

VERİ MADENCİLİĞİ

Temel Sınıflandırma Yöntemleri

Prof. Dr. Şule Gündüz Öğüdücü

1

Konular

- Sınıflandırma işlemi
 - Sınıflandırma tanımı
 - Sınıflandırma uygulamaları
- Sınıflandırma yöntemleri
 - Karar ağaçları
 - Yapay sinir ağları
 - Bayes sınıflandırıcılar
 - Bayes ağları

2

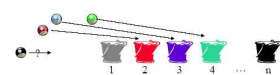
Sınıflandırma

- Sınıflandırma (classification) problemi:
 - nesnelerden oluşan veri kümesi (öğrenme kümesi):
 $D=\{t_1, t_2, \dots, t_n\}$
 - her nesne niteliklerden oluşuyor, niteliklerden biri sınıf bilgisi
- Sınıf niteliğini belirlemek için diğer nitelikleri kullanarak bir model bulma
- Öğrenme kümesinde yer almayan nesneleri (sınama kümesi) mümkün olan en iyi şekilde doğru sınıflara atamak
- sınıflandırma=ayrık değişkenler için öngöründe (prediction) bulunma

3

Sınıflandırma

- Amaç: Bir niteliğin değerini diğer nitelikleri kullanarak belirlemek
 - verinin dağılımına göre bir model bulunur
 - bulunan model, başarımı belirlendikten sonra niteliğin gelecekteki ya da bilinmeyen değerini tahmin etmek için kullanılır
 - model başarımı: doğru sınıflandırılmış sınama kümesi örneklerinin oranı
- Veri madenciliği uygulamasında:
 - ayrık nitelik değerlerini tahmin etmek: sınıflandırma
 - sürekli nitelik değerlerini tahmin etmek: öngörü

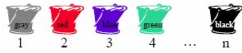


- Sınıflandırma: hangi topun hangi sepete koyulabileceği
- Öngörü: Topun ağırlığı

4

Gözetimli & Gözetimsiz Sınıflandırma

- Gözetimli (Supervised) sınıflandırma = sınıflandırma
 - Sınıfların sayısı ve hangi nesnenin hangi sınıfta olduğu biliniyor.
- Gözetimsiz (Unsupervised) sınıflandırma = demetleme (clustering)
 - Hangi nesnenin hangi sınıfta olduğu bilinmiyor. Genelde sınıf sayısı bilinmiyor.



5

Konular

- Sınıflandırma işlemi
 - Sınıflandırma tanımı
 - Sınıflandırma uygulamaları
- Sınıflandırma yöntemleri
 - Karar ağaçları
 - Yapay sinir ağları
 - Bayes sınıflandırıcılar
 - Bayes ağları

6

Sınıflandırma Uygulamaları

- Kredi başvurusu değerlendirme
- Kredi kartı harcamasının sahtekarlık olup olmadığına karar verme
- Hastalık teşhisi
- Ses tanıma
- Karakter tanıma
- Gazete haberlerini konularına göre ayırma
- Kullanıcı davranışları belirleme



7

Sınıflandırma için Veri Hazırlama

- Veri dönüşümü:
 - Sürekli nitelik değeri ayrık hale getirilir
 - Normalizasyon $[-1, \dots, 1]$, $[0, \dots, 1]$
- Veri temizleme:
 - gürültüyü azaltma
 - gereksiz nitelikleri silme

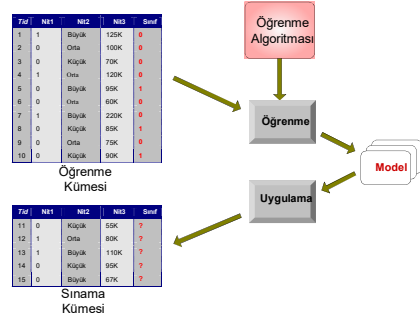
8

Sınıflandırma İşlemi

- Sınıflandırma işlemi üç aşamadan oluşur:
 1. Model oluşturma
 2. Model değerlendirme
 3. Modeli kullanma

9

Örnek



10

Sınıflandırma İşlemi: Model Oluşturma

1. Model Oluşturma:
 - Her nesnenin sınıf etiketi olarak tanımlanan niteliğinin belirlediği bir sınıfta olduğu varsayılır
 - Model oluşturmak için kullanılan nesnelerin oluşturduğu veri kümesi öğrenme kümesi olarak tanımlanır
 - Model farklı biçimlerde ifade edilebilir
 - IF – THEN – ELSE kuralları ile
 - Karar ağaçları ile
 - Matematiksel formüller ile

11

Sınıflandırma İşlemi: Model Değerlendirme

2. Model Değerlendirme:
 - Modelin başarımı (doğruluğu) sınama kümesi örnekleri kullanılarak belirlenir
 - Sınıf etiketi bilinen bir sınama kümesi örneği model kullanılarak belirlenen sınıf etiketiyle karşılaştırılır
 - Modelin doğruluğu, doğru sınıflandırılmış sınama kümesi örneklerinin toplam sınama kümesi örneklerine oranı olarak belirlenir
 - Sınama kümesi model öğrenirken kullanılmaz

12

Sınıflandırma İşlemi: Modeli Kullanma

3. Modeli kullanma:

- Model daha önce görülmemiş örnekleri sınıflandırmak için kullanılır
 - Örneklerin sınıf etiketlerini tahmin etme
 - Bir niteliğin değerini tahmin etme

13

Sınıflandırıcı Başarımını Değerlendirme

- Doğru sınıflandırma başarısı
- Hız
 - modeli oluşturmak için gerekli süre
 - sınıflandırma yapmak için gerekli süre
- Kararlı olması
 - veri kümesinde gürültülü ve eksik nitelik değerleri olduğu durumlarda da iyi sonuç vermesi
- Ölçeklenebilirlik
 - büyük miktarda veri kümesi ile çalışabilmesi
- Anlaşılabilir olması
 - kullanıcı tarafından yorumlanabilir olması
- Kuralların yapısı
 - birbiriyle örtüşmeyen kurallar

14

Konular

- Sınıflandırma işlemi
 - Sınıflandırma tanımı
 - Sınıflandırma uygulamaları
- Sınıflandırma yöntemleri
 - Karar ağaçları
 - Yapay sinir ağları
 - Bayes sınıflandırıcılar
 - Bayes ağları

15

Sınıflandırma Yöntemleri

- Karar ağaçları (decision trees)
- Yapay sinir ağları (artificial neural networks)
- Bayes sınıflandırıcılar (Bayes classifier)
- İlişki tabanlı sınıflandırıcılar (association-based classifier)
- k-en yakın komşu yöntemi (k- nearest neighbor method)
- Destek vektör makineleri (support vector machines)
- Genetik algoritmalar (genetic algorithms)
- ...

