Yalın Bayes Sınıflandırıcı

Prof. Dr. Hamdi Tolga KAHRAMAN

Sınıflandırma Problemleri

Olasılık

Olasılık ifadesinin birçok kullanım şekli vardır. Rasgele bir A olayının herhangi bir olaydan bağımsız olarak gerçekleşme ihtimalini ifade etmek için P(A) notasyonu kullanılır. A olayının olasılığı olarak bilinen bu ifade "önsel" (prior), "koşulsuz" (unconditional) veya "marjinal" (marginal) olasılık isimleriyle kullanılabilir.

Bayes Sınıflandırıcılar

- Bayes Sınıflayıcı Bayes teoremine göre istatistiksel kestirim yapar.

- Bir örneğin sınıf üyelik olasılığını kestirir.

- *Naïve Bayesian sınıflandırıcı* (simple Bayesian classifier) oldukça başarılı bir sınıflayıcıdır.

Bayes Kuralı

- $p(\mathbf{x}|Cj)$: Sınıf j'den bir örneğin x olma olasılığı
- $P(Cj)$: Sınıf j'nin ilk olasılığı
- $p(\mathbf{x})$: Herhangi bir örneğin x olma olasılığı
- $P(Cj|\mathbf{x})$: x olan bir örneğin sınıf j'den olma olasılığı (son olasılık)

$$P(C_j \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid C_j)P(C_j)}{p(\mathbf{x})} = \frac{p(\mathbf{x} \mid C_j)P(C_j)}{\sum_k p(\mathbf{x} \mid C_k)P(C_k)}$$

Naïve Bayes sınıflandırıcı

- T öğrenme kümesinde bulunan her örnek n boyutlu uzayda tanımlı olsun, $\mathbf{X} = (x_1, x_2, \ldots, x_n)$
- Veri kümesinde m adet sınıf bulunuyor olsun, $C_1, C_2, \ldots, C_m$
- Sınıflamada son olasılığı en büyütme aranır (the maximal $P(C_i|\mathbf{X})$)

- Bayes teoreminden türetilebilir

$$P(C_i|\mathbf{X}) = \frac{P(\mathbf{X}|C_i)P(C_i)}{P(\mathbf{X})}$$

- P(X) olasılıgı bütün sınıflar için sabit olduğuna göre, sadece olasılığı için en büyük değer aranır.

$$P(C_i|\mathbf{X}) = P(\mathbf{X}|C_i)P(C_i)$$

Naïve Bayes sınıflandırıcı

$$P(C_i|\mathbf{X}) = P(\mathbf{X}|C_i)P(C_i)$$

- Eğer bu basitleştirilmiş ifadede bütün özellikler bağımsız ise $P(X|C_i)$ aşağıdaki şekilde yazılabilir.

$$P(\mathbf{X}|C_i) = \prod_{k=1}^{n} P(x_k|C_i) = P(x_1|C_i) \times P(x_2|C_i) \times \ldots \times P(x_n|C_i)$$

- Böylece hesap karmaşıklığı büyük ölçüde azaltılmış olur.

Bayes Sınıflandırıcı için Örnek

Table 10.4 • **Data for Bayes Classifier**

Magazine Promotion	Watch Promotion	Life Insurance Promotion	Credit Card Insurance	Sex
Yes	No	No	No	Male
Yes	Yes	Yes	Yes	Female
No	No	No	No	Male
Yes	Yes	Yes	Yes	Male
Yes	No	Yes	No	Female
No	No	No	No	Female
Yes	Yes	Yes	Yes	Male
No	No	No	No	Male
Yes	No	No	No	Male
Yes	Yes	Yes	No	Female

Bayes Sınıflandırıcı için Örnek

- Sınıflandırılacak örnek:
  - Magazine Promotion = *Yes*
  - Watch Promotion = *Yes*
  - Life Insurance Promotion = *No*
  - Credit Card Insurance = *No*
  - Sex = ?

