Niemożliwe lub nieprawdopodobne - delikatne wprowadzenie do prawdopodobieństwa

W następnych kilku częściach będziemy badać zarówno prawdopodobieństwo, jak i statystyki jako metody badania zarówno sytuacji opartych na danych, jak i rzeczywistych scenariuszy. Zasady prawdopodobieństwa rządzą podstawami przewidywania. Używamy prawdopodobieństwa do określenia prawdopodobieństwa wystąpienia zdarzenia. W tej części przyjrzymy się następującym tematom:

• Co to jest prawdopodobieństwo?

• Różnice między podejściem Frequentystycznym a Bayesowskim

• Jak wizualizować prawdopodobieństwo

• Jak korzystać z reguł prawdopodobieństwa

• Używanie macierzy pomyłek do przyjrzenia się podstawowym metrykom

Prawdopodobieństwo pomoże nam w modelowaniu rzeczywistych wydarzeń, które zawierają poczucie przypadkowości i przypadku. W kolejnych dwóch częściach przyjrzymy się terminologii stojącej za twierdzeniami prawdopodobieństwa i sposobom ich zastosowania w sytuacjach modelowych, które mogą pojawić się nieoczekiwanie.

Podstawowe definicje

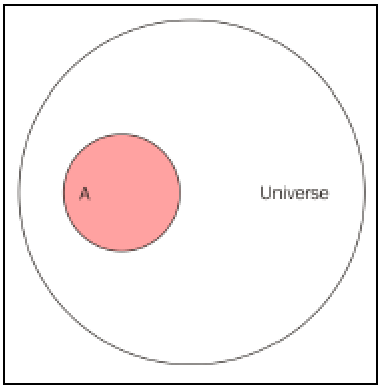
Jednym z podstawowych pojęć prawdopodobieństwa jest pojęcie procedury. Procedura to czynność, która prowadzi do rezultatu. Na przykład rzucanie kostką lub odwiedzanie strony internetowej. Zdarzenie to zbiór wyników procedury, takich jak trafienie orzełkiem w monetę lub opuszczenie strony internetowej po zaledwie 4 sekundach. Proste zdarzenie to wynik/zdarzenie procedury, którego nie można dalej rozbić. Na przykład rzut dwiema kośćmi można podzielić na dwa proste zdarzenia: rzut 1 kostką, i rzut 2. Przestrzeń próbki procedury to zbiór wszystkich możliwych prostych zdarzeń. Na przykład przeprowadzany jest eksperyment, w którym moneta jest rzucana trzy razy z rzędu. Jaka jest wielkość przestrzeni próbnej dla tego eksperymentu? Odpowiedzią jest osiem, ponieważ wynikiem może być dowolna z możliwości w następującej przestrzeni próbek - {OOO, OOR, ORR, ORO, RRR, RRO, ROO lub ROR}.

Prawdopodobieństwo

Prawdopodobieństwo zdarzenia reprezentuje częstotliwość lub szansę, że zdarzenie się wydarzy. Dla notacji, jeśli A jest zdarzeniem, P(A) jest prawdopodobieństwem wystąpienia zdarzenia. Możemy zdefiniować rzeczywiste prawdopodobieństwo zdarzenia A w następujący sposób:

P(A) = liczba razy występowania A / wielkość przestrzeni próbki

Tutaj A jest wydarzeniem, o którym mowa. Pomyśl o całym wszechświecie wydarzeń, w których wszystko jest możliwe, i przedstawmy to jako okrąg. Możemy myśleć o pojedynczym zdarzeniu A jako o mniejszym okręgu w tym większym wszechświecie, jak pokazano na poniższym diagramie:



Załóżmy teraz, że w naszym wszechświecie prowadzone są badania naukowe na ludziach, a wydarzeniem A są ludzie z tymi badaniami, którzy mają raka. Jeśli nasze badanie obejmuje 100 osób, a A ma 25 osób, prawdopodobieństwo wystąpienia A lub P(A) wynosi 25/100. Maksymalne prawdopodobieństwo wystąpienia dowolnego zdarzenia wynosi 1. Można to rozumieć, jako że czerwony okrąg powiększa się tak bardzo, że jest to rozmiar wszechświata (większy okrąg). Najbardziej podstawowe przykłady (obiecuję, że będą ciekawsze) to rzuty monetą. Powiedzmy, że mamy dwie monety i chcemy mieć prawdopodobieństwo, że wyrzucimy dwie głowy. Bardzo łatwo możemy policzyć, w jaki sposób dwie monety mogą być dwiema orłami. Jest tylko jeden! Obie monety muszą być głowami. Ale ile jest opcji? Mogą to być dwa orły, dwa reszki lub kombinacja orła/reszki. Najpierw zdefiniujmy A. Jest to zdarzenie, w którym występują dwie głowy. Liczba sposobów, w jakie A może wystąpić, wynosi 1. Przestrzeń próbki eksperymentu to {HH, HT, TH, TT}, gdzie każde dwuliterowe słowo wskazuje jednocześnie wynik pierwszej i drugiej monety. Wielkość przestrzeni próbki to cztery. Tak więc P (uzyskanie dwóch głów) = 1/4. Aby to udowodnić, odwołajmy się do szybkiej tabeli wizualnej. Poniższa tabela przedstawia opcje dla monety 1 jako kolumny i opcje dla monety 2 jako rzędy. W każdej komórce znajduje się prawda lub fałsz. Wartość True wskazuje, że spełnia warunek (oba orły), a False wskazuje inaczej.

	Moneta 1 to Orzeł	Moneta 1 to Reszka
Moneta 2 to Orzeł	True	False
Moneta 2 to Reszka	False	False

Mamy więc jeden z czterech możliwych wyników.

Bayesian kontra Frequentist

Poprzedni przykład był prawie zbyt prosty. W praktyce prawie nigdy nie jesteśmy w stanie naprawdę policzyć, na ile może się coś wydarzyć. Załóżmy na przykład, że chcemy poznać prawdopodobieństwo, że przypadkowa osoba pali papierosy przynajmniej raz dziennie. Gdybyśmy chcieli podejść do tego problemu w sposób klasyczny (poprzednia formuła), musielibyśmy dowiedzieć się, na ile różnych sposobów dana osoba jest palaczem - ktoś, kto pali przynajmniej raz dziennie - co nie jest możliwe! W obliczu takiego problemu, przy obliczaniu prawdopodobieństw w praktyce, rozważa się dwie główne szkoły myślenia: podejście Frequentystyczne i podejście bayesowskie. Tu skupimy się głównie na podejściu Frequentystycznym, podczas gdy kolejna część zagłębi się w analizę bayesowską.

Frequentystyczne podejście

W podejściu Frequentystycznym prawdopodobieństwo zdarzenia jest obliczane poprzez eksperymenty. Wykorzystuje przeszłość, aby przewidzieć przyszłą szansę na wydarzenie. Podstawowa formuła wygląda następująco:

P(A) = liczba razy występowania A / ile razy procedura została powtórzona

Zasadniczo obserwujemy kilka przypadków zdarzenia i liczymy, ile razy A było spełnione. Dzielenie tych liczb jest przybliżeniem prawdopodobieństwa. Podejście bayesowskie różni się tym, że dyktuje, że prawdopodobieństwa należy rozróżniać za pomocą środków teoretycznych. Stosując podejście Bayesa, musielibyśmy bardziej krytycznie myśleć o zdarzeniach i ich przyczynach. Żadna metodologia nie jest zawsze w 100% poprawną odpowiedzią. Zwykle sprowadza się to do problemu i trudności w stosowaniu obu podejść. Sednem podejścia Frequentist jest względna częstotliwość. Względna częstotliwość zdarzenia to częstotliwość występowania zdarzenia podzielona przez całkowitą liczbę

Przykład - statystyki marketingowe

Załóżmy, że interesuje Cię ustalenie, jak często osoba odwiedzająca Twoją witrynę prawdopodobnie wróci w późniejszym terminie. Jest to czasami nazywane wskaźnikiem powracających gości. W poprzedniej definicji nasze wydarzenie A zdefiniowalibyśmy jako odwiedzający powracający na stronę. Musielibyśmy wtedy obliczyć, na ile dana osoba może wrócić, co nie ma żadnego sensu! W tym przypadku wiele osób zwróciłoby się ku podejściu bayesowskiemu; jednak możemy obliczyć tak zwaną częstotliwość względną. Tak więc w tym przypadku możemy wziąć dzienniki odwiedzających i obliczyć względną częstotliwość zdarzenia A (powtarzający się odwiedzający). Załóżmy, że z 1458 unikalnych użytkowników w zeszłym tygodniu 452 było powracającymi użytkownikami. Możemy to obliczyć w następujący sposób:

P(A) RF(A) = 452/1458 = 0,31

Tak więc około 31% Twoich odwiedzających to powracający odwiedzający.