16

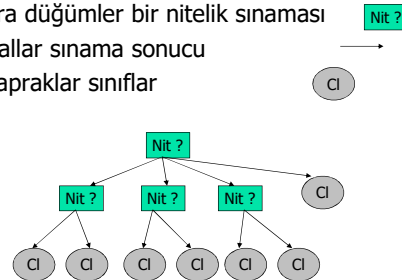
Konular

- Sınıflandırma işlemi
 - Sınıflandırma tanımı
 - Sınıflandırma uygulamaları
- Sınıflandırma yöntemleri
 - Karar ağaçları
 - Yapay sinir ağları
 - Bayes sınıflandırıcılar
 - Bayes ağları

17

Karar Ağaçları

- Akış diyagramı şeklinde ağaç yapısı
 - Ara düğümler bir nitelik sınaması
 - Dallar sınama sonucu
 - Yapraklar sınıflar



18

Örnek: Karar Ağacı

- J. Ross Quinlan'ın geliştirdiği ID3 modeline uyarlanmış:
 - hava tenis oynamaya uygun mu?

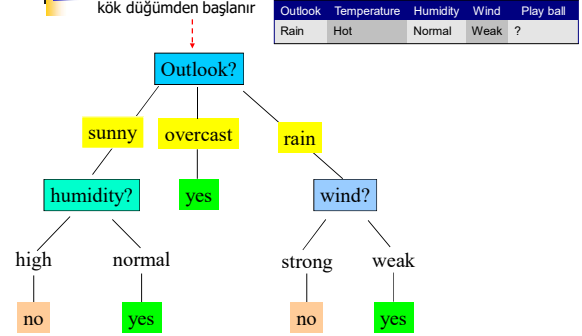
Day	Outlook	Temperature	Humidity	Wind	Play ball
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Hava durumu Verisi

19

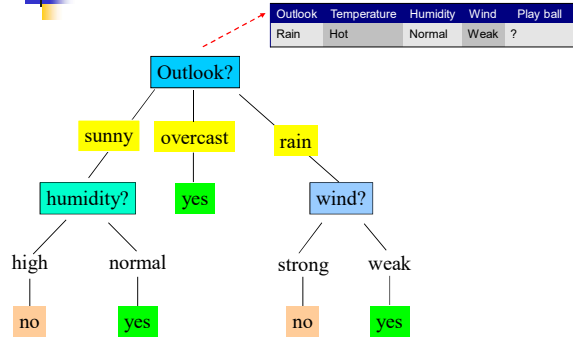
Örnek: Karar Ağacı

kök düğümden başlanır



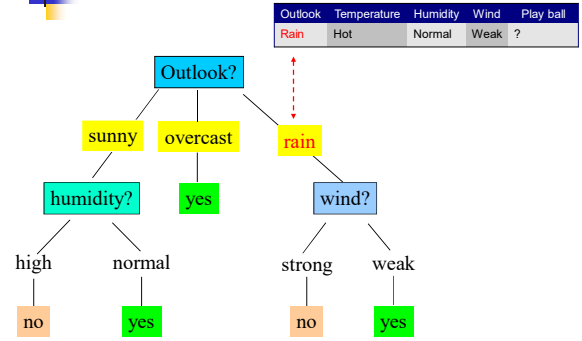
20

Örnek: Karar Ağacı



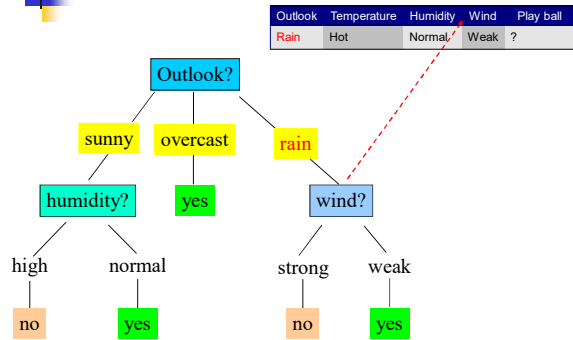
21

Örnek: Karar Ağacı



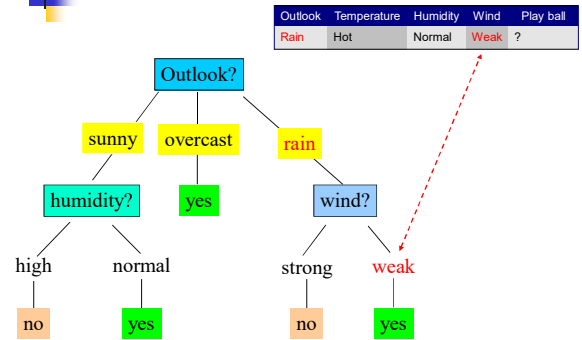
22

Örnek: Karar Ağacı



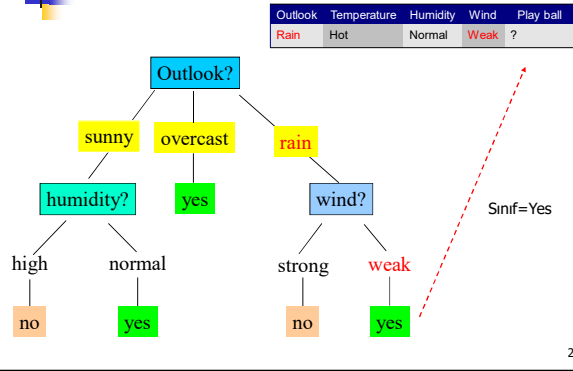
23

Örnek: Karar Ağacı



24

Örnek: Karar Ağacı



25

Karar Ağacı Yöntemleri

- Karar ağacı oluşturma yöntemleri genel olarak iki aşamadan oluşur:
 - ağaç oluşturma
 - en başta bütün öğrenme kümesi örnekleri kökte
 - seçilen niteliklere bağlı olarak örnek yinelemeli olarak bölünüyor
 - ağaç budama
 - öğrenme kümesindeki gürültülü verilerden oluşan ve sınam kümesinde hataya neden olan dalları silme (sınıflandırma başarımını artırır)

26

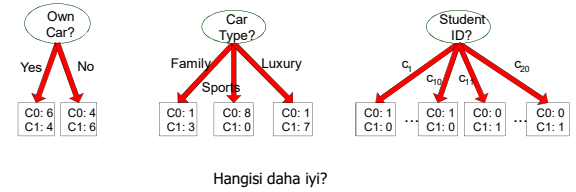
Karar Ağacı Oluşturma

- Yinelemeli işlem
 - ağaç bütün verinin oluşturduğu tek bir düğümle başlıyor
 - eğer örnekleri hepsi aynı sınıfa aitse düğüm yaprak olarak sonlanıyor ve sınıf etiketini alıyor
 - eğer değilse örnekleri sınıflara en iyi bölecek olan **nitelik seçiliyor**
 - işlem sona eriyor
 - örneklerin hepsi (çoğunluğu) aynı sınıfa ait
 - örnekleri bölecek nitelik kalmamış
 - kalan niteliklerin değerini taşıyan örnek yok
- nasıl?