Bayes Sınıflandırıcı için Örnek

Table 10.5 • **Counts and Probabilities for Attribute Sex**

	Magazine Promotion		Watch Promotion		Life Insurance Promotion		Credit Card Insurance	
Sex	Male	Female	Male	Female	Male	Female	Male	Female
Yes	4	3	2	2	2	3	2	1
No	2	1	4	2	4	1	4	3
Ratio: yes/total	4/6	3/4	2/6	2/4	2/6	3/4	2/6	1/4
Ratio: no/total	2/6	1/4	4/6	2/4	4/6	1/4	4/6	3/4

Bayes Sınıflandırıcı için Örnek

- Sex = Male için olasılık hesabı

$$P(sex = male \mid E) = \frac{P(E \mid sex = male)\, P(sex = male)}{P(E)}$$

Bayes Sınıflandırıcı için Örnek

- Sex = Male için koşullu olasılıklar;
  - $P(magazine\ promotion = yes \mid sex = male) = 4/6$
  - $P(watch\ promotion = yes \mid sex = male) = 2/6$
  - $P(life\ insurance\ promotion = no \mid sex = male) = 4/6$
  - $P(credit\ card\ insurance = no \mid sex = male) = 4/6$
  - $P(E \mid sex = male) = (4/6)\ (2/6)\ (4/6)\ (4/6) = 8/81$

$$P(sex = male \mid E) \approx (8/81)\ (6/10)\ /\ P(E)$$

$$P(sex = male \mid E) \approx 0{,}0593\ /\ P(E)$$

Bayes Sınıflandırıcı için Örnek

- Sex = Female için olasılık hesabı

$$P(sex = female \mid E) = \frac{P(E \mid sex = female)\,P(sex = female)}{P(E)}$$

Bayes Sınıflandırıcı için Örnek

- Sex = Female için koşullu olasılıklar;

  - $P(magazine\ promotion = yes \mid sex = female) = 3/4$
  - $P(watch\ promotion = yes \mid sex = female) = 2/4$
  - $P(life\ insurance\ promotion = no \mid sex = female) = 1/4$
  - $P(credit\ card\ insurance = no \mid sex = female) = 3/4$

  - $P(E \mid sex = female) = (3/4)\ (2/4)\ (1/4)\ (3/4) = 9/128$

$$P(sex = female \mid E) \approx (9/128)\ (4/10)\ /\ P(E)$$

$$P(sex = female \mid E) \approx 0{,}0281\ /\ P(E)$$

Bayes Sınıflandırıcı için Örnek

- $P(sex = male \mid E) \approx 0{,}0593 \,/\, P(E)$
- $P(sex = female \mid E) \approx 0{,}0281 \,/\, P(E)$

Bayes sınıflayıcı 0,0593 > 0,0281 olduğu için E davranışını gösteren kart sahibi erkektir.

Bayes Sınıflayıcı : Sayısal özellik

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

where

e = the *exponential function*

μ = the class *mean* for the given numerical attribute

σ = the class *standard deviation* for the attribute

x = the attribute value

Sınıflandırma Problemleri

Olasılık

Bir para atma olayının iki kez tekrarlanması durumunda ardarda iki defa da tura gelme ihtimalini bulalım.

$$P(A \cap B) = P(A)P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

Sınıflandırma Problemleri

Koşullu Olasılık

Rasgele bir A olayının, farklı bir rasgele B olayına bağlı gerçekleşmesi ihtimalini ifade etmek için önsel olasılıklar yeterli olmaz. Bu yüzden "koşullu" (conditional) veya "sonsal" (posterior) olasılık olarak isimlendirilen P(A|B) notasyonu kullanılır. Bilinen bir B olayına göre A olayının koşullu olasılığı şöyle gösterilir.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Sınıflandırma Problemleri

Bayes Teoremi

Birbirinden bağımsız ve rasgele iki olayın (A ve B) birbiri ardı sıra gerçekleştiği durumlarda bu iki olaydan birinin gerçekleşmesi durumunda ikinci olayın gerçekleşme olasılığı P(A,B) veya P(B,A) yada $P(A \cap B)$ ifadesi ile gösterilebilir. Değişme özelliği sayesinde aşağıdaki çarpım kuralı iki farklı ifade ile yazılabilir.

$$P(A \cap B) = P(A \mid B)P(B)$$

$$P(A \cap B) = P(B \mid A)P(A)$$

Bayes Sınıflandırıcı

- Bir olasılıksal sınıflandırma olan Yalın Bayes sınıflandırmanın ana fikri, bir belgenin sınıfının olasılığını tahmin etmek için verilen bir kelimenin sınıfının koşullu olasılıklarını kullanmaktır. Belge sınıflandırma gibi bazı öğrenme problemlerinde yaygın olarak kullanılan en pratik yaklaşımdır.

- Bu yaklaşımın "Yalın" (Naive) kısmı içindeki kelime bağımsızlığı varsayımından kaynaklanmaktadır. Çünkü kelime kombinasyonlarının olasılıklarını tahminci olarak kullanmaz.