Prawo wielkich liczb

Powodem, dla którego nawet podejście Frequentystyczne może to zrobić, jest prawo wielkich liczb, które mówi, że jeśli będziemy powtarzać procedurę w kółko, prawdopodobieństwo względnej częstotliwości zbliży się do prawdopodobieństwa rzeczywistego. Spróbujmy to zademonstrować za pomocą Pythona. Gdybym zadał ci średnią z liczb 1 i 10, bardzo szybko odpowiedziałbyś na około 5. To pytanie jest identyczne z prośbą o wybranie średniej liczby z przedziału od 1 do 10. Zaprojektujmy eksperyment w następujący sposób: Python wybierze n losowych liczb od 1 do 10 i znajdzie ich średnią. Powtórzymy ten eksperyment kilka razy, używając za każdym razem większej liczby n, a następnie wykreślimy wynik. Kroki są następujące:

1. Wybierz losową liczbę od 1 do 10 i znajdź średnią.

2. Wybierz dwie losowe liczby od 1 do 10 i znajdź ich średnią.

3. Wybierz trzy losowe liczby od 1 do 10 i znajdź ich średnią.

4. Wybierz 10 000 losowych liczb od 1 do 10 i znajdź ich średnią.

5. Wykres wyników.

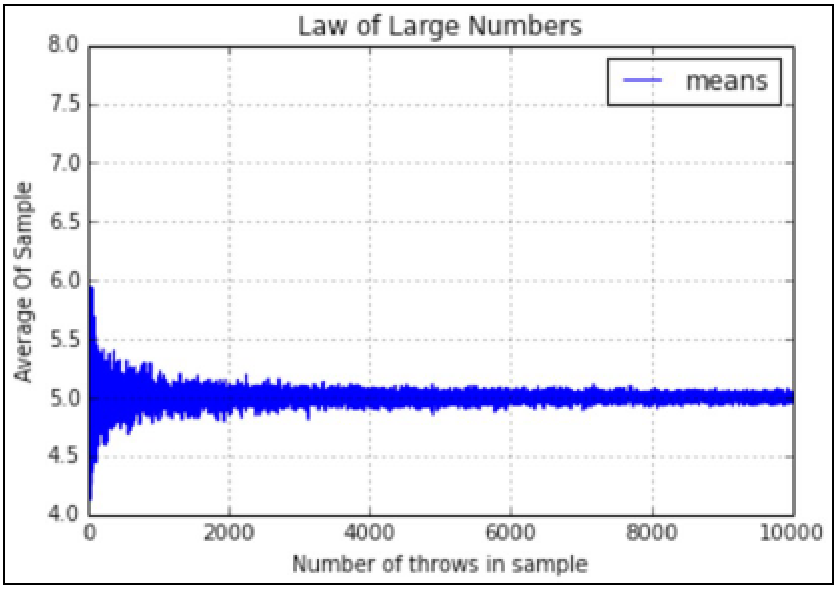
Rzućmy okiem na kod:

```
import numpy as np
```

```
import pandas as pd

from matplotlib import pyplot as plt

%matplotlib inline

results = []

for n in range(1,10000):

nums = np.random.randint(low=1,high=10, size=n) # choose n numbers

between 1 and 10

mean = nums.mean() # find the average

of these numbers

results.append(mean) # add the average

to a running list

# POP QUIZ: How large is the list results?

len(results) # 9999

# This was tricky because I took the range from 1 to 10000 and usually

we do from 0 to 10000

df = pd.DataFrame({ 'means' : results})

print df.head() # the averages in the beginning are all over the place!

# means

# 9.0

# 5.0

# 6.0

# 4.5

# 4.0

print df.tail() # as n, our size of the sample size, increases, the averages get closer to 5!

# means

# 4.998799

# 5.060924

# 4.990597

# 5.008802

# 4.979198

df.plot(title='Law of Large Numbers')
```

plt.xlabel("Number of throws in sample")

plt.ylabel("Average Of Sample")



Fajnie, prawda? Zasadniczo pokazuje nam to, że gdy zwiększamy wielkość próbki naszej względnej częstości, częstotliwość zbliża się do rzeczywistej średniej (prawdopodobieństwa) równej 5. W naszych rozdziałach dotyczących statystyk będziemy pracować nad zdefiniowaniem tego prawa znacznie bardziej rygorystycznie, ale na razie , po prostu wiedz, że jest używany do powiązania względnej częstotliwości zdarzenia z jego rzeczywistym prawdopodobieństwem.