27

Örnekleri En İyi Bölen Nitelik Hangisi?

- Bölmeden önce:
 - 10 örnek C0 sınıfında
 - 10 örnek C1 sınıfında



28

En İyi Bölme Nasıl Belirlenir?

- "Greedy" yaklaşım
 - çoğunlukla aynı sınıfa ait örneklerin bulunduğu düğümler tercih edilir
- Düğümün kalitesini ölçmek için bir yöntem



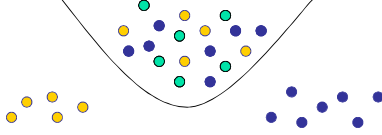
29

En İyi Bölen Nitelik Nasıl Belirlenir?

- İyilik Fonksiyonu (Goodness Function)
- Farklı algoritmalar farklı iyilik fonksiyonları kullanabilir:
 - bilgi kazancı (information gain): ID3, C4.5
 - bütün niteliklerin ayrı değerler aldığı varsayılıyor
 - sürekli değişkenlere uygulamak için değişiklik yapılabilir
 - gini index (IBM IntelligentMiner)
 - her nitelik ikiye bölünüyor
 - her nitelik için olası bütün ikiye bölünmeler sınanıyor

30

Bilgi Kazancı



sepetteki toplar farklı renklerde belirsizlik fazla
topların hepsi aynı renkte ise belirsizlik daha az

31

Bilgi / Entropi

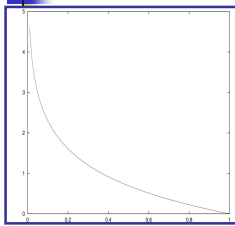
- p_1, p_2, \dots, p_s toplamı 1 olan olasılıklar. Entropi (Entropy)

$$H(p_1, p_2, \dots, p_s) = - \sum_{i=1}^s p_i \log(p_i)$$

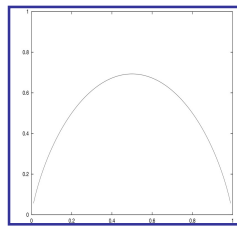
- Entropi rastgeleliği, belirsizliği ve beklenmeyen durumun ortaya çıkma olasılığını gösterir
- Sınıflandırmada
 - olayın olması beklenen bir durum
 - entropi=0

32

Entropi



log (p)



H(p, 1-p)

- örnekler aynı sınıfa aitse entropi=0
- örnekler sınıflar arasında eşit dağılmışsa entropi=1
- örnekler sınıflar arasında rastgele dağılmışsa $0 < \text{entropi} < 1$

33

Örnek

- S veri kümesinde 14 örnek: C0 sınıfına ait 9, C1 sınıfına ait 5 örnek.

- Entropi

$$H(p_1, p_2, \dots, p_s) = - \sum_{i=1}^s p_i \log(p_i)$$

- $H(p_1, p_2) = - (9/14) \log_2 (9/14) - (5/14) \log_2 (5/14) = 0.940$

34

Bilgi Kazancı (ID3 / C4.5)

- Bilgi kuramı kavramlarını kullanarak karar ağacı oluşturulur. Sınıflandırma sonucu için en az sayıda karşılaştırma yapmayı hedefler.
- Ağaç bir niteliğe göre dallandığında entropi ne kadar düşer?
- A niteliğinin S veri kümesindeki bilgi kazancı

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

Values(A), A niteliğinin alabileceği değerler, S_v , A=v olduğu durumda S'nin altkümesi.

35

Örnek

- Bilgi kazancına göre nitelik seçme

toplam örnek sayısı s=14, iki sınıfa ayrılmış

s1=9(yes), s2=5(no)

Entropy(S) = - (9/14) Log₂ (9/14) - (5/14) Log₂ (5/14) = 0.940

wind için: weak=8, strong=6

weak: no=2, yes=6

strong: no=3, yes=3

Entropy(S_{weak}) = - (6/8)*log₂(6/8) - (2/8)*log₂(2/8) = 0.811

Entropy(S_{strong}) = - (3/6)*log₂(3/6) - (3/6)*log₂(3/6) = 1.00

Entropy_{wind}(S) = (8/14)*0.811 + (6/14)*1.00

Gain(wind)=0.940 - (8/14)*0.811 - (6/14)*1.00

Gain(Outlook) = 0.246
Gain(Humidity) = 0.151
Gain(wind)=0.048
Gain(Temperature) = 0.029

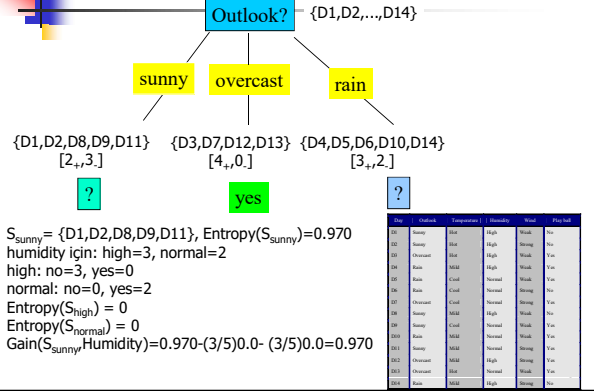
$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$H(p_1, p_2, \dots, p_s) = - \sum_{i=1}^s p_i \log(p_i)$$

ID	Outlook	Temperature	Humidity	Wind	Play
001	Sunny	Hot	High	Weak	No
002	Sunny	Hot	High	Strong	No
003	Overcast	Hot	High	Weak	Yes
004	Rain	Mild	High	Weak	Yes
005	Rain	Cool	Normal	Weak	Yes
006	Rain	Cool	Normal	Strong	No
007	Overcast	Cool	Normal	Strong	Yes
008	Sunny	Mild	High	Weak	No
009	Sunny	Cool	Normal	Weak	Yes
010	Rain	Mild	Normal	Weak	Yes
011	Sunny	Mild	Normal	Strong	Yes
012	Overcast	Hot	High	Strong	Yes
013	Overcast	Hot	Normal	Weak	Yes
014	Rain	Mild	High	Strong	No

36

Örnek



Gini Index (IBM IntelligentMiner)

- Veri kümesi S içinde n sınıf varsa ve p_j C_j sınıfının olasılığı ise

$$gini(S) = 1 - \sum_{j=1}^n p_j^2$$

- Eğer veri kümesi S_1 ve S_2 altkümelere bölünüyorsa ve her altkümede sırasıyla N_1 ve N_2 örnek varsa:

$$gini_{split}(S) = \frac{N_1}{N} gini(S_1) + \frac{N_2}{N} gini(S_2)$$

- Gini Index değeri en küçük olan nitelik seçilir.

