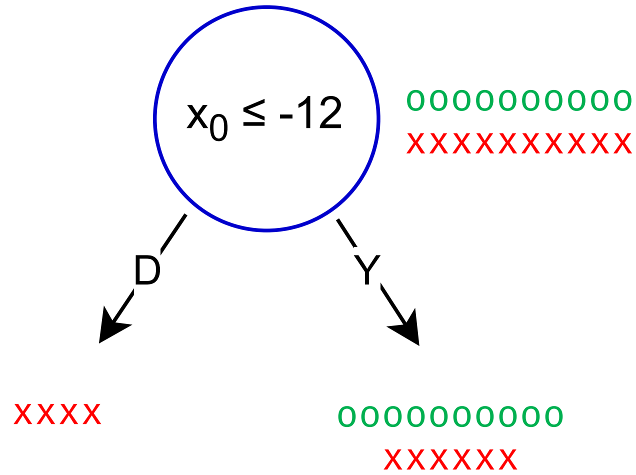
- Bir veri setinde karar ağaçları ile sınıflandırma (classification) yaparken birden çok çözüm olabilir.
- Diğer bir deyişle, sınıfları uygun şekilde birbirinden ayırabilen birbirinden farklı, birden çok karar ağacı bulunabilir.
- Karar ağaçları karar düğümlerinde hangi koşulu kullanması gerektiğini nasıl hesaplar?

Karar Ağacı (Decision Tree)

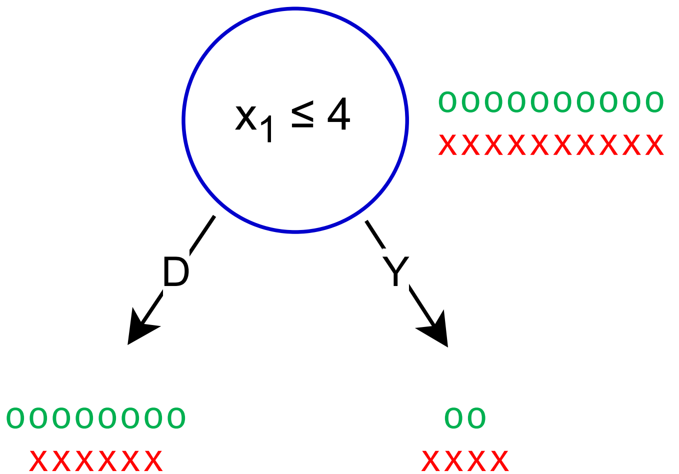
- Karar ağaçları karar düğümlerinde hangi koşulu kullanması gerektiğini nasıl hesaplar?
- Veri setinden rastgele 2 örnek seçilir ve bu örneklerin nitelikleri (features) karar mekanizmalarında kullanılarak bilgi kazanımları (information gain) hesaplanır.
- Bilgi kazanımı yüksek olan karar mekanizması kullanılarak işlem olabildiğince saf düğümler elde edilene kadar tekrarlanır.

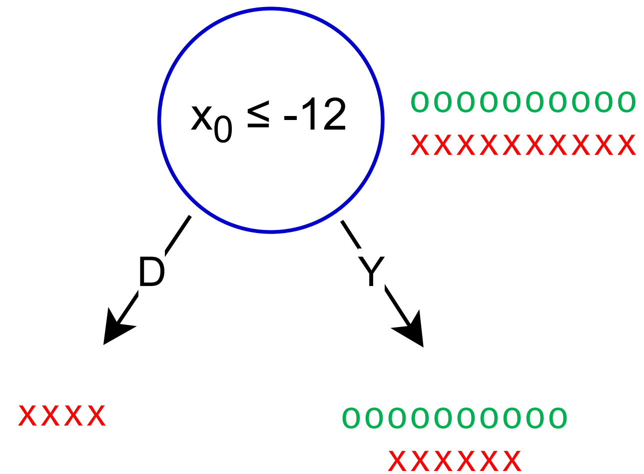
Karar Ağacı (Decision Tree)