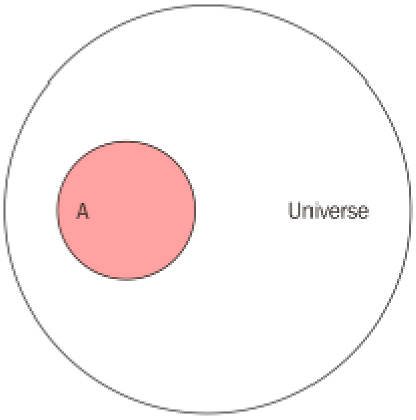
str. 93

Wydarzenia złożone

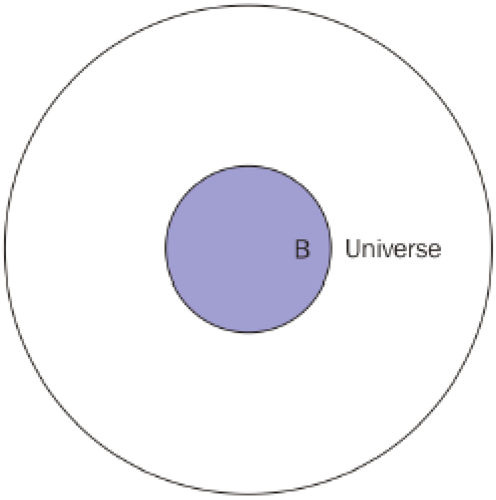
Czasami musimy mieć do czynienia z dwoma lub więcej zdarzeniami. Są to tak zwane zdarzenia złożone. Zdarzenie złożone to dowolne zdarzenie, które łączy dwa lub więcej prostych zdarzeń. Kiedy tak się dzieje, potrzebujemy specjalnej notacji. Biorąc pod uwagę zdarzenia A i B:

- Prawdopodobieństwo wystąpienia A i B wynosi P(A ∩ B) = P(A i B)
- Prawdopodobieństwo wystąpienia A lub B wynosi P(A $\&cu \cup p$; B) = P(A lub B)

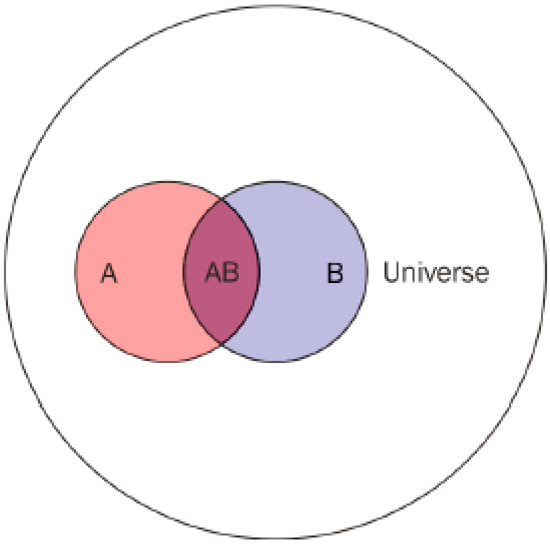
Zrozumienie, dlaczego używamy notacji zbiorowej dla tych złożonych zdarzeń, jest bardzo ważne. Pamiętasz, jak wcześniej przedstawialiśmy wydarzenia we wszechświecie za pomocą kół? Załóżmy, że nasz Wszechświat to 100 osób, które zgłosiły się na eksperyment, w którym opracowywany jest nowy test na raka:



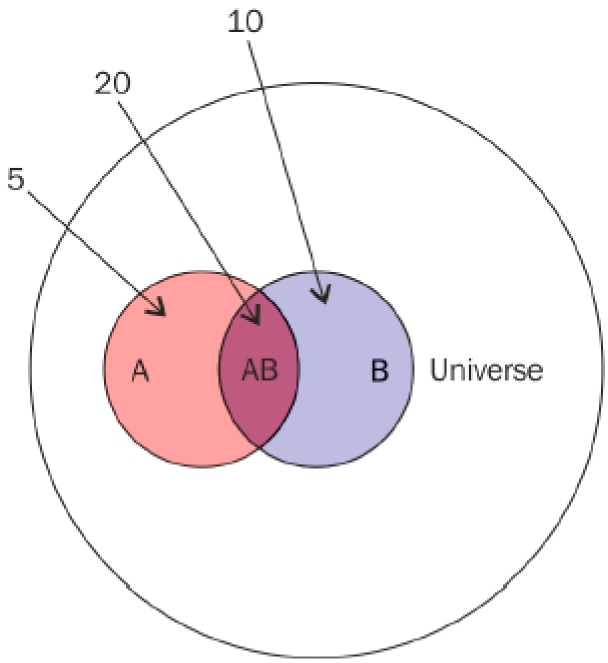
Na poprzednim schemacie czerwone kółko A przedstawia 25 osób, które rzeczywiście mają raka. Stosując podejście względnej częstotliwości, możemy powiedzieć, że P(A) = liczba osób z rakiem/liczba osób w badaniu, czyli 25/100 = ¼ = 0,25. Oznacza to, że istnieje 25% szans, że ktoś ma raka. Wprowadźmy drugie zdarzenie, nazwane B, jako show , które zawiera osoby, u których wynik testu był pozytywny (twierdził, że mają raka). Powiedzmy, że to jest dla 30 osób. Tak więc P(B) = 30/100 = 3/10 = 0,3. Oznacza to, że istnieje 30% szans na to, że test wykaże pozytywny wynik dla danej osoby:



Są to dwa oddzielne wydarzenia, ale wzajemnie na siebie oddziałują. Mianowicie mogą się przecinać lub mieć wspólnych ludzi, jak pokazano tutaj:



Każdy w przestrzeni, którą zajmują zarówno A, jak i B, inaczej znany jako A przecina B lub A ∩ B, to ludzie, dla których test stwierdził, że mają pozytywny wynik na raka (A) i faktycznie mają raka. Powiedzmy, że to 20 osób. Test wykazał pozytywny wynik dla 20 osób, to znaczy, że mają raka, jak pokazano tutaj:



Oznacza to, że P(A i B) = 20/100 = 1/5 = 0,2 = 20%. Jeśli chcemy powiedzieć, że ktoś ma raka lub wynik testu wypadł pozytywnie. Byłaby to łączna suma (lub suma) dwóch zdarzeń, a mianowicie suma 5, 20 i 10, czyli 35. Zatem 35/100 osób albo ma raka, albo miało pozytywny wynik testu. Oznacza to, że P(A lub B) = 35/100 = 0,35 = 35%. W sumie mamy ludzi w następujących czterech różnych klasach:

- Różowy: dotyczy osób z rakiem i z negatywnym wynikiem testu
- Fioletowy (A przecięcie B): Ci ludzie mają raka i mają pozytywny wynik testu
- Niebieski: dotyczy osób bez raka i z pozytywnym wynikiem testu
- Biały: dotyczy osób bez raka i z negatywnym wynikiem testu

Tak więc, w praktyce, test był dokładny tylko wtedy, gdy był w obszarach białych i fioletowych. W obszarach niebieskim i różowym test był nieprawidłowy.

Warunkowe prawdopodobieństwo

Wybierzmy dowolną osobę z badania 100 osób. Załóżmy również, że powiedziano ci, że ich wynik testu był pozytywny. Jakie jest prawdopodobieństwo, że rzeczywiście zachorują na raka? Tak więc powiedziano nam, że wydarzenie B już się odbyło i że ich test wypadł pozytywnie. Teraz pytanie brzmi: jakie jest prawdopodobieństwo, że mają raka, czyli P(A)? Nazywa się to prawdopodobieństwem warunkowym danego B lub P(A|B). W rzeczywistości jest to prośba o obliczenie prawdopodobieństwa zdarzenia, biorąc pod uwagę, że inne zdarzenie już się wydarzyło. Możesz myśleć o prawdopodobieństwie warunkowym jako o zmianie odpowiedniego wszechświata. P(A|B) (nazywane prawdopodobieństwem A przy danym B) to sposób na powiedzenie, biorąc pod uwagę, że cały mój wszechświat jest teraz B, jakie jest prawdopodobieństwo A? Jest to również znane jako przekształcanie przestrzeni próbki. Wzór może być podany w następujący sposób:

P(A|B) = P(A i B) / P(B) = (20/100) / (30/100) = 20/30 = 0,66 = 66%

Istnieje 66% szans, że jeśli wynik testu będzie pozytywny, ta osoba ma raka. W rzeczywistości jest to główne prawdopodobieństwo, którego chcą eksperymentatorzy. Chcą wiedzieć, jak dobry jest test w przewidywaniu raka.

