VERİ BİLİMİ İÇİN TEMEL İSTATİSTİK

hafta-6

CEMİLE YILDIZÇAKAR

19.01.2021

İçindekiler

Bütüne Tamamlayan Olaylar (Collectively Exhaustive Events)

- E1, E2, … Ek olaylarının kesişimleri boş küme, birleşimleri örnek uzaya eşit ise bu olaylar bütüne tamamlayan olaylardır. (E1 U E2 U . . . U Ek = S)

Mutually exclusive and exhaustive system of events : Let S be the sample space associated with a random experiment. Let A_1, A_2 ...... A_n be subsets of S such that

(i) $A_i \cap A_j = \phi$ for $i \neq j$ and

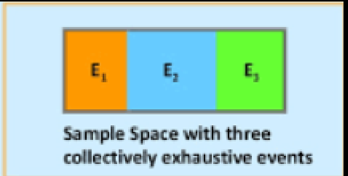
(ii) $A_1 \cup A_2 \cup \ . \ . \ . \ \cup A_n = S$

Then the collection of events A_1, A_2, ..., A_n is said to form a mutually exclusive and exhaustive system of events.

In this sytem,

$$P(A_1 \cup A_2 \cup \ . \ . \ . \ \cup A_n) = P(A_1) + P(A_2) + \ . \ . \ . \ + P(A_n) = 1$$

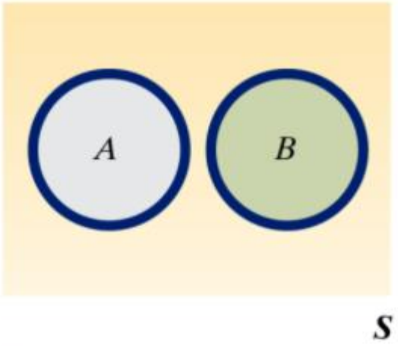
$$E_1 \cup E_2 \cup E_3 \cup \cdots \cup E_n = \bigcup_{i=1}^{n} E_i = S$$



Karşılıklı Ayrık Olaylar (Mutually Exclusive Events)

Events A and B are **mutually exclusive** if $A \cap B$ contains no sample points—that is, if A and B have no sample points in common. For mutually exclusive events,

$$P(A \cap B) = 0$$



Probability of Union of Two Mutually Exclusive Events

If two events A and B are *mutually exclusive*, the probability of the union of A and B equals the sum of the probability of A and the probability of B; that is, $P(A \cup B) = P(A) + P(B)$.

Olasılık Teorisi

- Klasik (A Priori) Olasılık
- Frekans (A Posteriori) Olasılığı
- Aksiyom Olasılığı.

Olasılığın tarihsel gelişim aşamalarıdır.

KLASİK OLASILIK

- Eğer bir örnek uzayı $n(S)$ adet ayrık ve eşit olasılıkla ortaya çıkan basit olaylardan oluşuyor ve örnek uzayındaki basit olaylardan $n(A)$ adedi A olayının özelliğine sahip ise A'nın olasılığı:

$$P(A) = n(A) / n(S)$$

kesri ile elde edilir

KLASİK OLASILIK NEDEN YETERSİZDİR?

- Örnek uzayının eleman sayısı sonsuz olduğu durumlarda,
- Eşit olasılıklı olay varsayımı yapılamadığı durumlarda ,
  - *Tümdengelim çıkarımları yapılamadığında*

klasik olasılık ile hesaplama yapılamayacağından dolayı yetersizdir.

FREKANS OLASILIĞI

- Araştırılan anakütle üzerinde n adet deney uygulanır. Yapılan bu deneylerde ilgilenilen A olayı n(A) defa gözlenmiş ise A olayının göreli frekansı (yaklaşık olasılığı):

$$P(A) = n(A) / n$$

olarak bulunur.

ÖRNEK:

Bir fabrikanın üretmiş olduğu televizyonların hatalı olma olasılığı p nedir?

Önce örnek uzayı oluşturulur:

S={sağlam,hatalı}

Klasik olasılığa göre (eşit olasılıklı olaylar) p=0.5 olup gerçeği yansıttığı şüphelidir.

Yapılması gereken örneklem alarak

$$p = n(H) / n$$

olasılığını hesaplamaktır.

