Olasılık ve İstatistik

Fırat İsmailoğlu, PhD

BAYES Teoremi

BAYES Teoremi

Çarpım kuralında $P(A \cap B) = P(B)P(A|B)$ idi. $P(A \cap B)$, benzer şekilde $P(A)$ ile $P(B|A)$ olasılıklarının çarpımı şeklinde de yazılır: $P(A \cap B) = P(A)P(B|A)$. Buradan

$$P(B)P(A|B) = P(A)P(B|A)$$

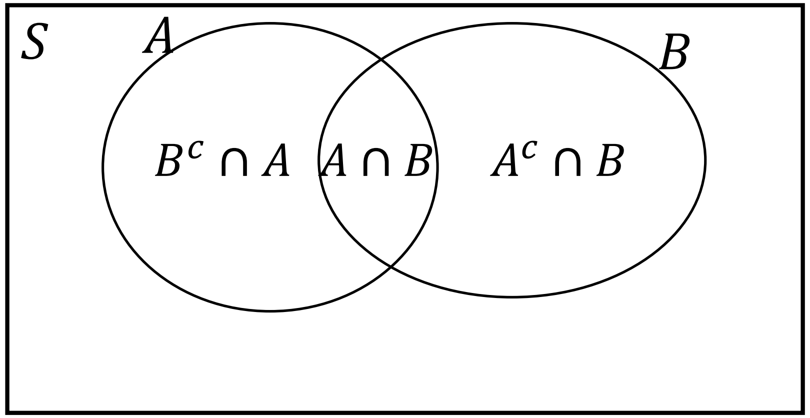
olur. $P(A|B)$ yalnız bırakılırsa:

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

olur. Bu eşitliğe Bayes teoremi denir.

Burada, A yerine neden, B yerine sonuç yazılırsa, Bayes teoreminin neden bu kadar önemli olduğunu anlayabiliriz. Bu durumda teorem şöyle bir hal alır:

BAYES Teoremi



B kümesi ayrık olan (kesişimleri olmayan) $A \cap B$ ve $A^c \cap B$'nin birleşimi şeklinde yazılabilir:

$B = (A \cap B) \cup (A^c \cap B)$.

O halde B'nin olasılığı:

$$P(B) = P(A \cap B) + P(A^c \cap B)$$

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

olur.

Bayes teoremi yeniden düzenlenirse:

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}}$$

ör. Bir klinikte yapılan kanser testi gerçekte kanser olan hastaların %98’inde pozitif olarak sonuç veriyor, yani kanser diyor. Bu test, gerçekte kanser olmayan hastaların %97’sinde negatif olarak sonuç veriyor, yani kanser değilsin diyor. Ayrıca toplumdaki kişilerin 0.008’nin kanser olduğu biliniyor.

Buna gore kanser testi pozitif çıkmış bir kişinin gerçekte kanser olma olasılığı nedir?

Çözüm.

$P(kanser) = 0.008$ (kişi hakkında hiçbir bilgi yokken, kisinin kanser olma olasiligi).

$P(\neg kanser) = 0.992$ (kisinin kanser olmama olasılığı)

(kişi hakkında hiçbir bilgi yokken, kisinin kanser olma olasiligi).

$P(test\ pozitif|kanser) = 0.98$ (kişinin kanser olduğu bilinirken, testin pozitif çıkması olasılığı).
$P(test\ negatif|kanser) = 0.02$

