MINGGU KE 10

Pengklasifikasian Naive Bayes

Dalam machine learning, suatu pengklasifikasi Bayes adalah pengklasifikasi probabilistik sederhana, yang berdasarkan pada pengaplikasian teorema Bayes. Model fitur digunakan oleh suatu pengklasifikasi naive Bayes membuat asumsi kebebasan kuat. Ini berarti bahwa eksistensi dari suatu bagian fitur dari suatu class adalah bebas atau tidak terkait pada eksistensi dari setiap fitur lain.

Definisi dari kejadian bebas:

Dua kejadian E dan F adalah bebas, jika keduanya E dan F memiliki kemungkinan positif dan jika P(E|F)=P(E) dan P(F|E)=P(F). Teorema Bayes berdasar atas probabilitas kondisional, yang akan didefinisikan berikut ini.

Probabilitas Kondisional

P(A|B) adalah probabilitas kondisional A jika sebelumnya diberikan B, atau kemungkinan dari A jika sebelumnya B. Ketika dalam suatu percobaan acak kejadian B diketahui telah terjadi, kemungkinan hasil percobaan dikurangi menjadi B, dan karenanya probabilitas terjadinya A diubah dari probabilitas tanpa syarat ke dalam probabilitas bersyarat yang diberikan B. Probabilitas Bersama adalah probabilitas dua kejadian yang berhubungan. Artinya, ini adalah probabilitas dari kejadian bersama. Ada tiga notasi untuk probabilitas gabungan dari A dan B. Ini dapat ditulis sebagai

- P(A∩B)
- P(AB) atau
- P(A,B)

Probabilitas kondisional didefinisikan oleh:
$$P(A|B)=P(A\cap B)/P(B)$$

Contoh dari probabilitas kondisional

German Swiss Speaker

Ada sekitar 8.4 juta orang tinggal di Switzerland. Sekitar 64% dari penduduk itu berbicara German. Ada sekitar 7500juta orang di bumi. Jika beberapa alien secara acak menembaki bumi, berapa kemungkinan orang tertembak adalah orang Swiss berbahasa Jerman?

Kita memiliki kejadian
S:orang Swiss
GS:berbicara Jerman

Peluang untuk suatu pemilihan orang acak adalah orang Swiss:

$P(S) = 8.4/7500 = 0.00112$

Jika kita tahu bahwa seseorang adalah Swiss, peluangnya berbicara Jerman adalah 0.64. Ini terkait dengan peluang bersyarat

$P(GS|S) = 0.64$

Maka peluang dari penduduk bumi adalah Swiss dan berbicara Jerman dapat dihitung dengan rumus:

$P(GS|S) = P(GS \cap S)/P(S)$

Nilai yang diperoleh menjadi:

$0.64 = P(GS \cap S)/0.00112$

$P(GS \cap S) = 0.0007168$

Jadi alien kita berakhir dengan kesempatan 0,07168% mendapatkan orang Swiss berbahasa Jerman.

False Positif dan False Negatif

Suatu lab riset medis mengajukan suatu screening untuk menguji suatu grup luas dari orang untuk suatu penyakit X. Suatu argumen pada beberapa screening adalah masalah dari hasil screening false positif.

Misalnya 0.1% dari grup terkena penyakit ini, sisanya adalah sehat:

$P(\text{"}sick\text{"}) = 0{,}1$
dan
$P(\text{"}well\text{"}) = 99{,}9$

Berikut ini adalah benar untuk suatu tes screening, Jika Anda memiliki penyakit ini, tes akan positif 99% dari waktu, dan jika Anda tidak memilikinya, tesnya akan menjadi negatif 99% dari waktu:

P("test positive" | "well") = 1 %

and

P("test negative" | "well") = 99 %.

Akhirnya, anggaplah bahwa saat tes ini diterapkan pada seseorang yang menderita penyakit ini, ada kemungkinan 1% hasil negatif palsu (dan 99% kemungkinan mendapatkan hasil positif yang sebenarnya), yaitu.