FREKANS OLASILIĞIN KARARLILIK ÖZELLİĞİ

- Gerçekleştirilen deney sayısı arttıkça P(A) olasılık değerindeki değişkenlik azalacak ve giderek bir sabit değere yaklaşacaktır. Bu duruma ***kararlılık özelliği*** adı verilir.
- Bir olayın olasılığı deneyin tekrarlama sayısı sonsuza yaklaşırken o olayın göreli frekansının alacağı limit değer olarak tanımlanır:

$$p = P(A) = \lim_{n \to \infty} n(A) / n$$

FREKANS OLASILIĞI NİÇİN YETERSİZDİR

- Olasılığın kararlılık değerine ulaştığı deneme sayısı kaçtır?
- Sonsuz adet deneme yapmak mümkün değildir.
- Aynı deney iki defa aynı tekrar sayısı ile gerçekleştirildiğinde elde edilen olasılıklardan hangisi olayın olasılığı olarak kabul görecektir?

Aksiyom Olasılığı Nedir?

- Olasılığın matematiksel teorisini tanımlar.
- Bu teorinin oluşturduğu ideal modeller yaşadığımız dünyanın problemlerini çözmede kullanılır.
- Klasik ve frekans Olasılık teorisi için ortak nokta; her ikisinin de, benzer koşullarda (teorikte aynı koşullarda) uygulanan deneylere gereksinim duyar. Buna karşın tekrarlı olarak uygulanamayan durumlarda olasılıkların hesaplanmasında AKSİYON OLASILIĞI yardımcı olur.

BENZER KOŞULLARDA TEKRARLI OLARAK UYGULANAMAYAN DURUMLARA ÖRNEK;

- Çok hoşlandığınız bir kişi ile çıkma olasılığı nedir?
- Karşıyaka – Göztepe maçının 6-0 bitme olasılığı nedir?
- 3. Dünya savaşının çıkma olasılığı nedir?

KOLMOGOROV AKSİYOMLARI

- BİRİNCİ AKSİYOM:

Bir olayın olasılığı bir negatif-olmayan reel sayıdır ve bu sayı şöyle ifade edilir:

$$P(E) \geq 0 \qquad \forall E \subseteq F$$

Burada F olay uzayıdır.

İKİNCİ AKSİYOM

Bu **birim-ölçüsü** varsayımıdır: Örnekleme uzayının tümünü kapsayan bir basit olay ortaya çıkması için olasılık **1**dir. Daha belirli bir şekilde ifadeyle; Örneklem uzayını taşan hiçbir basit olay mümkün değildir:

$$P(\Omega) = 1.$$

Bu aksiyom bazı hatalı olasılık hesaplamalarında çok kere temel bir hatanın ortaya çıkmasına neden olmuştur. Eğer tüm örneklem uzayı kesinlikle tanımlanamıyorsa bunun herhangi bir alt setinin tanımlanması da imkânsızdır.

ÜÇÜNCÜ AKSİYOM

Bu σ-toplanabilirlik varsayımıdır. Herhangi bir ikişerli bağlantısız ortaya çıkan sayılabilir olaylar dizisi, $E_1, E_2, \ldots$ şu eşitliği tatmin eder:

$$P(E_1 \cup E_2 \cup \cdots) = \sum_i P(E_i).$$

Bazı yazarlar sadece sonsuz olmayan-toplanabilir olasılık uzaylarını ele alırlar.

KOŞULLU OLASILIK(CONDITIONAL PROBABILITY)

- Koşullu olasılık, bir olayın, başka bir olayın gerçekleştiği biliniyor iken, gerçekleşmesi olasılığıdır.

- A olayının gerçekleştiği bilindiğinde B olayının olması olasılığına, B olayının A olayına göre koşullu olasılığı ya da B olayının koşullu olasılığı denir.

$$P(B|A)$$

- $(B|A)$ olayı, A ile kısıtlanmış örneklem uzayında tanımlanmış bir olaydır.

Koşullu olasılık P(A|B) biçiminde gösterilir. P(A|B)'nın anlamı, B gibi bir olayın gerçekleştiği bilindiğinde A olayının olasılığı olarak ifade edilir.

A ve B, aynı örneklem uzayında tanımlanmış iki olay ve P(B)>0 olmak üzere B olayının gerçekleştiği varsayımı altında A olayının koşullu olasılığı

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

Biçiminde tanımlanır. Bu tanımdan yararlanılarak A ve B olaylarının birlikte gerçekleşme olasılığını koşullu olasılık yardımı ile

$$P(A \cap B) = P(A|B).P(B)$$

biçiminde bulabiliriz.

Örnek: Okulumuz öğrencilerinden %45'i istatistik, %35'i bilgisayar derslerinden ve %25'i hem istatistik hem de bilgisayar derslerinden başarısızdır. Rasgele seçilen bir öğrencinin,

a) Bilgisayardan başarısız ise, istatistikten de başarısız olma olasılığını,
b) İstatistikten başarısız ise, bilgisayardan da başarısız olma olasılığını,
c) Bu iki dersten en az birinden başarısız olma olasılığını bulunuz.