38

Örnek

$$GINI(S) = 1 - \sum_j [p_j]^2$$

C1	0
C2	6

$P(C1) = 0/6 = 0$ $P(C2) = 6/6 = 1$
 $Gini = 1 - P(C1)^2 - P(C2)^2 = 1 - 0 - 1 = 0$

C1	1
C2	5

$P(C1) = 1/6$ $P(C2) = 5/6$
 $Gini = 1 - (1/6)^2 - (5/6)^2 = 0.278$

C1	2
C2	4

$P(C1) = 2/6$ $P(C2) = 4/6$
 $Gini = 1 - (2/6)^2 - (4/6)^2 = 0.444$

39

Sürekli Nitelikleri Bölme

- Bölmeleme:

- Statik: En başta bölmelerin
 - Bölmeler eşit genişlik, eşit derinlik veya demetleme yöntemi ile bulunur.
- Dinamik:
 - Sürekli nitelik A sıralanır. Birbirini izleyen ancak sınıf etiketi farklı olan nitelik değerleri bulunur. En fazla kazanç sağlayan bölme seçilir.

Temperature	40	48	60	72	80	90
Play tennis	No	No	Yes	Yes	Yes	No

40

Ağaç Oluşturmada Temel Yaklaşımlar

- Bölme kriteri:
 - ağacın bir düğümünde karşılaştırma yapılacak niteliğin seçilmesi
 - farklı algoritmalar farklı iyilik fonksiyonları kullanabilir: bilgi kazancı, gini index,...
- Dallanma kriteri:
 - bir örneğin hangi dala ait olduğunu belirleme
 - ikiye dallanma (gini index), çoklu dallanma (bilgi kazancı)
- Durma kararı:
 - dallanma işleminin devam edip etmeyeceğine karar verme
- Etiketleme kuralı:
 - yaprak düğüm en çok örneği olan sınıfla etiketleniyor

41

Ağaç Oluşturma:

- Parçala ve çöz (divide and conquer)
 - kökten yapraklara
 - düğümü dallara ayır
 - 'Greedy' algoritma
 - her adımda en iyi çözümü bul: her düğümde dallanmak için en iyi niteliği bul
 - her dal için algoritmayı uygula

42

Örnek Algoritma: ID3

- Bütün nitelikler ayrık
- Bir düğüm oluştur N:
 - Eğer örneklerin hepsi C sınıfına ait ise, N düğümü C etiketli yaprak
 - Eğer karşılaştırma yapılacak nitelik yoksa N düğümü en çok örneği olan sınıf
- En büyük bilgi kazancı olan niteliği bölmek için seç
 - N'yi seçilen nitelik ile etiketle
 - niteliğin her A_i değeri için bir dal oluştur
 - S_i , örneklerin hepsinin A_i değeri aldığı dal
 - S_i boş \rightarrow bir yaprak oluşturup en çok örneği olan sınıfa etiketle
 - S_i boş değil \rightarrow algoritmayı S_i düğümü üzerinde yinele
- Yaprak düğümlere kadar

Ayrıntılı bilgiler: http://dms.irb.hr/tutorial/tut_dtrees.php

43

Örnek Algoritma: C4.5

- Kökten yapraklara doğru ağaç oluşturma
- Bilgi kazancı yöntemini kullanıyor
- Bütün veri kümesini bellekte tutuyor
 - Büyük veri kümeleri için uygun değil

<http://www.rulequest.com/Personal/c4.5r8.tar.gz>

44

Karar Ağacı Kullanarak Sınıflandırma

- Doğrudan
 - sınıflandırmak istenilen örneğin nitelikleri ağaç boyunca sinanır
 - ulaşılan yaprağın etiketi sınıf bilgisini verir
- Dolaylı
 - karar ağacı sınıflandırma kurallarına dönüştürülür
 - kökten yaprakların herbirine giden yollar için ayrı bir kural oluşturulur.
 - IF-THEN şeklinde kuralları insanlar daha kolay anlıyor
 - Örnek: IF Outlook="sunny" AND humidity="normal" THEN play tennis

45

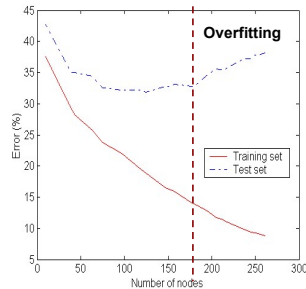
Karar Ağacı Kullanarak Sınıflandırma

- Avantajları:
 - Karar ağacı oluşturmak zahmetsiz
 - Küçük ağaçları yorumlamak kolay
 - Anlaşılabilir kurallar oluşturulabilir
 - Sürekli ve ayrık nitelik değerleri için kullanılabilir
- Dezavantajları:
 - Sürekli nitelik değerlerini tahmin etmekte çok başarılı değil
 - Sınıf sayısı fazla ve öğrenme kümesi örnekleri sayısı az olduğunda model oluşturma çok başarılı değil
 - Zaman ve yer karmaşıklığı öğrenme kümesi örnekleri sayısına (q), nitelik sayısına (h) ve oluşan ağacın yapısına bağlı.
 - Ağaç oluşturma karmaşıklığı fazla, ağaç budama karmaşıklığı fazla
 - ağaç oluşturma için zaman karmaşıklığı: $O(h \cdot q \cdot \log q)$

46

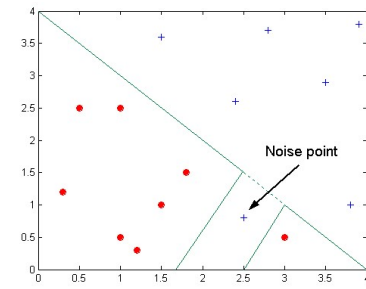
Karar Ağaçlarında Aşırı Öğrenme

- Öğrenme kümesindeki örneklerin azlığı veya gürültülü olması
- Aşırı öğrenmeyi engelleyen iki yaklaşım
 - işlemi erken sona erdirmek
 - işlemi sona erdirmek için eşik değeri belirlemek gerekiyor
 - karar ağacı oluştuktan sonra ağacı küçültme



47

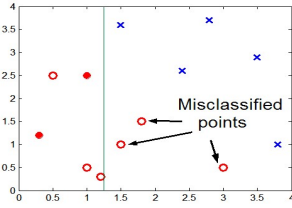
Aşırı Öğrenme:Gürültülü Örnekler



Gürültülü örnekler nedeniyle sınıfları ayıran düzlemin bozulması

48

Aşırı Öğrenme: Yetersiz Öğrenme Kümesi



- Öğrenme kümesindeki örnek sayısının yetersiz olması nedeniyle sınama kümesindeki örneklerin yanlış sınıflandırılması

49

Aşırı Öğrenme

- Gereğinden fazla karmaşık karar ağaçları aşırı öğrenmeye neden oluyor.
- Karar ağacının yeni örnekler üzerindeki başarımını tahmin etmek için öğrenme kümesi örnekleri yeterli olmuyor.
- Hatayı tahmin etmek için farklı yöntemler gerekli.

