

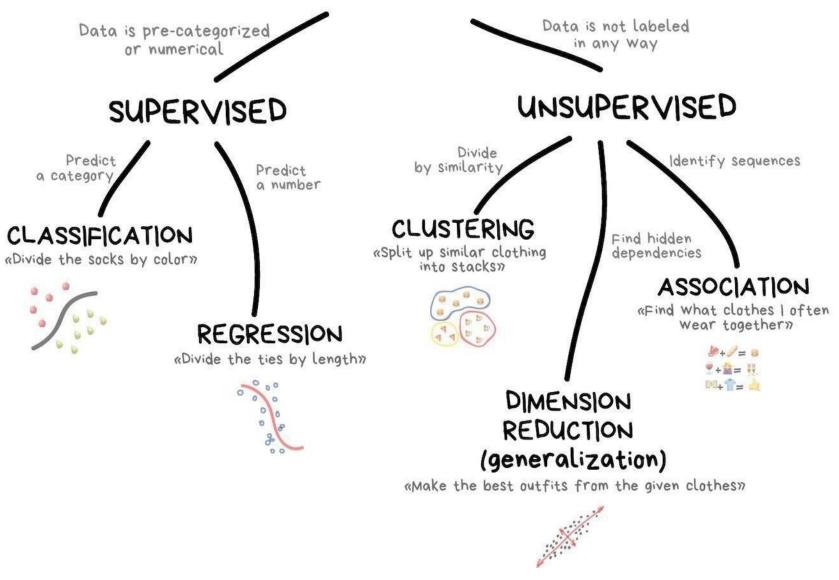
Makine Öğrenmesi ve İmge Tanıma

Karar Ağaçları ve Birleştirme Modelleri

İKİNCİ HAFTANIN KAYNAK KİTABI

• Sullivan, William. "Machine Learning For Beginners: Algorithms, Decision Tree & Random Forest Introduction, CreateSpace Independent Publishing Platform (August 20, 2017).

CLASSICAL MACHINE LEARNING



THE MAIN TYPES OF MACHINE LEARNING When quality is Simple data Complicated data a real problem Unclear features Clear features Belief in a miracle ENSEMBLES CLASSICAL ML NEURAL NETWORKS No data, AND but We have an environment DEEP LEARNING to interact With REINFORCEMENT LEARNING

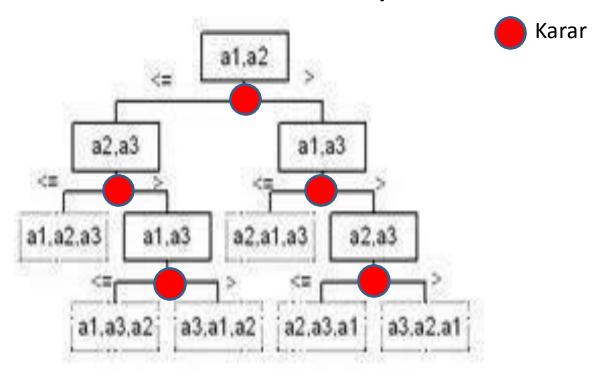


a1,a2,a3 sıralamak istiyoruz.



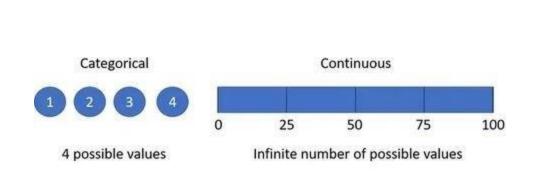


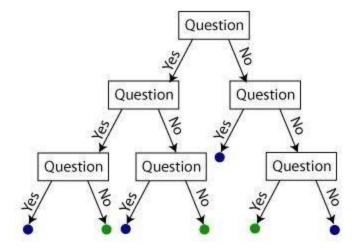
a1,a2,a3 sıralamak istiyoruz.





- Bir Karar Ağacı(decision tree), <u>denetimli ve</u> <u>sınıflandırma problemleri</u>nde yaygın olarak kullanılan **önceden tanımlanmış bir hedef değişken**e sahip olan bir öğrenme algoritması türüdür.
- Karar ağaçları hem **sürekli(!)** hem de **kategorik** girdi çıktı değişkenlerinde kullanılabilir.