İ, istatistik dersinden başarısız öğrencileri; ve B, bilgisayar dersinden başarısız öğrencileri göstersin.

$$P(İ)=0.45, \qquad P(B)=0.35 \qquad \text{ve } P(İ\cap B) = 0.25$$

a) $$P(İ|B) = \frac{p(İ \cap B)}{P(B)} = \frac{0.25}{0.35} = \frac{5}{7}$$

b) $$P(B|İ) = \frac{P(B \cap İ)}{P(İ)} = \frac{0.25}{0.45} = \frac{5}{9}$$

c) $$P(İ \cup B) = P(İ) + P(B) - P(İ \cap B) = 0.45 + 0.35 - 0.25 = 0.55$$

Bulunur.

İki tabak dolusu bisküvi düşünülsün; tabak #1 içinde 10 tane çikolatalı bisküvi ve 30 tane sade bisküvi bulunduğu kabul edilsin. Tabak #2 içinde ise her iki tip bisküviden 20şer tane olduğu bilinsin. Evin küçük çocuğu bir tabağı rastgele seçip bu tabaktan rastgele bir bisküvi seçip alsın. Çocuğun bir tabağı diğerine ve bir tip bisküviyi diğerine tercih etmekte olduğuna dair elimizde hiçbir gösterge bulunmamaktadır. Çocuğun seçtiği bisküvinin sade olduğu görülsün. Çocuğun bu sade bisküviyi tabak #1 den seçmiş olmasının olasılığının ne olacağı problemi burada incelenmektedir.

Sezgi ile, tabak #1de sade bisküvi sayısının çikolatalı bisküvi sayısına göre daha fazla olduğunu göz önüne alınırsak incelenen olasılığın %50'den daha fazla olacağı hemen algılanır. Bu soruya cevap Bayes teoremi kullanarak kesin olarak verilebilir.

Önce soruyu değiştirip Bayes teoremi uygulanabilecek şekle sokmak gerekmektedir: Çocuğun bir sade bisküvi seçmiş olduğu bilinmektedir; o halde bu koşulla birlikte tabak #1den seçim yapması olasılığı ne olacaktır?

Böylece Bayes teoremi formülüne uymak için *A* olayı çocuğun tabak #1den seçim yapması; *B* olayı ise çocuğun bir sade bisküvi seçmesi olsun. İstenilen olasılık böylece Pr(*A*|*B*) olacaktır ve bunu hesaplamak için şu olasılıkların bulunması gerekir:

- Pr(*A*) veya hiçbir diğer bilgi olmadan çocuğun tabak #1'den seçim yapması olasılığı;

İki tabak arasında tercih olmayıp seçimin eşit olasılığı olduğu kabul edilmektedir.

- Pr(*B*) veya hiçbir diğer bilgi olmadan çocuğun bir sade bisküvi seçmesi olasılığı: Diğer bir ifade ile, bu çocuğun her bir tabaktan bir sade bisküvi seçme olasılığıdır. Bu olasılık, önce her iki tabaktan ayrı ayrı olarak seçilen bir tabaktan bir sade bisküvi seçme olasılığı ile bu tabağı seçme olasılığının birbirine çarpılması ve sonra bu iki çarpımın toplanması suretiyle elde edilir. Tabaklarda olan sade bisküvinin sayısının toplama orantısından bilinmektedir ki tabak #1'den bir sade bisküvi seçme olasılığı (30/40=) 0,75; tabak #2'den sade bisküvi seçme olasılığı (20/40=) 0,5 olur. Her iki tabaktan seçme olasılığı ise her tabak aynı şekilde uygulama gördüğü için 0,50 olur. Böylece bu problemin tümü için bir sade bisküvi seçme olasılığı 0.75×0.5 + 0.5×0.5 = 0.625 olarak bulunur.
- Pr(*B*|*A*), veya çocuğun tabak #1'den seçim yaptığı bilirken bir sade bisküvi seçmesi.: Bu 0,75 olarak bilinmektedir çünkü tabak #1'deki toplam 40 bisküviden 30'u sade bisküvidir.

Şimdi bu açıklanan tüm olasılık değerleri Bayes teoremi formülüne konulabilir:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.75 \times 0.5}{0.625} = 0.6$$

Böylece çocuğun sade bisküvi seçimi bilindiğine göre tabak #1'den alma olasılığı %60'tır ve sezgimize göre seçtiğimiz %50'den daha büyüktür.