- Bu yüzden Karar Ağacı (Decision Tree) gibi algoritmaların üstel karmaşıklığından öte verimli bir yaklaşımdır ve performansı Yapay Sinir Ağları(Neural Networks) ve Karar Ağacı ile karşılaştırılabilir.

Sınıflandırma Problemleri

Bayes Teoremi

Bayes teoremi (kuralı veya kanunu) stokastik (rassal) bir sürece bağlı olarak ortaya çıkan rasgele bir A olayı ile diğer bir rasgele B olayı için koşullu olasılıklar ve marjinal olasılıklar arasındaki ilişkiyi tanımlar. Bu ilişkiyi ilk kez Thomas Bayes (1702-1761) ortaya atmış ve aşağıdaki eşitliği önermiştir.

$$P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B)}$$

Bayes Sınıflandırıcı

A ve B rasgele olaylar olsun;

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

$P(A)$: A olayının bağımsız olasılığı prior (öncül) olasılık

$P(B)$: B olayının bağımsız olasılığı

$P(B \mid A)$: A olayının olduğu bilindiğinde B olayının olasılığı likelihood (şartlı olasılık)

$P(A \mid B)$: B olayının olduğu bilindiğinde A olayının olasılığı posterior (artçıl) olasılık

Bayes kuralına dayanarak $P(A \mid B)$ yi maksimum yapan durumlar hesaplanabilir.

Bayes Sınıflandırıcı

- Yalın Bayes'de Artımlı (Incremantal) olarak tabir edilen online bir öğrenme durumu vardır; her bir talim örneği artımlı olarak bir hipotezin doğru olma olasılığını arttırır veya azaltır.

- Bu durum, klasikleşmiş bir örnek olan, bir tenis maçının oynanma olasılığı örneği ile ele alınmıştır. Verilen hava şartlarına göre bir tenis maçının oynanıp oynanmayacağı belirlenmektedir.

- örnek veri kümesi kullanılarak frekans ve olasılıklar hesaplanabilir.

Bayes Sınıflandırıcı

Örnek veri kümesi

GÖRÜNÜM	SICAKLIK	NEM	RÜZGAR	OYUN
Güneşli	sıcak	yüksek	Yok	hayır
Güneşli	sıcak	yüksek	Var	hayır
Bulutlu	sıcak	yüksek	Yok	evet
Yağmurlu	ılık	yüksek	Yok	evet
Yağmurlu	serin	normal	Yok	evet
Yağmurlu	serin	normal	Var	hayır
Bulutlu	serin	normal	Var	evet
Güneşli	ılık	yüksek	Yok	hayır
Güneşli	serin	normal	Yok	evet
Yağmurlu	ılık	normal	Yok	evet
Güneşli	ılık	normal	Var	evet
Bulutlu	ılık	yüksek	Var	evet
Bulutlu	sıcak	normal	Yok	evet
Yağmurlu	ılık	yüksek	Var	hayır

Bayes Sınıflandırıcı

Örnek veri kümesi

GÖRÜNÜM	SICAKLIK	NEM	RÜZGAR	OYUN
Güneşli	sıcak	yüksek	Yok	hayır
Güneşli	sıcak	yüksek	Var	hayır
Bulutlu	sıcak	yüksek	Yok	evet
Yağmurlu	ılık	yüksek	Yok	evet
Yağmurlu	serin	normal	Yok	evet
Yağmurlu	serin	normal	Var	hayır
Bulutlu	serin	normal	Var	evet
Güneşli	ılık	yüksek	Yok	hayır
Güneşli	serin	normal	Yok	evet
Yağmurlu	ılık	normal	Yok	evet
Güneşli	ılık	normal	Var	evet
Bulutlu	ılık	yüksek	Var	evet
Bulutlu	sıcak	normal	Yok	evet
Yağmurlu	ılık	yüksek	Var	hayır

Frekanslar

GÖRÜNÜM	Oyun evet	Oyun hayır	SICAKLIK	Oyun evet	Oyun hayır	NEM	Oyun evet	Oyun hayır	RÜZGAR	Oyun evet	Oyun hayır	OYUN evet	OYUN hayır
Güneşli	2	3	Sıcak	2	2	Yüksek	3	4	Yok	6	2	9	5
Bulutlu	4	0	Ilık	4	2	Normal	6	1	Var	3	3		
Yağmurlu	3	2	Serin	3	1								