Guess the Animal

Boolean Programming Activity





FOX



PANDA



COW







MOUSE TIGER



HAMSTER





CAT



FROG

RABBIT



MONKEY



RACOON



SHEEP





DONKEY



HORSE



CHICKEN



KOALA



OWL

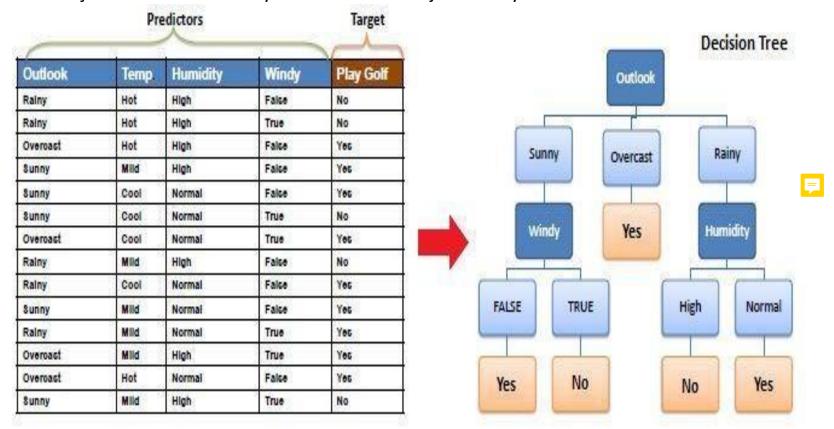




 Bu yöntemde, örnek(instance/sample), girdi değişkenlerindeki en önemli ayırıcı veya farklılaştırıcı tarafından iki veya daha fazla homojen kümeye bölünür.

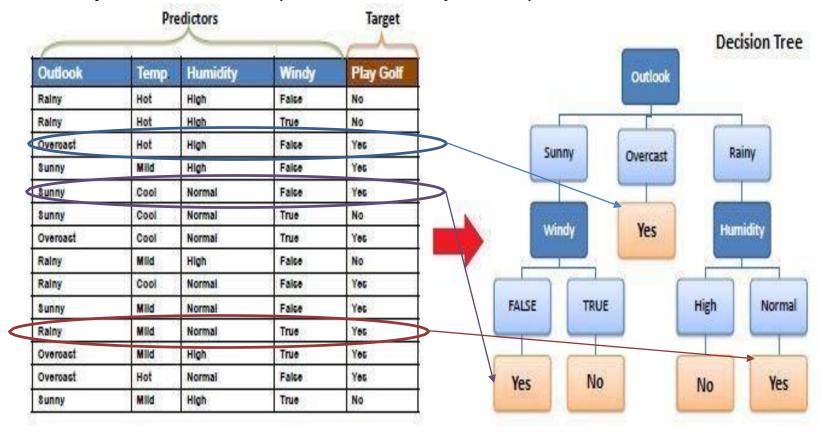


• Bu yöntemde, örnek(instance/sample), girdi değişkenlerindeki en önemli ayırıcı veya farklılaştırıcı tarafından iki veya daha fazla homojen kümeye bölünür.





• Bu yöntemde, örnek(instance/sample), girdi değişkenlerindeki en önemli ayırıcı veya farklılaştırıcı tarafından iki veya daha fazla homojen kümeye bölünür.



KARAR AĞAÇLARI – TERMİNOLOJİ

• Kök Düğüm(root node):

örneklerin hepsi veya tüm popülasyon anlamına gelir. Daha sonra iki veya daha fazla homojen kümeye ayrılır.

• Karar Düğümü(decision node):

Alt düğümlerin daha fazla alt düğüme bölünmesi. , bu süreçte oluşturulan düğüm karar düğümü olarak bilinir.

Yaprak /Terminal Düğümü (Leaf/ Terminal Node):

Bölünemeyen düğümler Yanrak yeva Terminal düğümü

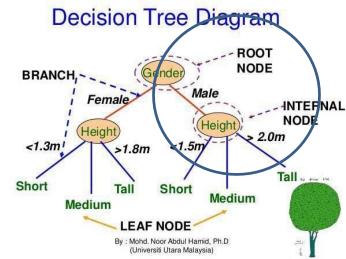
Bölünemeyen düğümler Yaprak veya Terminal düğümü olarak bilinir.

Dal / Alt Ağaç(Branch / Sub-Tree):

Adından da anlaşılacağı gibi, bir ağacın alt bölümü

• Üst ve Alt Düğüm(ebeveyn/çocuk):

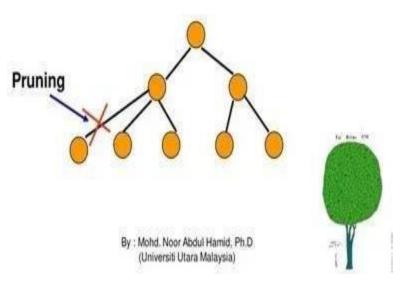
Parent and Child Node: Düğümlerin bölünmesinde, bölünen düğüme üst ebeveyn düğüm denir, bu bölümleme nedeniyle oluşan düğümlere, yani alt düğümlere alt çocuk düğüm denir.



KARAR AĞAÇLARI – TERMİNOLOJİ

- Bölümleme (splitting): Bu, düğümleri iki veya daha fazla alt düğüme sınıflandırmak için kullanılan işlemdir.
- Budama (pruning): Ağaç metaforundan devamla, karar düğümünün alt düğümlerinin kaldırılması işlemi budama olarak bilinir. Dolayısıyla bölünmeyi önleyici bir süreçtir.