Çarpım Kuralı (Multiplication Rule)

- A ve B gibi birlikte ortaya çıkan olayların olasılığı $P(A \cap B)$ şeklinde ifade edilir. İki olayın kesişiminin olasılığı, bir olayın olasılığı ile ikinci olayın koşullu olasılığının çarpımına eşittir ve bu kurala çarpma kuralı denir.

$$P(A \cap B) = P(A) * P(B|A)$$

Koşullu olasılık $P(B|A)$ biçiminde gösterilir. $P(B|A)$: A olayının gerçekleştiği bilindiğinde B olayının olma olasılığını ifade eder.

Koşullu Olasılık Ve Bağımsız Olaylar

Olasılık konusunda iki olayın bağımsız olduğunu söylediğimizde, bir olayın olmasının diğer olayın olasılığını değiştirmediğini söylüyoruz.

Örneğin, hilesiz bir paranın yazı tura atışında "tura" gelme olasılığı $1/2$'dir. Peki ya yazı tura atılan günün Salı günü olduğunu bilseydik? Bu, "tura" gelmesi olasılığı değiştirir miydi? Tabii ki değiştirmezdi. Salı olduğu bilindiğinde, "tura" gelme olasılığı hala $1/2$'dir. O zaman, yazı tura atışının sonucuyla günün Salı olması bağımsız olaylardır; Salı olduğunu bilmek, "tura" gelme olasılığını değiştirmez.

Her durum bu kadar açık değildir. Ya cinsiyet veya hangi elin kullanıldığı (solak ve sağlak)? Bir kişinin cinsiyetiyle solak veya sağlak olması tamamen bağımsız olaylar gibi görünmektedir. Yine de olasılıklara baktığımızda, tüm kişilerin $\%10$'unun solak, ama erkeklerin yaklaşık $\%12$'sinin solak olduğunu görürüz. O zaman, bu olaylar bağımsız değildir, çünkü rastgele bir kişinin erkek olduğunu bilmek onun solak olma olasılığını artırır.

Buradaki temel fikir şu: Olasılıkta olayların bağımsız olup olmadığına bakarız.

Eğer $P(\mathrm{A} \mid \mathrm{B}) = P(\mathrm{A})$ ve $P(\mathrm{B} \mid \mathrm{A}) = P(\mathrm{B})$ ise, iki olay bağımsızdır.

- A ve B olaylarından birinin elde edilmesi, diğerinin elde edilmesini etkilemiyorsa A ve B bağımsız olaylardır.
- Eğer aşağıdaki eşitlik sağlanıyorsa, A ve B olayları istatistiksel olarak bağımsızdır.

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

- Gerçek hayattan veri setlerinde olayların bağımsızlığına baktığımızda, mükemmel eşitlikte olasılıklar elde etmek çok nadir görülür. Şans oyunları içermeyen neredeyse tüm gerçek olaylar bir derece bağımlıdır.
- Pratikte, genelde olayların bağımsız olduğunu varsayarız ve örneklem verisi üzerinde bu varsayımı test ederiz.
- Son olarak, veriler iyi tasarlanmış bir deneyden gelmediği sürece, neden sonuç ilişkisi hakkında bir sonuca varmaktan kaçının.st ederiz. Olasılıklar çok farklıysa, olayların bağımsız olmadığı sonucuna varırız.

BAYES TEOREMİ

Bayes ağları, bir rastlantı değişkenleri kümesinin çok değişkenli olasılık dağılımlarını etkili bir biçimde göstermeye ve modellemeye yarayan bir kavramdır. Bir olayın meydana gelmesinde birden fazla etkenin olması koşulunda, olayın hangi etkenin etkinliği ile ortaya çıktığını gösteren bu teoremde, rassal bir sürece bağlı olarak ortaya çıkan rasgele bir *X* olayı ile diğer bir rasgele *Y* olayı için koşullu olasılıklar ve marjinal olasılıklar arasındaki ilişki tanımlanır. Bu ilişkiyi ilk kez Thomas Bayes ortaya atmış ve aşağıdaki eşitliği önermiştir .

$$P(Y|X) = \frac{P(Y \cap X)}{P(X)} = \frac{P(X|Y)P(Y)}{P(X)}$$

Naive Bayes Sınıflandırıcı

Sınıflandırıcı nedir?

Sınıflandırıcı, belirli özellikleri temel alan farklı nesneleri ayırt etmek için kullanılan bir makine öğrenmesi modelidir.

Naive Bayes Sınıflandırma Modeli

Bir sınıflandırma problemi bir çok özellikten ve bir sonuç (hedef) değişkeninden oluşur.

$$p(C|F_1,...F_n) = \frac{P(C)p(F_1,...F_n|C)}{p(F_1,...F_n)}$$

C verilen hedef ve F özelliklerimiz temsil eder. Naive bayes sınıflandırıcı basitçe bütün koşullu olasılıkların çarpımıdır.