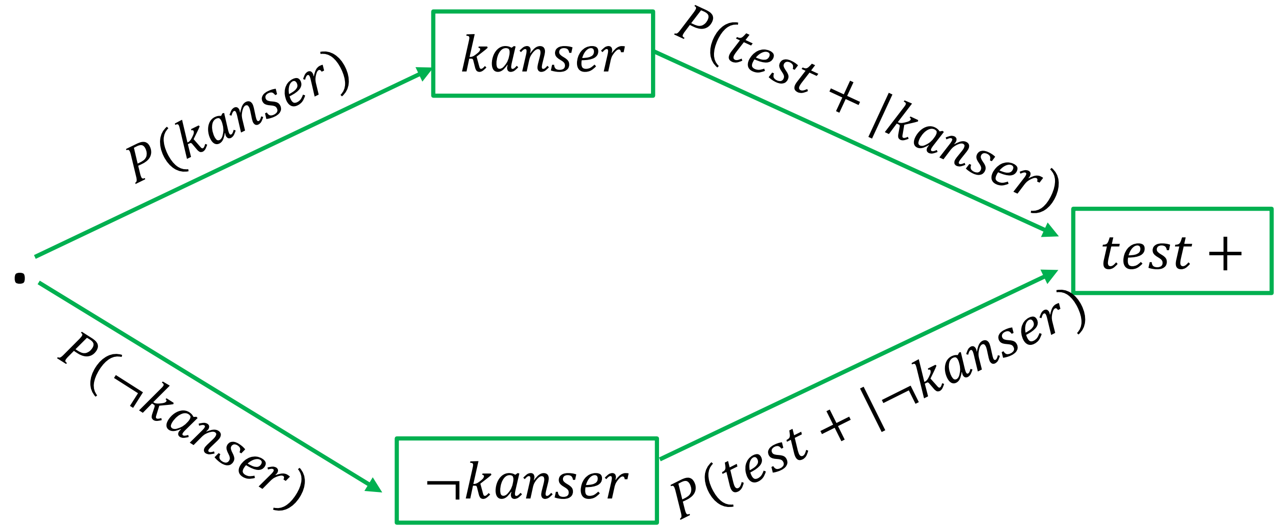
$P(test\ negatif|\neg kanser) = 0.97$ (kişi kanser degilken, testin negatif olması)

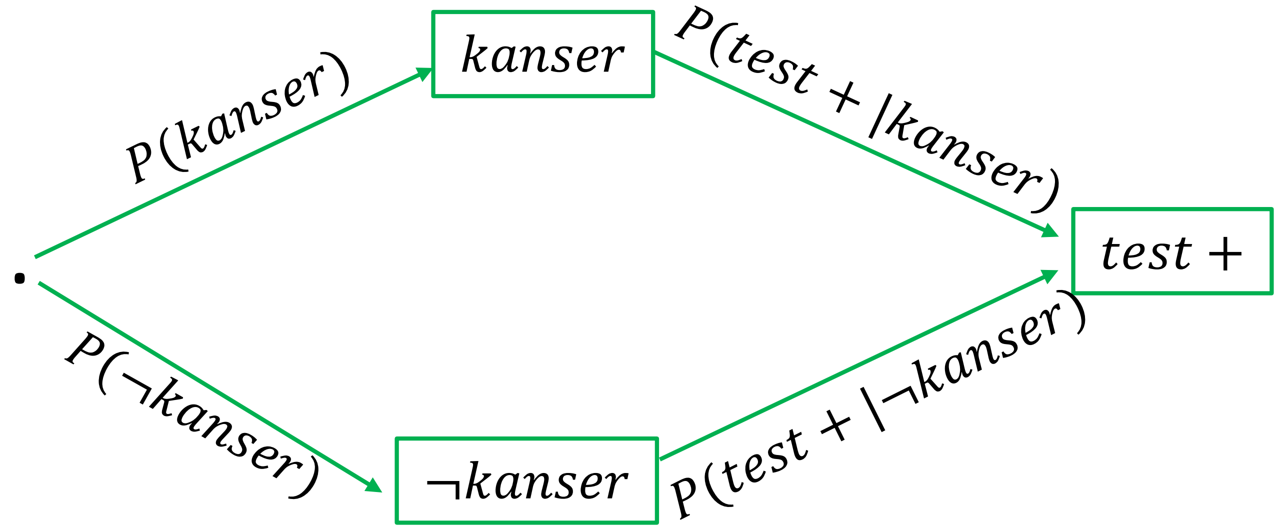
$P(test\ pozitif|\neg kanser) = 0.03$ (kisi kanser degilken testi pozitif)
$P(kanser|test\ pozitif) = ?$

Bayes teoreminden:

$$P(kanser|test\ pozitif) = \frac{P(test\ pozitif|kanser)P(kanser)}{P(test\ pozitif)}$$

Yukarıda bilmediğimiz şey $P(test\ pozitif)$. Bununla ilgili direkt bir bilgi yok, ama testin pozitif çıkması kanserli kişilerde veya kanserli olmayan kişilerde görülür.





$$P(test\ pozitif) = P(test\ pozitif, kanser) + P(test\ pozitif, \neg kanser) =$$
$$P(test\ pozitif|kanser)\cdot P(kanser) + P(test\ pozitif|\neg kanser)\cdot P(\neg kanser)$$
$$= 0.98 \times 0.008 + 0.03 \times 0.992$$
$$= 0.0376$$

$$P(kanser|test\ pozitif) = \frac{0.98 \times 0.008}{0.0376} = 0.2$$

Şu halde testi pozitif çıksada bir kişi endişe etmesine gerek yoktur, test güvenilir değildir.