Zasady prawdopodobieństwa

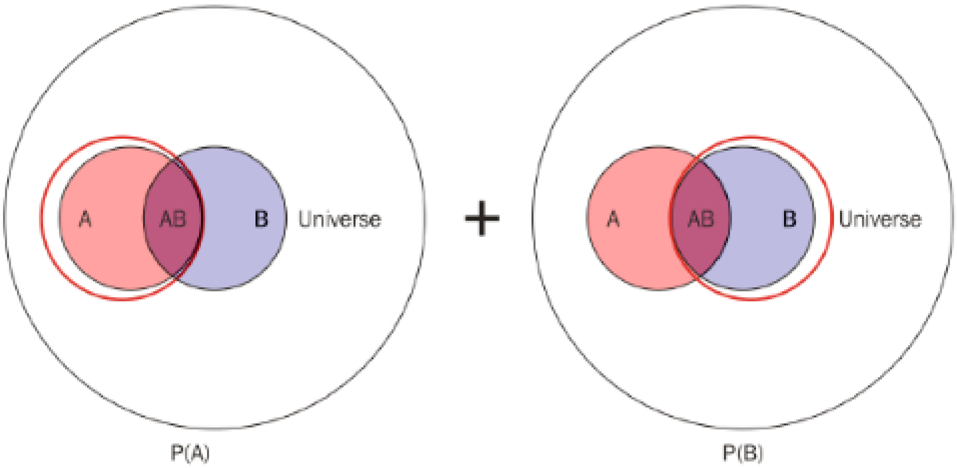
W prawdopodobieństwie mamy pewne zasady, które stają się bardzo przydatne, gdy wizualizacja staje się zbyt uciążliwa. Te reguły pomagają nam z łatwością obliczać prawdopodobieństwa złożone.

Zasada dodawania

Reguła dodawania służy do obliczania prawdopodobieństwa zdarzenia lub zdarzeń. Licząc $P(A \cup B)$ = P(A lub B), posługujemy się następującym wzorem:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Pierwsza część wzoru (P(A) + P(B)) ma sens. Aby uzyskać połączenie tych dwóch wydarzeń, musimy zsumować obszar okręgów we wszechświecie. Ale dlaczego odejmowanie P(A i B)? Dzieje się tak, ponieważ dodając dwa okręgi, dwukrotnie dodajemy obszar przecięcia, jak pokazano na poniższym diagramie:



Widzisz, jak oba czerwone kółka obejmują przecięcie A i B? Tak więc, kiedy je dodajemy, musimy odjąć tylko jedną z nich, aby to uwzględnić, pozostawiając nam naszą formułę. Przypomnij sobie, że chcieliśmy uzyskać liczbę osób, które albo miały raka, albo miały pozytywny wynik testu? Jeśli A to zdarzenie, że ktoś ma raka, a B to, że wynik testu był pozytywny, mamy:

$P(A \text{ lub } B) = P(A) + P(B) – P(A \text{ i } B) = 0,25 + 0,30 - 0,2 = 0,35$

Zostało to obliczone wcześniej wizualnie na schemacie.

Wzajemna wyłączność

Mówimy, że dwa zdarzenia wykluczają się wzajemnie, jeśli nie mogą wystąpić w tym samym czasie. Oznacza to, że $A \cap B = \emptyset$ lub po prostu przecięcie zdarzeń jest zbiorem pustym. Kiedy tak się dzieje, $P(A \cap B) = P(A \text{ i } B) = 0$. Jeżeli dwa zdarzenia wzajemnie się wykluczają, to:

$P(A \cup B) = P(A \text{ lub } B) = P(A) + P(B) − P(A \cap B) = P(A) + P(B)$

To znacznie ułatwia regułę dodawania. Oto kilka przykładów wzajemnie wykluczających się wydarzeń:

• Klient widzi Twoją witrynę po raz pierwszy zarówno na Twitterze, jak i na Facebooku

• Dzisiaj jest sobota, a dziś jest środa

• Oblałem Econ 101 i zdałem Econ 101

Żadne z tych zdarzeń nie może wystąpić jednocześnie.