Hava durumu üzerinden futbol oynayıp oynamayacağımıza karar vermeye çalışalım.

Özellikler	Hedef
Hava Durumu	Futbol Oyna
Yağmurlu	Hayır
Yağmurlu	Hayır
Bulutlu	Evet
Güneşli	Evet
Güneşli	Evet
Güneşli	Hayır
Bulutlu	Evet
Yağmurlu	Hayır
Yağmurlu	Evet
Güneşli	Evet
Yağmurlu	Evet
Bulutlu	Evet
Bulutlu	Evet
Güneşli	Hayır

		Futbol Oyna	
		Evet	Hayır
Hava Durumu	Güneşli	3	2
	Bulutlu	4	0
	Yağmurlu	2	3

		Futbol Oyna		
		Evet	Hayır	
Hava Durumu	Güneşli	3/9	2/5	5/14
	Bulutlu	4/9	0/5	4/14
	Yağmurlu	2/9	3/5	5/14
		9/14	5/14	

Hava güneşli olduğunda futbol oynama olasılığını tahmin edelim. Bunun için yukarıdaki tablolardan hesaplamalar yapacağız.

Beklenti 1: Güneşliyken Futbol Oyna Evet = P(Evet | Güneşli) = P(Güneşli | Evet) * P(Evet) / P (Güneşli)

P (Güneşli |Evet) = 3/9 = 0.333, P(Güneşli) = 5/14 = 0.357, P(Evet)= 9/14 = 0.643

P (Evet | Güneşli) = 0.333 * 0.643 / 0.357 = 0.600

Beklenti 2: Güneşliyken Futbol Oyna Hayır = P(Hayır | Güneşli) = P(Güneşli | Hayır) * P(Hayır) / P (Güneşli)

P (Güneşli | Hayır) = 2/9 = 0.222, P(Güneşli) = 5/14 = 0.357, P(Hayır)= 5/14 = 0.357

P (Hayır | Güneşli) = 0.222 * 0.357 / 0.357 = 0.222

Son aşamada beklenti 1 ile beklenti 2 kıyaslanır. Beklenti 1 daha büyük değere sahip olduğu seçili
Naive Bayesian Classifier havayı güneşli gördüğünde futbol oynaya izin verir.

Örnek Bir tez;

- “E-Ticaret Sistemlerinde Reklam Ürünlerinin Bayes Teoremine Göre Yerleştirilmesi” - Mehmet Akif BÜLBÜL
- Bu tez çalışmasında e-ticaret ile alışveriş yapan üye ziyaretçilerin ürünler üzerindeki her bir hareketi kayıt altına alınmış ve bu kayıtlar analiz edilerek kullanıcının bir sonraki ürün satın alma veya ürün inceleme ihtimali yüksek olan ürün grubu Bayes Teoremi ile hesaplanmıştır.
- Sunulan çalışmada, reklam ürünlerinin dinamik yerleştirilmesi ve öngörü sistemi ile web tabanlı kişiselleştirilmiş prototip yapısı Bayes Teoremi kullanılarak oluşturulmaktadır. Ayrıca, önceki araştırmalarda kullanıcıların tercihleri/girişleri manuel olarak gerçekleştirilmiştir. Bu çalışma sayesinde kullanıcıların site içi anlık hareketleri ile ürün inceleme veya ürün satın alma arasındaki ilişki otomatik olarak Bayes Teoremi kullanarak hesaplanmakta ve kişiselleştirilmiş dinamik ürün yerleştirme ile özgün bir model ortaya konulmaktadır.

1. Başla
2. Eğer kullanıcı sisteme giriş yaptı ise 4. Adıma git
3. Rastgele reklam ürünü yerleştir 10. Adıma git
4. Kullanıcı hareketlerini veritabanından süz
5. Reklam ürün gruplarını Bayes Teoremi ile hesapla
6. En yüksek oranlı ürün grubunu belirle
7. Web arayüzündeki reklam alanlarını belirlenen ürün grubu ürünleri ile güncelle
8. Eğer ziyaretçi kategori değiştirdi ise 4. Adıma geri dön
9. Eğer ziyaretçi ürün aldı veya ürün baktı ise 4. Adıma geri dön
10. Bitir

Teşekkür Ederim

LinkedIn
https://www.linkedin.com/in/cemile-yıldızçakar-34782248/

Email
yildizcakar.cemile@gmail.com

Cemile YILDIZÇAKAR