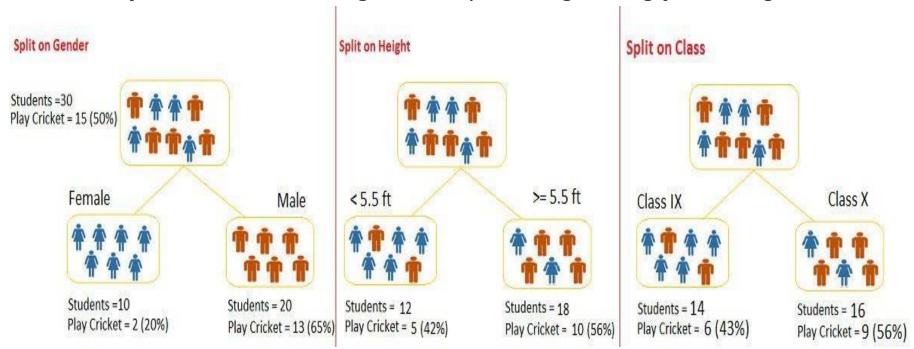
- ÖRNEK:
- Üç farklı özellikle tanımlanan 30 çocuğa bakalım.
 - Cinsiyet (E veya K),
 - Sınıf(IX. veya X.)
 - Boy (1.50 1.80 cm).



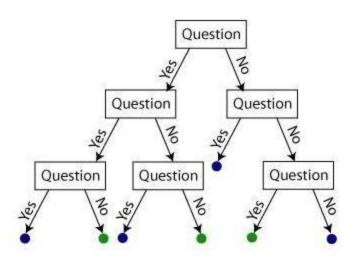
- Bu 30 kişiden 15'i boş zamanlarında kriket* oynuyor.
- Boş zamanlarında kriket oynayanları tahmin edecek bir model oluşturmak istiyoruz.
- Böyle bir senaryoda bir karar ağacı "üç değişkenin tümünü kullanarak çocukları ayırıp sınıflandırdığı ve birbirleriyle heterojen olacak en homojen öğrenci gruplarını oluşturacağı" için çok elverişlidir.
- Bu problem için çocukları kategorize etmek ve kriket oynayan çocukları bir sınıfta toplamak gerekecek..
- Peki üç öznitelikten en belirgin/belirleyici olan girdi değişkeni hangisi?



- Üç farklı özellikle tanımlanan 30 çocuğa bakalım.
 - Cinsiyet (E veya K),
 - Sınıf(IX. veya X.)
 - Boy (1.50 1.80 cm).
- Peki üç öznitelikten en belirgin/belirleyici olan girdi değişkeni hangisi?



BÖLÜMLEME KAR



BAŞARILI - BAŞARISIZ

Araştırma: Gelir Dağılımındaki Eşitsizlik Değerleri (Gini Katsayıları)

1. Gini Endeksi

Gini: «Bir popülasyondan rastgele iki öğe seçildiğinde, aynı sınıfta olmaları gerekir»
Bu durumda popülasyonlardan birinin saf olması olasılığı yüksektir.

"Başarı" (success) veya "Başarısızlık" (failure) gibi kategorik hedef değişkenle çalışır.

Yalnızca İkili bölümleme yapabilir.

Gini değeri arttıkça, homojenlik değeri de artar.

Gini yöntemi, CART (Sınıflandırma ve Regresyon Ağacı) tarafından ikili bölünmeler yapmak için kullanılır.



BÖLÜMLEME KARARI



1. Gini Endeksi

Başarı ve başarısızlık için olasılık karesinin toplamını hesapla, yani (p²+q²) Şimdi oluşturulan bölümün her bir düğümünün ağırlıklı Gini bölünümü için Gini'yi hesaplayın.



BÖLÜMLEME KARARI

Gini değeri arttıkça, homojenlik değeri de artar.



1. Gini Endeksi

Başarı ve başarısızlık için olasılık karesinin toplamını hesapla, yani (p²+q²) Şimdi oluşturulan bölümün her bir düğümünün ağırlıklı Gini bölünümü için Gini'yi hesaplayın.

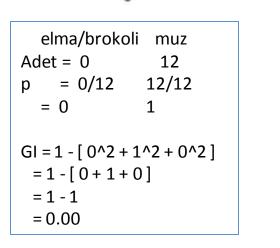


BÖLÜMLEME KARARI



1. Gini Endeksi

Başarı ve başarısızlık için olasılık karesinin toplamını hesapla, yani (p²+q²) Şimdi oluşturulan bölümün her bir düğümünün ağırlıklı Gini bölünümü için Gini'yi hesaplayın.



F