Zasada mnożenia

Reguła mnożenia służy do obliczania prawdopodobieństwa zdarzeń i zdarzeń. Aby obliczyć P(A ∩ B) = P(A i B), używamy następującego wzoru:

$P(A \cap B) = P(A \text{ i } B) = P(A) \cdot P(B|A)$

Dlaczego używamy B|A zamiast B? Dzieje się tak, ponieważ możliwe jest, że B zależy od A. Jeśli tak jest, to samo pomnożenie P(A) i P(B) nie daje pełnego obrazu. W naszym przykładzie z rakiem znajdźmy P(A i B). Aby to zrobić, przedefiniujmy A jako zdarzenie, w którym badanie jest pozytywne, a B jako osobę z rakiem (ponieważ nie ma znaczenia, jak nazywamy zdarzenia). Równanie będzie wyglądało następująco:

$P(A \cap; B) = P(A \text{ i } B) = P(A) \cdot P(B|A) = 0,3 * 0,6666 = 0,2 = 20\%$

Zostało to obliczone wcześniej wizualnie. Trudno dostrzec prawdziwą konieczność korzystania z prawdopodobieństwa warunkowego, spróbujmy więc inny, trudniejszy problem. Na przykład z losowo wybranego zestawu 10 osób 6 ma iPhone'y, a 4 Androidy. Jakie jest prawdopodobieństwo, że jeśli losowo wybiorę dwie osoby, obie będą miały iPhone'y? Ten przykład można powtórzyć za pomocą przestrzeni zdarzeń w następujący sposób:

Mam następujące dwa wydarzenia:

• O: To zdarzenie pokazuje prawdopodobieństwo, że jako pierwszy wybiorę osobę z iPhonem

• B: To zdarzenie pokazuje prawdopodobieństwo, że wybiorę drugą osobę z iPhonem

Więc zasadniczo chcę, co następuje:

• P(A i B): P(wybieram osobę z iPhone'em i osobę z iPhone'em)

Więc mogę użyć mojego wzoru P(A i B) = P(A) · P(B|A). P(A) jest proste, prawda? Osób z iPhone'ami jest 6 na 10, więc mam 6/10 = 3/5 = 0,6 szansy na A. To oznacza, że P(A) = 0,6. Tak więc, jeśli mam 0,6 szansy na wybranie kogoś z iPhonem, prawdopodobieństwo wyboru dwóch powinno wynosić 0,6 * 0,6, prawda?

Ale poczekaj! Zostało nam tylko 9 osób do wyboru naszej drugiej osoby, bo jedna została zabrana. Tak więc w naszej nowej przekształconej przestrzeni próbnej mamy w sumie 9 osób, 5 z iPhone'ami i 4 z Androidami, co daje P(B) = 5/9 = 0,555. Tak więc prawdopodobieństwo wyboru dwóch osób z iPhone'ami wynosi 0,6 * 0,555 = 0,333 = 33%. Mam 1/3 szansy na wybranie dwóch osób z iPhone'ami spośród 10. Prawdopodobieństwo warunkowe jest bardzo ważne w regule mnożenia, ponieważ może drastycznie zmienić twoją odpowiedź.

Niezależność

Dwa zdarzenia są niezależne, jeśli jedno zdarzenie nie wpływa na wynik drugiego, czyli P(B|A) = P(B) i P(A|B) = P(A).

Jeżeli dwa zdarzenia są niezależne, to:

P(A ∩ B) = P(A) ·; P(B|A) = P(A) · P(B)

Oto kilka przykładów niezależnych wydarzeń:

• W San Francisco padało, a w Indiach urodził się szczeniak

• Rzuć monetą i zdobądź orły i rzucaj kolejną monetą i zdobądź reszki

Żadna z tych par wydarzeń nie wpływa na siebie nawzajem.

Wydarzenia uzupełniające

Uzupełnienie A jest przeciwieństwem lub negacją A. Jeśli A jest wydarzeniem, $\overline{A}$ reprezentuje uzupełnienie A. Na przykład, jeśli A jest wydarzeniem, w którym ktoś ma raka, $\overline{A}$ jest wydarzeniem, w którym ktoś jest wolny od raka. Aby obliczyć prawdopodobieństwo $\overline{A}$, użyj następującego wzoru:

$$P(\overline{A}) = 1 - P(A)$$

Na przykład, kiedy rzucasz dwiema kostkami, jakie jest prawdopodobieństwo, że wyrzuciłeś więcej niż 3? Niech A reprezentuje toczenie wyższe niż 3.

$\overline{A}$ oznacza wyrzucenie 3 lub mniej.