50

Genel Hatayı Tahmin Etme

- Yerine koyma (Resubstitution) Hatası: öğrenme kümesi kullanılarak hesaplanan hata ($\sum e(t)$)
- Genel (Generalization) hata: sınama kümesi kullanılarak hesaplanan hata ($\sum e'(t)$)
- Genel hatayı tahmin etme yöntemleri:
 - İyimser yaklaşım: $e'(t) = e(t)$
 - Kötümser yaklaşım:
 - Her yaprak düğüm için: $e'(t) = (e(t) + 0.5)$
 - Toplam hata: $e'(T) = e(T) + N \times 0.5$ (N: yaprak düğüm sayısı)
 - 30 yaprak düğümü olan bir karar ağacı, 1000 öğrenme kümesi örneğinden 10 örneği yanlış sınıflandırırsa
 - Yerine koyma hatası: $10/1000 = \%1$
 - Genel hata: $(10 + 30 \times 0.5)/1000 = \%2.5$
 - Ağaç budama: Genel hatayı tahmin etmek için geçerleme kümesi kullanılır

51

Occam's Razor

- Genel hatası aynı olan iki modelden karmaşıklığı daha az olan seçilmeli
- Karmaşık modellerin veri içindeki gürültüyü öğrenme ihtimalleri daha fazla

52

Karar Ağacı Boyutunu Belirleme

- Veri kümesi öğrenme ve sınama kümesi olarak ayrılır
- Çapraz geçerleme kullanılır.
- Veri kümesinin tümü ağacı oluşturmak için kullanılır
 - istatistiksel bir test ile (chi-square) düğüm eklemenin ya da ağacı küçültmenin katkısı sınanır
 - MDL (Minimum Description Length) yöntemi kullanılır: kodlama en aza indirildiğinde ağacın büyümesi durdurulur

53

Karar Ağacı: Erken Durdurma

- Karar ağacını tam oluşturmadan işlemi bitirme
 - İşlemi sona erdirmeye için ek kurallar
 - Örneklerin sayısı kullanıcı tarafından belirlenen bir eşik değerinden daha az ise
 - Örneklerin sınıf dağılımı niteliklerden bağımsız ise (chi-square testi ile belirlenebilir)
 - Ağaca yeni bir düğüm ekleyince iyilik fonksiyonu yeterince artmıyorsa

54

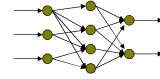
Konular

- Sınıflandırma işlemi
 - Sınıflandırma tanımı
 - Sınıflandırma uygulamaları
- Sınıflandırma yöntemleri
 - Karar ağaçları
 - Yapay sinir ağları
 - Bayes sınıflandırıcılar
 - Bayes ağları

55

Yapay Sinir Ağları ile Sınıflandırma

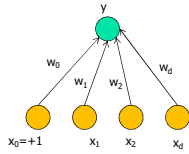
- İnsan beynindeki sinir hücrelerinin işlevini modelleyen bir yapı
- Birbiri ile bağlantılı katmanlardan oluşur.
 - katmanlar hücrelerden oluşur
- Katmanlar arasında iletim
- İletim katmanlar arasındaki bağın ağırlığına ve her hücrenin değerine bağlı olarak değişebilir



56

Algılayıcı

- Algılayıcı (perceptron)



$$y = \sum_{j=1}^d w_j x_j + w_0 = \mathbf{w}^T \mathbf{x}$$

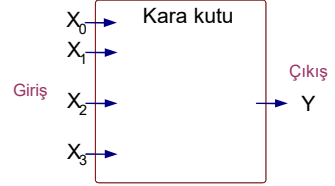
$$\mathbf{w} = [w_0, w_1, \dots, w_d]^T$$

$$\mathbf{x} = [1, x_1, \dots, x_d]^T$$

57

Örnek:

X ₁	X ₂	X ₃	Y
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1
0	0	1	0
0	1	0	0
0	1	1	1
0	0	0	0

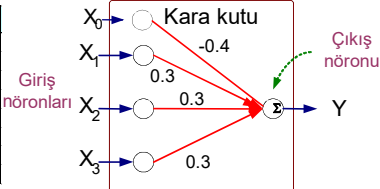


- En az iki giriş 1 ise çıkış 1, diğer durumlarda çıkış 0

58

Örnek

X ₁	X ₂	X ₃	Y
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1
0	0	1	0
0	1	0	0
0	1	1	1
0	0	0	0



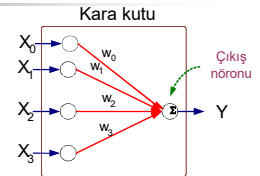
$$Y = I(0.3X_1 + 0.3X_2 + 0.3X_3 - 0.4 > 0)$$

$$I(z) = \begin{cases} 1 & \text{eğer } z > 0 \\ 0 & \text{diğer} \end{cases}$$

59

Yapay Sinir Ağları

- Birbiri ile bağlantılı nöronlar ve ağırlıklar
- Çıkış nöronu kendisine gelen girişleri ağırlıklı olarak topluyor
- Çıkış nöronu bir eşik değeri ile karşılaştırılıyor

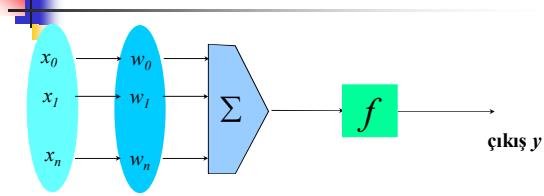


$$Y = I\left(\sum_{i=0}^3 w_i x_i\right)$$

$$Y = \text{sign}\left(\sum_{i=0}^3 w_i x_i\right)$$

60

Yapay Sinir Hücresi



Giriş vektörü x **ağırlık vektörü w** **ağırlıklı toplam** **Aktivasyon fonksiyonu**

- n-boyutlu giriş vektörü x
- Her giriş değeri kendi ağırlığıyla çarpılarak toplanır
- Aktivasyon fonksiyonu ile hücrenin girişine gelen değere karşılık çıktıya üreteceği değer belirlenir.