Karar Ağacı (Decision Tree)



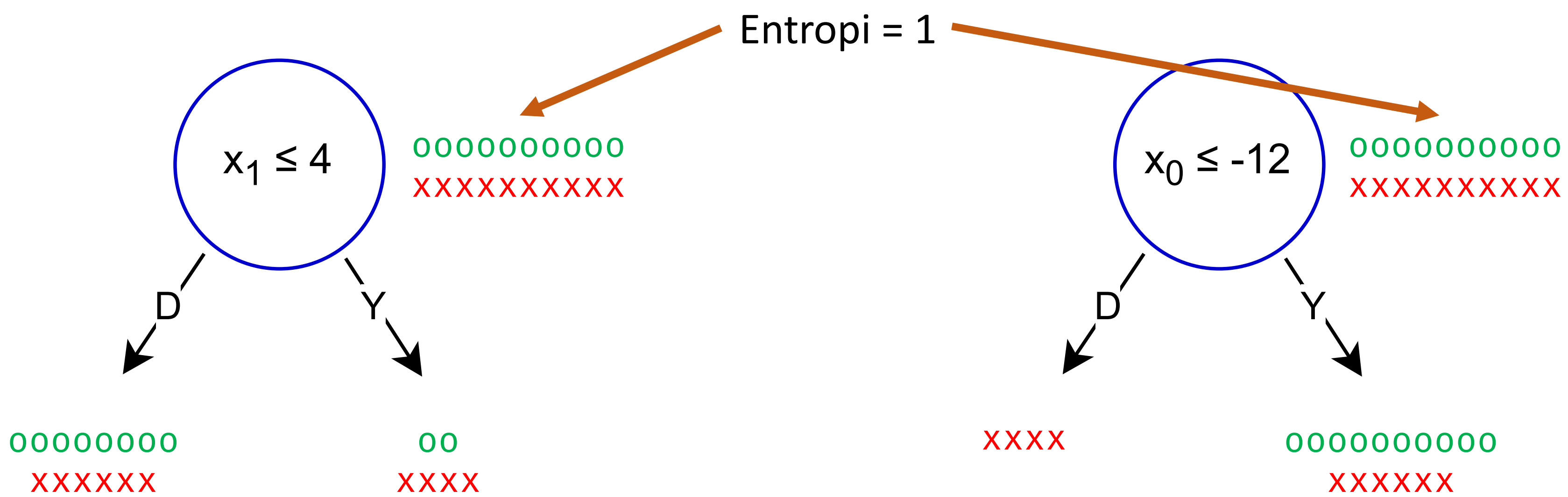


Entropi = $\sum -p_i \log(p_i)$, p_i: p sınıfının olasılığı

oooooooooo
XXXXXXXXXX

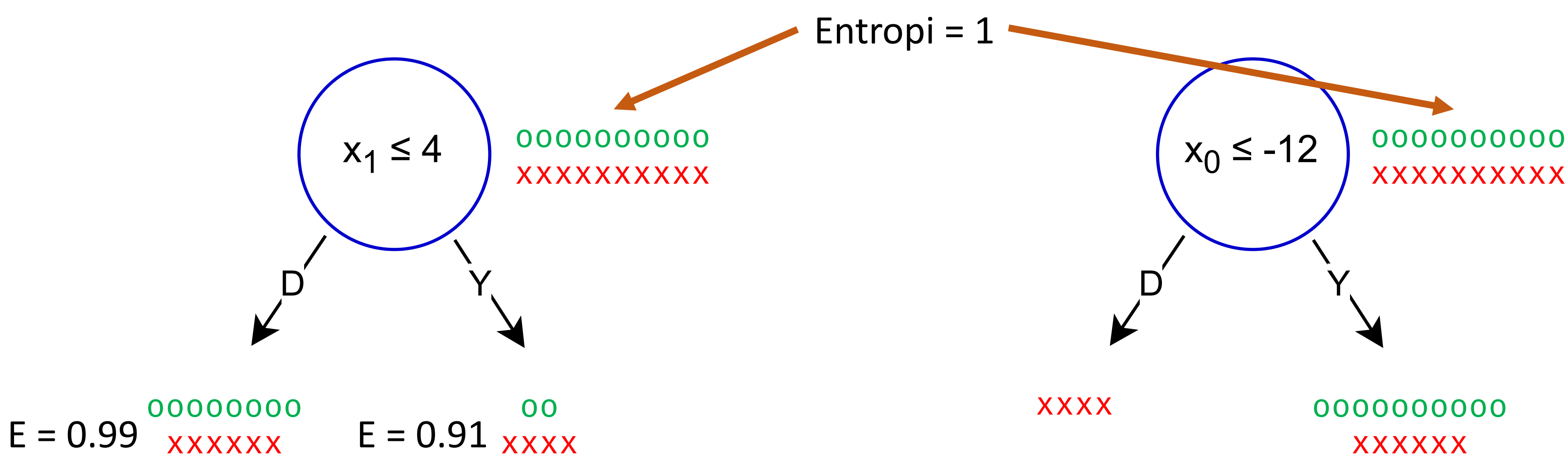
Entropi (E) = (-0.5 x log(0.5)) + (-0.5 x log(0.5)) = -0.5 x -1 + -0.5 x -1 = 1

Karar Ağacı (Decision Tree)