P("test negative" | "sick") = 1 %

and

P("test positive" | "sick") = 99 %

	Sick	Healthy	Totals
Test result positive	99	999	1098
Test result negative	1	98901	98902
Totals	100	99900	100000

Ada 999 Positif palse (False Positive) dan 1 Negatif palsu

Masalah:
Dalam banyak kasus, profesional medis sekalipun berasumsi bahwa "jika Anda menderita penyakit ini, tesnya akan positif dalam 99% dari waktu dan jika Anda tidak memilikinya, tesnya akan menjadi negatif 99% dari waktu. Kasus yang melaporkan hasil positif hanya 99 (9%) kasus yang benar dan 999 kasus positif palsu (91%), yaitu jika seseorang mendapat hasil tes positif, kemungkinan dia benar-benar memiliki penyakit ini hanya sekitar 9% . P("sick" | "test positive") = 99 / 1098 = 9.02 %

Teorema Bayes

Kita menghitung peluang bersyarat P(GS|S), yang mana peluang bahwa seseorang bicara Jerman, jika dia diketahui Swiss. Untuk menghitungnya kita menggunakan persamaa berikut:

$$P(GS|S) = P(GS,S)/P(S)$$

Bagaimana jika kita menghitung peluang P(S|GS), yaitu peluang bahwa seseorang itu adalah Swiss dengan asumsi bahwa orang tersebut berbicara Jerman?

Persamaannya akan menjadi:

$$P(S|GS)=P(GS,S)/P(GS)$$

Mari kita isolasi kedua persamaan tersebut P(GS,S):

$$P(GS,S)=P(GS|S)P(S)$$
$$P(GS,S)=P(S|GS)P(GS)$$

Dengan mengubah sisi kiri maka diperoleh persamaan:

$$P(GS|S)*P(S)=P(S|GS)P(GS)$$

Persamaan di atas akan diubah ke dalam:

$$P(S|GS)=P(GS|S)P(S)/P(GS)$$

Hasil ini merujuk pada Teorema Bayes

Untuk menyelesaikan masalah kita sebagai contoh peluang bahwa seseorang adalah Swiss, jika kita tahu bahwa dia berbicara Jerman. Yang kita harus lakukan adalah menghitung sisi kanan. Kita telah mengetahui dari latihan sebelumnya bahwa

$$P(GS|S)=0.64$$

Dan

Jumlah orang Jerman native speak di dunia mencapai 101 juta, maka kita tahu bahwa

$$P(GS)=101/7500=0.0134667$$

Akhirnya kita dapat menghitung P(S|GS) dengan mensubstitusi nilai dalam persamaan kita:

$$P(S|GS)=P(GS|S)P(S)/P(GS)=0.64*0.00112/0.0134667=0.0532276$$

Ada sekitar 8.4 juta orang tinggal di Swiss. Sekitar 64% berbicara Jerman. Ada sekitar 7500 juta orang di bumi. Jika beberapa alien secara acak menembaki bumi, berapa kemungkinan orang tertembak adalah orang Swiss berbahasa Jerman?

Kita memiliki kejadian

S:orang Swiss
GS:berbicara Jerman

$$P(S)=8.4/7500=0.00112$$
$$P(A|B)=P(B|A)P(A)/P(B)$$

P(A|B) adalah probabilitas bersyarat dari A, jika B (probabilitas posterior), P (B) adalah probabilitas sebelumnya dari B dan P (A) probabilitas sebelumnya dari A. P(B|A) adalah kondisional probabilitas B diberikan A, dipanggil kemungkinan besar.

Keuntungan dari pengklasifikasi Naive Bayes adalah bahwa ia hanya memerlukan sejumlah kecil data pelatihan untuk memperkirakan parameter yang diperlukan untuk klasifikasi. Karena variabel independen diasumsikan, hanya varians variabel untuk masing-masing kelas yang harus ditentukan dan bukan keseluruhan matriks kovarians.