61

Aktivasyon Fonksiyonları

Unit step



$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

Sigmoid



$$f(x) = \frac{1}{1 + e^{-x}}$$

62

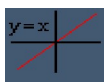
Aktivasyon Fonksiyonları

Gaussian



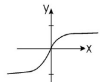
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Identity



$$f(x) = x$$

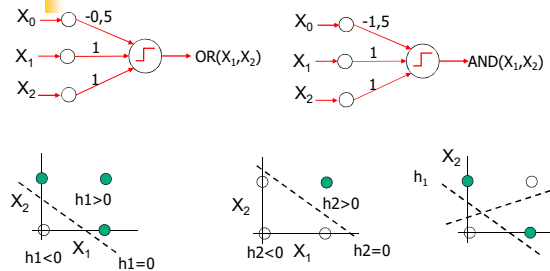
Hiberbolik Tanjant



$$f(x) = \frac{1 - \exp(-2x)}{1 + \exp(-2x)}$$

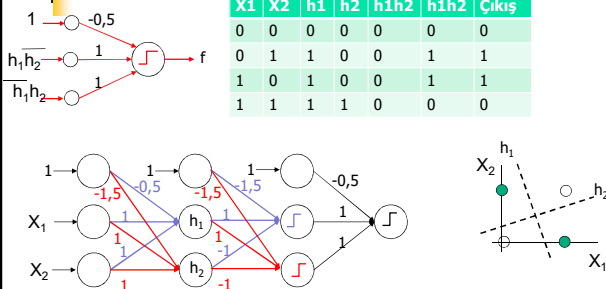
63

Çok Katmanlı YSA



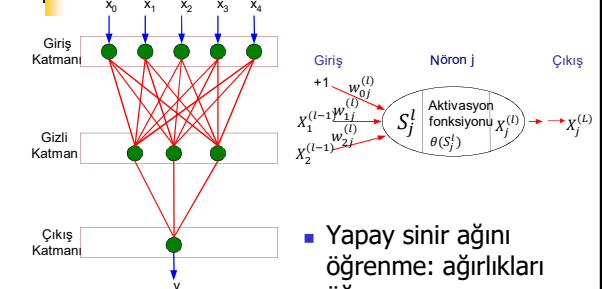
64

Katman Ekleme



65

Çok Katmanlı

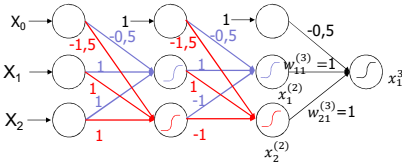


66

Yapay Sinir Ağı ile Öğrenme

$$w_{ij}^{(l)} \begin{cases} 1 \leq l \leq L & \text{Katman} \\ 0 \leq i \leq d^{(l-1)} & \text{Girişler} \\ 1 \leq j \leq d^{(l)} & \text{Çıkışlar} \end{cases} \quad x_j^{(l)} = \theta(S_j^{(l)}) = \theta\left(\sum_{i=0}^{d^{(l-1)}} w_{ij}^{(l)} x_i^{(l-1)}\right)$$

Giriş katmanı: $x_1^{(0)} \dots x_{d^{(0)}} \rightarrow x_1^{(L)} = f(x)$



67

Yapay Sinir Ağı ile Öğrenme

- Yapay sinir ağı oluşturma
 - giriş verisini modelleme
 - gizli katman sayısını, gizli katmanlardaki nöron sayısını belirleme
- Yapay sinir ağını eğitme
 - Sinir ağını küçültme
 - Sonucu yorumlama

68

Yapay Sinir Ağını Oluşturma

- Giriş nöron sayısı
 - Öğrenme kümesindeki verilerin nitelik sayısı
- Gizli nöron sayısı
 - öğrenme sırasında ayarlanır
- Çıkış nöron sayısı
 - sınıf sayısı

69

Yapay Sinir Ağını Eğitme

- Amaç: Veri kümesindeki örneklerin hepsini doğru sınıflandıracak ağırlıkları belirlemek
 - ağırlıklara rasgele değerler ata
 - öğrenme kümesindeki giriş değerlerini teker teker sinir ağına uygula
 - çıkışı hesapla
 - hata değerini hesapla
 - ağırlıkları hata fonksiyonunu en küçültecek şekilde düzelt

$$\Delta w_i = \eta \frac{\partial E}{\partial w_i}$$

$$w_i \leftarrow w_i + \Delta w_i$$

70

Backpropagation

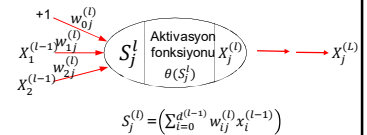
- Hata hesaplama
 - ağırlıklar $w = \{w_{ij}^{(l)}\}$
 - çıkış $f(x)$
 - bir örnek x_n, y_n için hata değeri: $e(f(x_n), y_n) = e(w)$
 - Hata fonksiyonunu en küçültmek için gradyan hesaplama:
 - $\nabla e(w) = \frac{\partial e(w)}{\partial w_{ij}^{(l)}}$ her i, j, l

71

Backpropagation

$\frac{\partial e(w)}{\partial w_{ij}^{(l)}}$ hesaplama:

$$\frac{\partial e(w)}{\partial w_{ij}^{(l)}} = \frac{\partial e(w)}{\partial S_j^{(l)}} \times \frac{\partial S_j^{(l)}}{\partial w_{ij}^{(l)}}$$



$$\frac{\partial S_j^{(l)}}{\partial w_{ij}^{(l)}} = x_i^{(l-1)}$$

$$\frac{\partial e(w)}{\partial S_j^{(l)}} = \delta_j^{(l)}$$

72

Backpropagation

$$\delta_j^{(l)} = \frac{\partial e(\mathbf{w})}{\partial S_j^{(l)}}$$

Son katman için $l = L, j = 1 \delta_1^{(L)} = \frac{\partial e(\mathbf{w})}{\partial S_1^{(L)}}$

$$e(\mathbf{w}) = e(f(\mathbf{x}_n), y_n)$$

73

Backpropagation

$$\delta_j^{(l)} = \frac{\partial e(\mathbf{w})}{\partial S_j^{(l)}}$$

Son katman için $l = L, j = 1 \delta_1^{(L)} = \frac{\partial e(\mathbf{w})}{\partial S_1^{(L)}}$

$$e(\mathbf{w}) = e(x_1^{(L)}, y_n)$$

74

Backpropagation

$$\delta_j^{(l)} = \frac{\partial e(\mathbf{w})}{\partial S_j^{(l)}}$$

Son katman için $l = L, j = 1 \delta_1^{(L)} = \frac{\partial e(\mathbf{w})}{\partial S_1^{(L)}}$

$$e(\mathbf{w}) = e(x_1^{(L)} - y_n)^2$$

$$x_1^{(L)} = \theta(S_1^{(L)})$$

$$\theta'(S) = 1 - \theta^2(S) \text{ (hiperbolik tanjant için)}$$

75

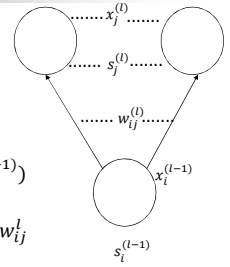
Backpropagation

$$\delta_i^{(l-1)} = \frac{\partial e(\mathbf{w})}{\partial S_i^{(l-1)}}$$

$$= \sum_{j=1}^{d^{(l)}} \frac{\partial e(\mathbf{w})}{\partial S_j^{(l)}} \times \frac{\partial S_j^{(l)}}{\partial x_i^{(l-1)}} \times \frac{\partial x_i^{(l-1)}}{\partial S_i^{(l-1)}}$$

$$= \sum_{j=1}^{d^{(l)}} \delta_j^{(l)} \times w_{ij}^{(l)} \times \theta'(S_i^{(l-1)})$$

$$\delta_i^{(l-1)} = \left(1 - (x_i^{(l-1)})^2\right) \sum_{j=1}^{d^{(l)}} \delta_j^{(l)} \times w_{ij}^{(l)}$$