Bayes Sınıflandırıcı

Frekanslar

GÖRÜNÜM	Oyun evet	Oyun hayır	SICAKLIK	Oyun evet	Oyun hayır	NEM	Oyun evet	Oyun hayır	RÜZGAR	Oyun evet	Oyun hayır	OYUN	
												evet	hayır
Güneşli	2	3	Sıcak	2	2	Yüksek	3	4	Yok	6	2	9	5
Bulutlu	4	0	Ilık	4	2	Normal	6	1	Var	3	3		
Yağmurlu	3	2	Serin	3	1								

Olasılıklar

GÖRÜNÜM	Oyun evet	Oyun hayır	SICAKLIK	Oyun evet	Oyun hayır	NEM	Oyun evet	Oyun hayır	RÜZGAR	Oyun evet	Oyun hayır	OYUN	
												evet	hayır
Güneşli	2/9	3/5	Sıcak	2/9	2/5	Yüksek	3/9	4/5	Yok	6/9	2/5	9/14	5/14
Bulutlu	4/9	0/5	Ilık	4/9	2/5	Normal	6/9	1/5	Var	3/9	3/5		
Yağmurlu	3/9	2/5	Serin	3/9	1/5								

Bayes Sınıflandırıcı

GÖRÜNÜM	Oyun evet	Oyun hayır	SICAKLIK	Oyun evet	Oyun hayır	NEM	Oyun evet	Oyun hayır	RÜZGAR	Oyun evet	Oyun hayır	OYUN evet	OYUN hayır
Güneşli	2/9	3/5	Sıcak	2/9	2/5	Yüksek	3/9	4/5	Yok	6/9	2/5	9/14	5/14
Bulutlu	4/9	0/5	Ilık	4/9	2/5	Normal	6/9	1/5	Var	3/9	3/5		
Yağmurlu	3/9	2/5	Serin	3/9	1/5								

Olasılıklar

Bu örnek veriler eşliğinde aşağıdaki yeni durum sınıflandırılırsa şekildeki duruma ulaşılır.

Görünüm	Sıcaklık	Nem	Rüzgar	Oyun
Günesli	Serin	Yüksek	Var	?

$$v_{YB} = \underset{v_j \in \{evet, hayır\}}{\arg\max} P(v_j) \prod_i P(a_i \mid v_j)$$

$$= \underset{v_j \in \{evet, hayır\}}{\arg\max} P(v_j) P(görünüm = güneşli \mid v_j) P(sıcaklık = serin \mid v_j) \ldots\ldots\ldots$$

Öncül Olasılıklar (2)

Şartlı Olasılıklar (8)

Bayes Sınıflandırıcı

Olasılıklar

GÖRÜNÜM	Oyun evet	Oyun hayır	SICAKLIK	Oyun evet	Oyun hayır	NEM	Oyun evet	Oyun hayır	RÜZGAR	Oyun evet	Oyun hayır	OYUN evet	OYUN hayır
Güneşli	2/9	3/5	Sıcak	2/9	2/5	Yüksek	3/9	4/5	Yok	6/9	2/5	9/14	5/14
Bulutlu	4/9	0/5	Ilık	4/9	2/5	Normal	6/9	1/5	Var	3/9	3/5		
Yağmurlu	3/9	2/5	Serin	3/9	1/5								

Bu örnek veriler eşliğinde aşağıdaki yeni durum sınıflandırılırsa şekildeki duruma ulaşılır.

Görünüm	Sıcaklık	Nem	Rüzgar	Oyun
Güneşli	Serin	Yüksek	Var	?

Anafikir: talim kümesindeki olasılık dağılımına dayanarak her bir sınıf için olasılık hesaplanır. Önce her bir özelliğin olasılığı hesaba katılır, bütün özellikler eşdeğer önemde ele alınır ve olasılıklar çarpılır.

P(EVET) = 2/9 3/9 3/9 3/9 = 0.0082

P(HAYIR) = 3/5 1/5 4/5 3/5 = 0.0577

Verilen bir sınıfın toplam olasılığını hesaba katılır ve özelliklerin olasılıklarıyla çarpılır ;

P(EVET) = 0.0082 . 9/14 =0.0053

P(HAYIR) =0.0577. 5/14 =0.0206

Bu olasılığı maksimum yapan sınıf seçilirse, yeni durum “HAYIR” olarak sınıflandırma anlamına gelecektir.