BK (IG) = $E(anne) - \sum w_i \, E(çocuk)$, w_i = çocuğun anneye oranlı büyüklüğü

Karar Ağacı (Decision Tree)



BK (IG) = $E(anne) - \sum w_i \, E(çocuk)$, w_i = çocuğun anneye oranlı büyüklüğü

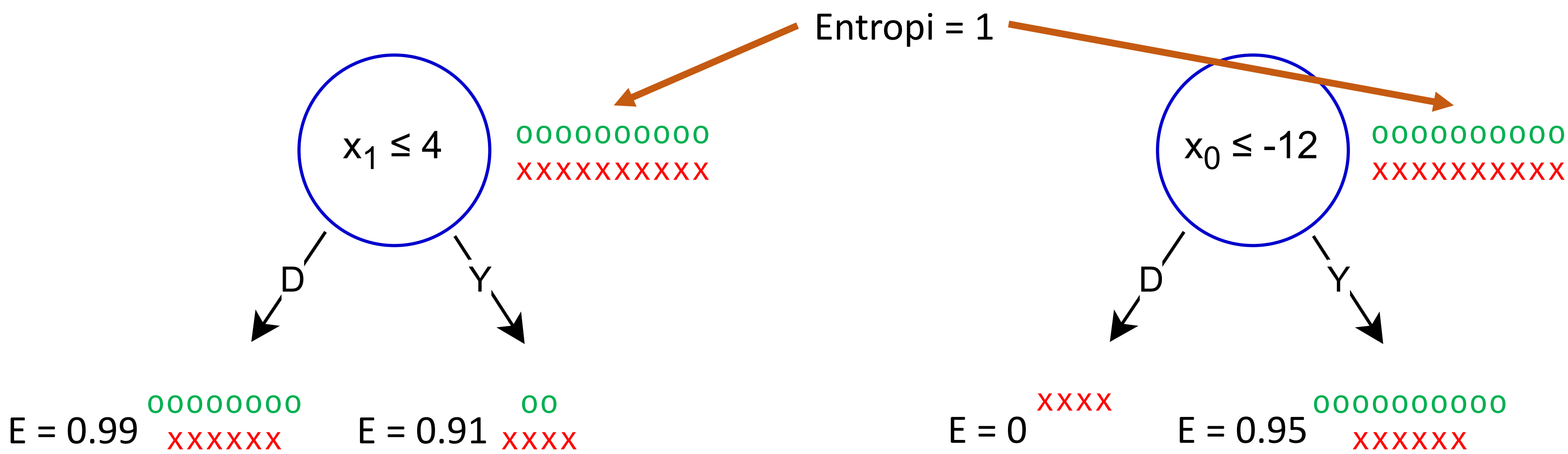
oooooooo
xxxxxx

Entropi = $\frac{-8}{14}\log\left(\frac{8}{14}\right) + \frac{-6}{14}\log\left(\frac{6}{14}\right) = 0.99$