76

Backpropagation

Initialize all weights $w_{ij}^{(l)}$ at random

for $t=0,1,2,\dots$ Do

pick $n \in \{1,2,\dots,N\}$

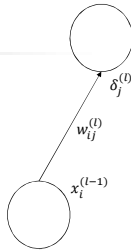
compute all $x_j^{(l)}$

compute all $\delta_j^{(l)}$

update the weights: $w_{ij}^{(l)} \leftarrow w_{ij}^{(l)} - \eta x_i^{(l-1)} \delta_j^{(l)}$

Iterate to the next step until stop

Return the final weights $w_{ij}^{(l)}$



77

Yapay Sinir Ağını Küçültme

- Tam bağlı ağıın anlaşılması çok güç
- n giriş nöron, d gizli nöron, l çıkış nöronu $d(l+n)$ ağırlık
- Küçültme: ağırlıklardan bazıları sınıflandırma sonucunu etkilemeyecek şekilde silinir

78

Yapay Sinir Ağları

- Avantajları
 - doğru sınıflandırma oranı genelde yüksek
 - kararlı – öğrenme kümesinde hata olduğu durumda da çalışıyor
 - çıkış ayırık, sürekli ya da ayırık veya sürekli değişkenlerden oluşan bir vektör olabilir
- Dezavantajları
 - öğrenme süresi uzun
 - öğrenilen fonksiyonun anlaşılması zor

79

Konular

- Sınıflandırma işlemi
 - Sınıflandırma tanımı
 - Sınıflandırma uygulamaları
- Sınıflandırma yöntemleri
 - Karar ağaçları
 - Yapay sinir ağları
 - Bayes sınıflandırıcılar
 - Bayes ağları

80

Bayes (İstatistiksel) Modelleme

- Bayes teoremini kullanan istatistiksel sınıflandırıcı
- Örneklerin hangi sınıfa hangi olasılıkla ait oldukları
- Naïve Bayes sınıflandırıcı
 - niteliklerin hepsi aynı derecede önemli
 - nitelikler birbirinden bağımsız
 - bir niteliğin değeri başka bir nitelik değeri hakkında bilgi içermiyor
 - sınıflandırma ve öğrenme problemleri

81

Bayes Teoremi

- X sınıflandırılacak örnek. Hipotez h , X örneğinin C sınıfına ait olduğu
- h hipotezinin sonrasal olasılığı (*posteriori probability*)

$$P(h|X) = \frac{P(X|h)P(h)}{P(X)}$$

- MAP (maximum posteriori) hipotez

$$h_{MAP} = \underset{h \in H}{\operatorname{argmax}} P(h|D) = \underset{h \in H}{\operatorname{argmax}} P(D|h)P(h)$$

- Çok sayıda olasılığı önceden kestirmek gerekiyor

82

Örnek

- $X \xrightarrow{H} C$
- $P(H) = P(\text{apple})$ $P(X) = P(\text{red} + \text{orange})$
- $P(X|H) = P(\text{apple} \text{ ise } \text{red} + \text{orange})$

$$P(h|X) = \frac{P(X|h)P(h)}{P(X)}$$

83

Örnek

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

84

Naïve Bayes Sınıflandırıcı

- $X=(x_1, x_2, \dots, x_n)$ örneğinin C sınıfında olma olasılığı $P(C|X)$ nedir?
- $\frac{P(X|C_i)P(C_i)}{P(X)}$ değerini enbüyütme
→ $P(X|C_i)P(C_i)$ değerini enbüyütme
- $P(C_i)=|S_i|/|S|$, S_i : C_i sınıfına ait örneklerin sayısı
- $P(X|C_i)=\prod_{k=1}^n P(x_k|C_i)$, $P(x_k|C_i)=s_{ik}/s_i$
- Hesaplama maliyetini azaltıyor, sadece sınıf dağılımları hesaplanıyor
- Naïve: nitelikler bağımsız

85

Hava Durumu Verisi için Olasılıklar

Outlook	Yes		No	Yes		No	Yes		No	Yes		No	Yes		No	Play
	Yes	No		Yes	No		Yes	No		Yes	No		Yes	No		
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5			
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3					
Rainy	3	2	Cool	3	1											
Sunny	2/9	3/5	Hot	2/9	2/5	High	3/9	4/5	False	6/9	2/5	9/14	5/14			
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	True	3/9	3/5					
Rainy	3/9	2/5	Cool	3/9	1/5											

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Hava Durumu Verisi için Olasılıklar

Outlook	Yes		No	Yes		No	Yes		No	Yes		No	Play
	Yes	No		Yes	No		Yes	No		Yes	No		
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/5	Hot	2/9	2/5	High	3/9	4/5	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	True	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

Yeni veri

$$P(C_i|X) = P(X|C_i)P(C_i) = \prod_{k=1}^n P(x_k|C_i)P(C_i)$$

Outlook	Temp	Humidity	Windy	Play
Sunny	Cool	High	True	?

İki Sınıf için olasılık:

$$P(\text{"yes"}|X) = 2/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14 = 0.0053$$

$$P(\text{"no"}|X) = 3/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14 = 0.0206$$

Normalize edilmiş olasılıklar:

$$P(\text{"yes"}) = 0.0053 / (0.0053 + 0.0206) = 0.205$$

$$P(\text{"no"}) = 0.0206 / (0.0053 + 0.0206) = 0.795$$

87

Sürekli Veriler için Olasılık

- Verinin normal dağılımdan geldiği varsayılıyor.
- Her sınıf-nitelik çifti için bir olasılık hesaplanıyor.

$$P(A_i|c_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(A_i - \mu_j)^2}{2\sigma_j^2}}$$

- Gelir için sınıf=-1
- ortalama=110
- varyans=2975

$$P(\text{Gelir} = 120 | -1) = \frac{1}{\sqrt{2\pi(54.54)}} e^{-\frac{(120-110)^2}{2(2975)}} = 0.0072$$

Tid	Geri Ödeme	Medeni Durum	Gelir	Dolan dinci
1	Evet	Bekar	125K	-1
2	Hayır	Evli	100K	-1
3	Hayır	Bekar	70K	-1
4	Evet	Evli	120K	-1
5	Hayır	Boşanmış	95K	1
6	Hayır	Evli	60K	-1
7	Evet	Boşanmış	220K	-1
8	Hayır	Bekar	85K	1
9	Hayır	Evli	75K	-1
10	Hayır	Bekar	90K	1