BAYES Teoremi

$$P(neden|sonuç) = \frac{P(sonuç|neden)P(neden)}{P(sonuç)}$$

Genelde elimizde sonuç (yani gözlem, yani veri) vardır; ve biz tüm nedenler içersinde bu sonuca neden olabilecek en olası nedeni bulmak isteriz. Öte yandan çoğu kez $P(sonuç|neden)'$i bulmak, $P(neden|sonuç)$'i bulmaktan daha kolaydır.

Diyelimki gözlemimiz (yani sonuç) öksürük. Buna neden olabilecek iki neden olsun: grip ve zatürre. Zatürre varken kişinin öksürmesi, grip varken kişinin öksürmesinden daha olasıdır. $P(öksürük|zatürre) > P(öksürük|grip)$

Fakat genel olarak grip görülme olasılığı, zatürre görülme olasılğından çok daha yüksektir. $P(grip) \gg P(zatürre)$.

Bu yüzden teoremdeki $P(neden)$ katsayısı $P(sonuç|neden)'$i dengeler.

BAYES Teoremi

$$P(neden|sonuç) = \frac{P(sonuç|neden)P(neden)}{P(sonuç)}$$

Genelde elimizde sonuç (yani gözlem, yani veri) vardır; ve biz tüm nedenler içersinde bu sonuca neden olabilecek en olası nedeni bulmak isteriz. Öte yandan çoğu kez $P(sonuç|neden)'$i bulmak, $P(neden|sonuç)$'i bulmaktan daha kolaydır.

Diyelimki aynı sonuca yol açabilcek iki farklı neden olsun: $neden\_1$ ve $neden\_2$. Ve biz bunlardan en olası nedeni ($P(neden|sonuç)$ değeri en yüksek nedeni) bulmak istiyoruz. Bu durumda elimizde

$$P(neden\_1|sonuç) = \frac{P(sonuç|neden\_1)P(neden\_1)}{P(sonuç)}$$

ve $P(neden\_2|sonuç) = \frac{P(sonuç|neden\_2)P(neden\_2)}{P(sonuç)}$

Karşılaştırmak istediğimiz iki değerinde paydası aynıdır: $P(sonuç)$. O halde bu paydayı nedenleri karşılaştırıp, en olası nedeni ararken yok sayabilriz. Nedenleri karşılaştırmada $P(sonuç|neden)P(neden)$ çarpımına bakmak yeterli olur. Sonuç olarak $P(neden|sonuç)$, $P(sonuç|neden) \times P(neden)$ ile orantılıdır:

$$\boxed{P(neden|sonuç) \propto P(sonuç|neden) \times P(neden)}$$

Diyelimki gözlemimiz (yani sonuç) öksürük. Buna neden olabilecek iki neden olsun: grip ve zatürre. Zatürre varken kişinin öksürmesi, grip varken kişinin öksürmesinden daha olasıdır. $P(\text{öksürük}|zatürre) > P(\text{öksürük}|grip)$

Fakat genel olarak grip görülme olasılığı, zatürre görülme olasılğından çok daha yüksektir. $P(grip) \gg P(zatürre)$.

Bu yüzden $P(neden)$ katsayısı, $P(sonuç|neden)$'i dengeler.

$P(\text{öksürük}|zatürre)P(zatürre)$ Vs $P(\text{öksürük}|grip)P(grip)$.

ör. Iki futbol takimi duşunun: T1 ve T2. Varsayalimki T1, T2 ile yaptigi maclarin %65'ni kazanmış olsun. T1'in T2'yi yendiği maclarin %70'i kendi sahasinda gerçekleşmiş olsun. Öte yandan T2'nin T1'i yendiği maclarin %75'i kendi sahasinda gerçeklemiş olsun. Eğer gelecek maç T1'in sahasinda olacaksa bu maçı hangi takimin kazanmasi daha olasidir?

Çözüm.