$P(A) = 1 - P(\overline{A})$

$P(A) = 1 - (P(2)+P(3))$

$= 1 - (2/36 + 2/36)$

$= 1 - (4/36)$

$= 32/36 = 8 / 9$

$= .89$

Na przykład zespół start-upów ma przed sobą trzy spotkania z inwestorami. Będziemy mieli następujące prawdopodobieństwa:

• 60% szans na zdobycie pieniędzy od pierwszego spotkania

• 15% szansy na otrzymanie pieniędzy od drugiego

• 45% szansy na zdobycie pieniędzy od trzeciego

Jakie jest prawdopodobieństwo, że zdobędą pieniądze z przynajmniej jednego spotkania? Niech A będzie zespołem otrzymującym pieniądze od co najmniej jednego inwestora i niech będzie zespołem, który nie otrzyma żadnych pieniędzy. P(A) można obliczyć w następujący sposób:

$P(A) = 1 - P(\overline{A})$

Aby obliczyć $P(\overline{A})$, musimy obliczyć:

$P(\overline{A}) = P$ (brak pieniędzy od inwestora 1 AND brak pieniędzy od inwestora 2 ORAZ brak pieniędzy od inwestora 3) Jeśli założymy, że te zdarzenia są niezależne (nie rozmawiają ze sobą), to:

$P(\overline{A}) = P$ (brak pieniędzy od inwestora 1) * P(brak pieniędzy od inwestora 2) * P(brak pieniędzy od inwestora 3) =

0,4 * 0,85 * 0,55 = 0,187

P(A) = 1 - 0,187 = 0,813 = 81 %

Tak więc startup ma 81% szans na zdobycie pieniędzy z przynajmniej jednego spotkania!

Trochę głębiej

Bez zbytniego zagłębiania się w terminologię uczenia maszynowego, ten test jest tak zwany klasyfikatorem binarnym, co oznacza, że próbuje przewidzieć tylko dwie opcje: mieć raka lub nie mieć raka. Kiedy mamy do czynienia z klasyfikatorami binarnymi, możemy narysować tak zwane macierze pomyłek, które są macierzami 2 x 2, które zawierają wszystkie cztery możliwe wyniki naszego eksperymentu. Spróbujmy różnych liczb. Powiedzmy, że na badanie przyszło 165 osób. Tak więc nasze n (wielkość próbki) to 165 osób. Wszystkim 165 osobom poddaje się test i pyta, czy ma raka (przeprowadzanego różnymi innymi środkami). Poniższa macierz pomyłek pokazuje nam wyniki tego eksperymentu:

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

Macierz pokazuje, że przewidziano, że 50 osób nie ma i nie ma raka, 100 osób miało raka i rzeczywiście go miało, i tak dalej. Ponownie mamy następujące cztery klasy, wszystkie o różnych nazwach:

• Prawdziwie pozytywne są testy prawidłowo przewidujące wynik pozytywny (rak) == 100

• Prawdziwie negatywy to testy prawidłowo przewidujące wynik negatywny (brak raka) == 50

• Fałszywie pozytywne to testy, które błędnie przewidują wynik pozytywny (rak) == 10

• Wyniki fałszywie ujemne to testy błędnie przewidujące wynik ujemny (brak raka) == 5

Pierwsze dwie klasy wskazują, gdzie test był poprawny lub prawdziwy. Dwie ostatnie klasy wskazują, gdzie test był błędny lub fałszywy. Fałszywe trafienia są czasami nazywane błędami typu I, podczas gdy fałszywie ujemne są nazywane błędami typu II.

Zajmiemy się tym w dalszych częściach. Na razie musimy tylko zrozumieć, dlaczego używamy notacji zbioru do oznaczania prawdopodobieństw zdarzeń złożonych. Dzieje się tak, ponieważ tym właśnie są. Kiedy zdarzenia A i B istnieją w tym samym wszechświecie, możemy użyć skrzyżowań i związków, aby przedstawić je w tym samym czasie lub przedstawić jedno i drugie. Zajmiemy się tym znacznie więcej w dalszych rozdziałach, ale dobrze jest przedstawić to teraz.

Podsumowanie

W tej części przyjrzeliśmy się podstawom prawdopodobieństwa i będziemy dalej zagłębiać się w tę dziedzinę w następnej części. Do większości naszego myślenia podeszliśmy jako częsty i przedstawiliśmy podstawy eksperymentowania i wykorzystywania prawdopodobieństwa do przewidywania wyników. W następnej części przyjrzymy się bayesowskiemu podejściu do prawdopodobieństwa, a także zbadamy wykorzystanie prawdopodobieństwa do rozwiązywania znacznie bardziej złożonych problemów. Te podstawowe zasady prawdopodobieństwa uwzględnimy w znacznie trudniejszych scenariuszach.