88

Örnek

$X=(\text{Geri Ödeme}=\text{"Hayır"}, \text{Medeni Durum}=\text{"Evli"}, \text{Gelir}=120k)$

$$\begin{aligned} P(\text{Geri Ödeme}=\text{"Evet"}|-1) &= 3/7 \\ P(\text{Geri Ödeme}=\text{"Hayır"}|-1) &= 4/7 \\ P(\text{Geri Ödeme}=\text{"Evet"}|1) &= 0 \\ P(\text{Geri Ödeme}=\text{"Hayır"}|1) &= 1 \\ P(\text{Medeni Durum}=\text{"Evli"}|-1) &= 4/7 \\ P(\text{Medeni Durum}=\text{"Bekar"}|-1) &= 2/7 \\ P(\text{Medeni Durum}=\text{"Boşanmış"}|-1) &= 1/7 \\ P(\text{Medeni Durum}=\text{"Evli"}|1) &= 0 \\ P(\text{Medeni Durum}=\text{"Bekar"}|1) &= 2/3 \\ P(\text{Medeni Durum}=\text{"Boşanmış"}|1) &= 1/3 \end{aligned}$$

Gelir:

Sınıf=-1

ortalama=110

varyans=2975

Sınıf=1

ortalama=90

varyans=25

Tid	Geri Ödeme	Medeni Durum	Gelir	Dolan dinci
1	Evet	Bekar	125K	-1
2	Hayır	Evli	100K	-1
3	Hayır	Bekar	70K	-1
4	Evet	Evli	120K	-1
5	Hayır	Boşanmış	95K	1
6	Hayır	Evli	60K	-1
7	Evet	Boşanmış	220K	-1
8	Hayır	Bekar	85K	1
9	Hayır	Evli	75K	-1
10	Hayır	Bekar	90K	1

89

Örnek

$X=(\text{Geri Ödeme}=\text{"Hayır"}, \text{Medeni Durum}=\text{"Evli"}, \text{Gelir}=120k)$

$$\begin{aligned} P(\text{Geri Ödeme}=\text{"Evet"}|-1) &= 3/7 \\ P(\text{Geri Ödeme}=\text{"Hayır"}|-1) &= 4/7 \\ P(\text{Geri Ödeme}=\text{"Evet"}|1) &= 0 \\ P(\text{Geri Ödeme}=\text{"Hayır"}|1) &= 1 \\ P(\text{Medeni Durum}=\text{"Evli"}|-1) &= 4/7 \\ P(\text{Medeni Durum}=\text{"Bekar"}|-1) &= 2/7 \\ P(\text{Medeni Durum}=\text{"Boşanmış"}|-1) &= 1/7 \\ P(\text{Medeni Durum}=\text{"Evli"}|1) &= 0 \\ P(\text{Medeni Durum}=\text{"Bekar"}|1) &= 2/3 \\ P(\text{Medeni Durum}=\text{"Boşanmış"}|1) &= 1/3 \end{aligned}$$

Gelir:

Sınıf=-1

ortalama=110

varyans=2975

Sınıf=1

ortalama=90

varyans=25

$$\begin{aligned} P(X|\text{Sınıf}=-1) &= \\ P(\text{Geri Ödeme}=\text{"Hayır"}|\text{Sınıf}=-1) &= 4/7 \\ \times P(\text{Medeni Durum}=\text{"Evli"}|\text{Sınıf}=-1) &= 4/7 \\ \times P(\text{Gelir}=120K|\text{Sınıf}=-1) &= 0.0072 \\ &= 4/7 \times 4/7 \times 0.0072 = 0.0024 \end{aligned}$$

$$\begin{aligned} P(X|\text{Sınıf}=1) &= \\ P(\text{Geri Ödeme}=\text{"Hayır"}|\text{Sınıf}=1) &= 1 \\ \times P(\text{Medeni Durum}=\text{"Evli"}|\text{Sınıf}=1) &= 0 \\ \times P(\text{Gelir}=120K|\text{Sınıf}=1) &= 1.2 \times 10^{-9} = 0 \end{aligned}$$

$$\begin{aligned} P(X|-1)P(-1) &> P(X|1)P(1) \\ P(-1|X) &> P(1|X) \\ \Rightarrow \text{Sınıf} &= -1 \end{aligned}$$

90

Olasılığın Sıfır Olması

- Her sınıfta bir niteliğin her değeri olmazsa
 - koşullu olasılıklardan biri 0
 - o sınıfa ait olma olasılığı 0
- Olasılıklar

$$Original : P(A_i | C) = \frac{N_{ic}}{N_c}$$

k: nitelik sayısı
Toplamları 1 olmak zorunda

$$Laplace : P(A_i | C) = \frac{N_{ic} + 1}{N_c + k}$$

91

Bayes Sınıflandırıcılar

- Avantajları:
 - gerçeklemesi kolay
 - çoğu durumda iyi sonuçlar
- Dezavantajları
 - varsayım: sınıf bilgisi verildiğinde nitelikler bağımsız
 - gerçek hayatta değişkenler birbirine bağımlı
 - değişkenler arası ilişki modellenemiyor
- Çözüm:
 - Bayes ağları

92

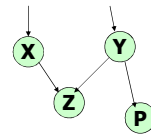
Konular

- Sınıflandırma işlemi
 - Sınıflandırma tanımı
 - Sınıflandırma uygulamaları
- Sınıflandırma yöntemleri
 - Karar ağaçları
 - Yapay sinir ağları
 - Bayes sınıflandırıcılar
 - Bayes ağları

93

Bayes Ağları

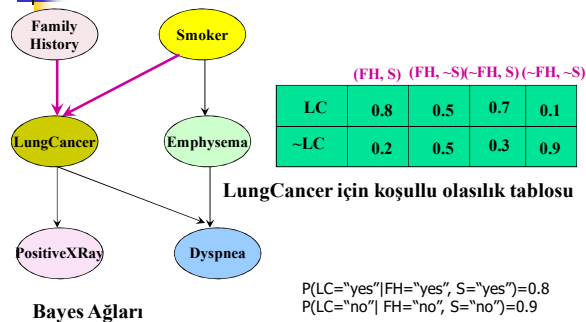
- Niteliklerin altkümesinin birbiri ile bağımsız olduğunu varsayıyor
- Yönlü çevrimsiz çizge (directed acyclic graph) ve koşullu olasılık tablolarından oluşur
- Her değişken A için bir tablo var
 - niteliğin ebeveynlerine olan koşullu olasılıkları



- düğüm: rasgele değişkenler
- ayrıntlar: olasılıklı bağımlılık
- X ve Y, Z değişkeninin ebeveyni
- Y, P değişkeninin ebeveyni
- Z ve P arasında bağ yok

94

Örnek



95

Bayes Ağlarının Eğitilmesi

- Ağ yapısı ve tüm değişkenler biliniyorsa koşullu olasılıklar hesaplanır
- Ağ yapısı belli ancak bazı değişkenler eksik ise yinelemeli öğrenme uygulanır
 - gradient descent algoritması
- D. Heckerman. [A Tutorial on Learning with Bayesian Networks](#). In *Learning in Graphical Models*, M. Jordan, ed.. MIT Press, Cambridge, MA, 1999. Also appears as Technical Report MSR-TR-95-06, Microsoft Research, March, 1995.

96