$P(T1'in\ kazanmasi) = 0.65$
$P(T2'nin\ kazanmasi) = 0.35$
$P(T1'in\ sahasi|T1'in\ kazanmasi) = 0.7$
$P(T2'nin\ sahasi|T2'nin\ kazanmasi) = 0.75$
$P(T1'in\ sahasi|T2'nin\ kazanmasi) = 0.25$

$P(T1'nin\ kazanmasi|T1'in\ sahasi)$ Vs $P(T2'nin\ kazanmasi|T1'in\ sahasi)$

$P(T1'nin\ kazanmasi|T1'nin\ sahasi),\quad P(T1'in\ sahası\ |T1'in\ kazanması) \times$
$P(T1'in\ kazanmasi)$ ile orantılıdır.

$$P(T1'nin\ kazanmasi|T1'nin\ sahasi) \propto 0.7 \times 0.65 = 0.455$$

$P(T2'nin\ kazanmasi|T1'nin\ sahasi)$, $P(T1'in\ sahası\ |T2'in\ kazanması) \times P(T2'in\ kazanmasi)$ ile orantılıdır.

$$P(T2'nin\ kazanmasi|T1'nin\ sahasi) \propto 0.25 \times 0.35 = 0.0875$$

olup,T1'in kazanması daha olasıdır.

ör. Diyelimki covid'li birinin PCR testinin pozitif çıkmasının olasılığı 0.8; covid'li olmayan birinin testinin negatif çıkma olasılığı 0.9 olsun. Ayrıca toplumdaki kişilerin %10'unun covid'li olduğu bilinsin.

Hasta olduğundan şüphelenen bir kişi üç farklı hastanede birbirinden bağımısız olarak PCR testi yaptırıyor. Testlerin ikisinin negatif birinin pozitif geldiği biliniyorsa, bu kişinin gerçekten covid'li olma olsalığı ne olur?

Çözüm.

$P(test + |\ covid) = 0.8$ ise aynı zamanda bize $P(test - |\ covid) = 0.2$ bilgisi de verilmişti; çünkü kişi ğer covid'li ise testi ya pozitiftir yada negatiftir, bu iki olay birbirni tamamlar.

$P(test - |\ \neg covid) = 0.9$ ise benzer şekilde $P(test + |\ \neg covid) = 0.1$ elde edilir.

$P(covid) = 0.1$ ve $P(\neg covid) = 0.9$. (Bu iki olasılık testten bağımısız, saf olasılıktır.)

Elimizdeki sonuç (gözlem): $+, -, -$. Işte biz bu sonuç varken pozitif olma olasılığını merak ediyoruz, yani $P(covid|+,-,-)$. Bayes teoreminden

$$P(covid|+,-,-) = \frac{P(+,-,-|covid)P(covid)}{P(+,-,-)}$$

Paydadaki $P(+,-,-)$ olasılığını hesaplarken şunu düşüneceğiz. Kişi covid iken bu sonucu gözlemleyebiliriz, bunun olasılığı:$P(+,-,-|covid)P(covid)$

Yada,

Kişi covid'li değilken bu sonucu gözlemleyebiliriz. Bunun olasılığı:$P(+,-,-|\neg covid)P(\neg covid)$.

Yukarıdaki iki durum (olay) aynı anda olması mümkün olmadığından (kişi ya covid'dir yada değildir), bu iki olaydan herhanagi birinin olma olaslığı, yani payda:

$$P(+,-,-) = P(+,-,-|covid)P(covid) + P(+,-,-|\neg covid)P(\neg covid).$$

Sonuç olarak aradığımız olasılık:

$$P(covid|+,-,-) = \frac{P(+,-,-|covid)P(covid)}{P(+,-,-|covid)P(covid) + P(+,-,-|\neg covid)P(\neg covid).}$$

Şimdi $P(+,-,-|covid)$ olasılığını düşünelm. Soruda testlerin biribirinden bağımısz olduğunu söylemiştik. Bağımısz olayların aynı anda (berbaer) olma olsılıkları ayrı ayrı olma olasıkıklarının çarpımına eşitti: $P(A \cap B) = P(A)P(B)$. Burada fazladan bir covid olma durumu verilmiş. Bunu olasılıkları çarparken koruyacağız. Yani şöyle bir şey:

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

O halde:

$$P(+,-,-|covid) = P(+|covid)P(-|covid)P(-|covid)$$

Benzer şekilde

$$P(+,-,-|\neg covid) = P(+|\neg covid)P(-|\neg covid)P(-|\neg covid)$$

$$P(covid|+,-,-) = \frac{0.8 \cdot 0.2 \cdot 0.2 \cdot 0.1}{0.8 \cdot 0.2 \cdot 0.2 \cdot 0.1 + 0.1 \cdot 0.9 \cdot 0.9 \cdot 0.9} \approx 0.04$$