oo
xxxx

Entropi = $\frac{-2}{6}\log\left(\frac{2}{6}\right) + \frac{-4}{6}\log\left(\frac{4}{6}\right) = 0.91$

Karar Ağacı (Decision Tree)



BK (IG) = $E(anne) - \sum w_i\, E(çocuk)$, w_i = çocuğun anneye oranlı büyüklüğü

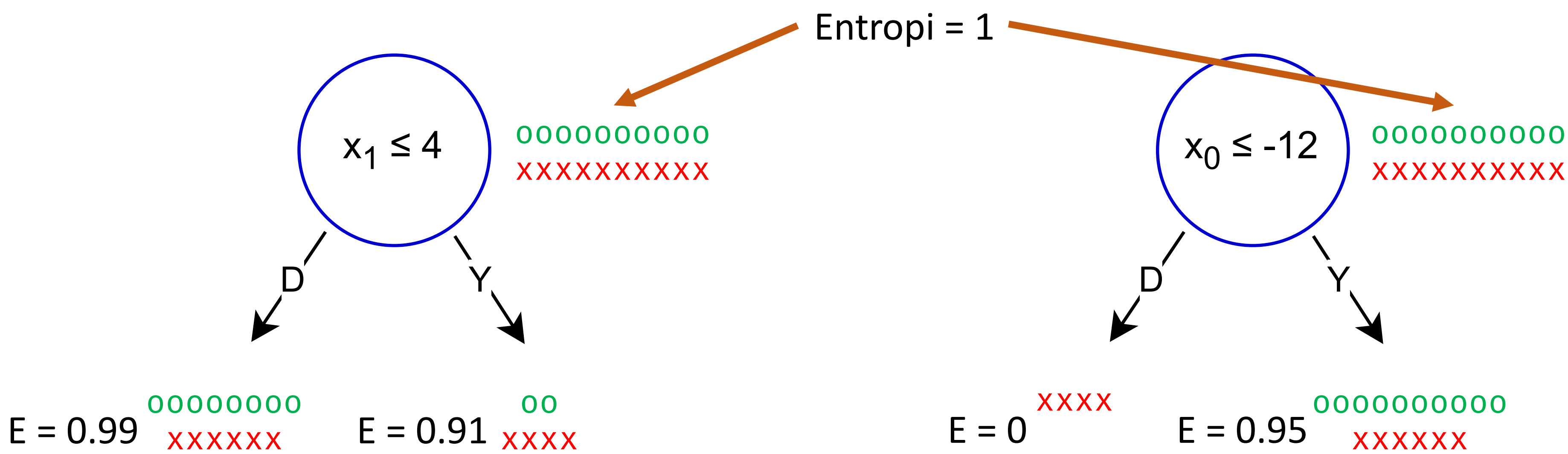
xxxx

Entropi = $-1\log(1) + 0\log(0) = 0$

oooooooooo
xxxxxx

Entropi = $\frac{-10}{16}\log\left(\frac{10}{16}\right) + \frac{-6}{16}\log\left(\frac{6}{16}\right) = 0.95$

Karar Ağacı (Decision Tree)



$$BK_1 = 1 - \left(\frac{14}{20}\,0.99 + \frac{6}{20}\,0.91\right) = 0.034$$

$$BK_2 = 1 - \left(\frac{4}{20}\,0 + \frac{16}{20}\,0.95\right) = 0.24$$

BK_2, BK_1'den büyük olduğu için sağ taraftaki karar mekanizmasını kullanmak sol taraftakinden daha faydalıdır diyoruz.