ör. Neden ikiden fazla olabilir. Örneğin kişinin öksürduğunü gözlemleyelim (sonuç). Buna neden olabilecek üç tane neden belirleyelim: Grip, zatürre, covid. Şu halde üç durumdan biri olabilir:

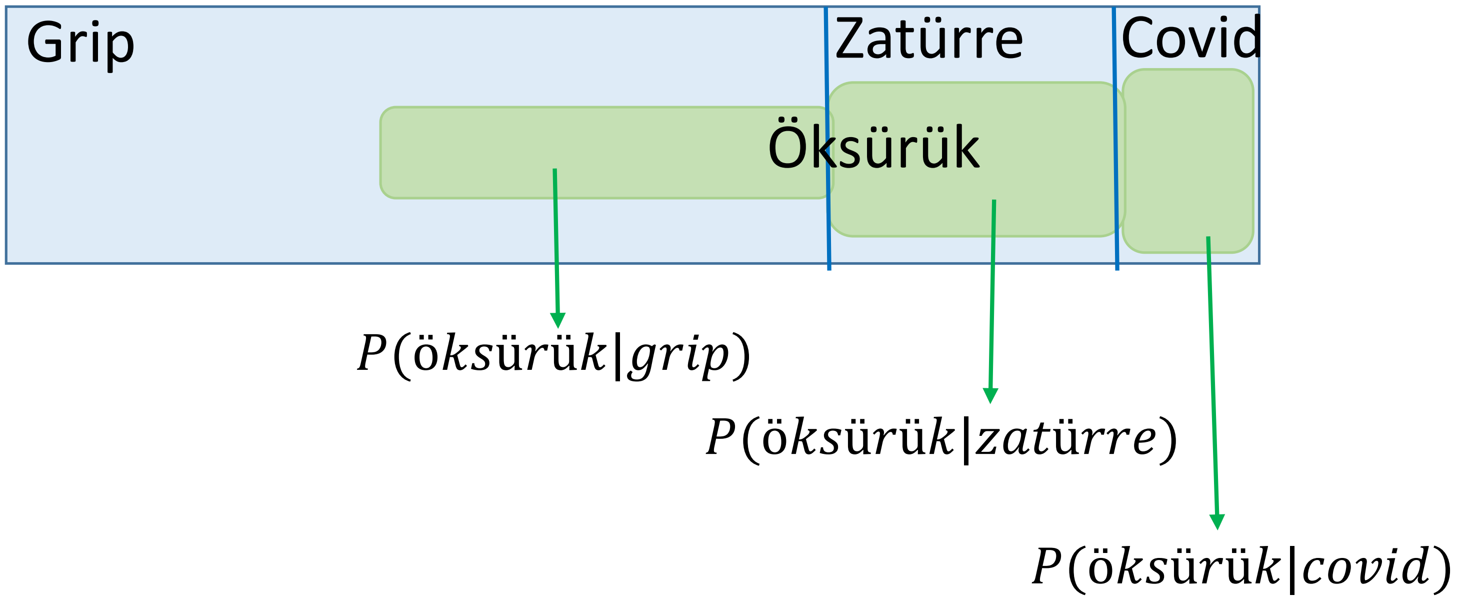
i. kişi griptir <u>ve</u> gripken öksürür: $P(grip)P(öksürük|grip)$

ii. kişi zatürredir <u>ve</u> zatürre iken öksürür: $P(zaturre)P(öksürük|zaturre)$

iii. kişi covid'dir <u>ve</u> covid iken öksürür: $P(covid)P(öksürük|covid)$

Bu üç durumdan herhangi biri ise kişi öksürür:

$$P(öksürük) = P(grip)P(öksürük|grip) + P(zaturre)P(öksürük|zaturre) + P(covid)P(öksürük|covid)$$



Sonuç (öksürük) yüzeyde nedenler altta.

Genellersek, diyelim ki bir A sonucuna (gözlemine) neden olabilecek n tane muhtemel neden olsun: $B_1, B_2, \dots, B_n$. Bu durumda A sonucunu gözlemleme olasılığımız (Bayes'de paydaya denk geliyor bu), her bir nedenin olma olasılığı ile bu nedenin elimizdeki sonuca yol açma olasılığının çarpımlarının toplanmasıyla bulunur:

$$P(A) = \sum_{i=}^{n} P(A|B_i)P(B_i)$$