Dasar-Dasar Statistika Ilmu Komputer : Teori Bayesan

Dasar teori

Mengenal Bayesian
- Nama teorema Bayes diambil dari nama penemu teorema tersebut, yaitu Reverend Thomas Bayes (1702 – 1761)
- Teorema Bayes digunakan untuk menghitung probabilitas terjadinya suatu peristiwa, berdasarkan pengaruh yang didapat dari hasil observasi peristiwa sebelumya
- Teorema Bayes menyempurnakan teorema probabilitas bersyarat yang hanya dibatasi oleh 2 buah kejadian sehingga dapat diperluas untuk n buah kejadian
- Dikembangkan secara luas dalam statistika inferensia / induktif
- Aplikasi teorema Bayes banyak ditemukan pada bidang komputer cerdas sebagai salah satu dasar dari metode machine learning dan data mining

Sebelum melanjutkan penjelasan dari teorema bayes, mari mulai dari contoh kasus lempar koin. Kita asumsikan dari 5 lemparan, kita peroleh data berikut : H, T, T, T, T. (*H untuk head dan T untuk tail*). Pertanyaannya, seberapa yakin kita bahwa koin tersebut seimbang?
Dalam bayesian, kita perlu memahami bagaimana suatu data diperoleh dan proses apa yang mendasarinya. Dalam hal ini serangkaian data keluaran pelemparan koin dapat kita modelkan dengan proses binomial. Kenapa? Dapat kita baca dari penjelasan wikipedia :
Dalam teori probabilitas dan statistika, distribusi binomial adalah distribusi probabilitas diskret jumlah keberhasilan dalam n percobaan ya/tidak (berhasil/gagal) yang saling bebas, dimana setiap hasil percobaan memiliki probabilitas p.

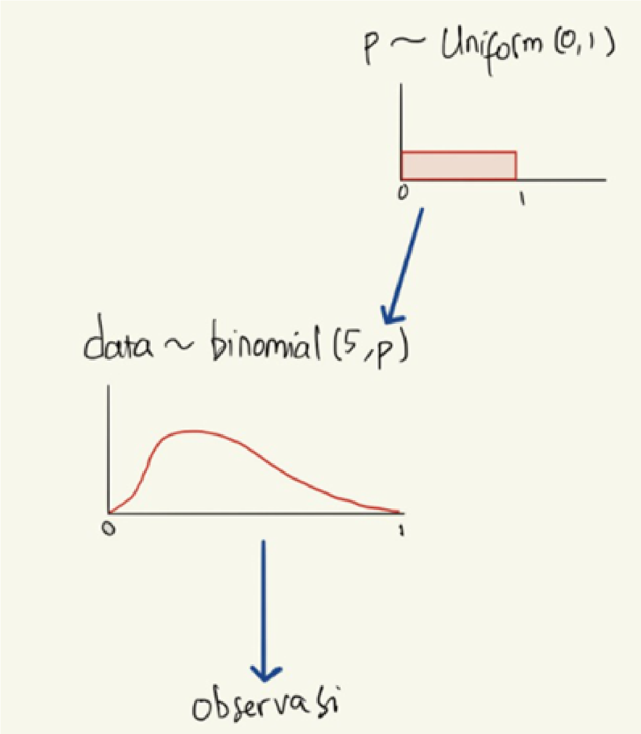
Jika kita analogikan bahwa hasil H (head) merupakan percobaan sukses, dan T (tail) merupakan percobaan gagal, tentu penjelasan diatas sangat cocok dengan kasus lemparan koin. Dalam kasus ini, kita memiliki 5 kali lemparan koin (n=5) dengan probabilitas sukses (p) yang tidak kita ketahui dan hasil lemparan adalah HTTTT (satu kali sukses). Jika dilihat dari kerangka proses standar (input, process, dan output), kita dapat menjelaskan bagaimana data dihasilkan:
- Input : n & p
- Proses : distribusi binomial(n, p)
- Output : HTTTT

Memahami bagaimana proses data dihasilkan merupakan hal yang esensial dalam melakukan inferensi bayesian. Dari proses diatas, parameter p tidak kita ketahui (yang tentu saja sejak awal probabilitas sukses ini yang ingin kita cari). Dan jika kita lihat lagi, hal ini bisa di tuliskan sebagai P(p|x) atau berapa probabilitas p jika kita ketahui data x adalah HTTTT. Bagaimana kita memperoleh ini? Tentu mari kita simulasikan.

Asumsikan kita sama sekali tidak yakin berapa nilai p sebenarnya, maka kita modelkan asumsi awal kita nilai p terdistribusi *uniform* dari angka 0 sampai 1, atau dapat dituliskan P(p) ~ Uniform(0,1), dan kita masukkan pada mesin simulasi kita dengan cara:
1. Ambil nilai random dari distribusi p, misal dari sini kita peroleh p=Uniform(0,1)=0.3
2. Masukkan ini pada proses kita : binomial(n, p) = binomial(5, 0.3), misal dari sini kita peroleh HHTTT (2 sukses dan 3 gagal).
3. Jika output sesuai denngan data (HTTTT) kita terima dan ambil nilai p, jika tidak maka kita ulang percobaan sebanyak mungkin.

Berikut simulasi agar lebih mudah di cerna:
- Percobaan 1 : p = Uniform(0,1) = 0.32 →binomial(5, 0.32) = 2 (2 kali sukses, atau HHTTT). Tolak
- Percobaan 2 : p = Uniform(0,1) = 0.11 →binomial(5, 0.11) = 1 (1 kali sukses, atau HTTTT). Ambil nilai p = 0.1
- Percobaan 3 : p = Uniform(0,1) =0.69 →binomial(5, 0.69) = 3. Tolak
- dst…

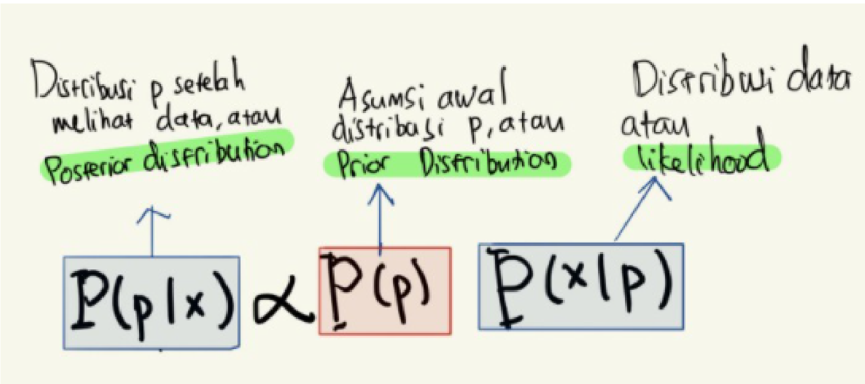
Jika kita sederhanakan, berikut kerangka berfikir proses yang sudah kita lakukan



Revisiting the Concept

Mari kita lihat kembali kasus di atas, pada dasarnya kita ingin mencari distribusi parameter p, dimana kita tahu data yang di peroleh ada 1 sukses dari 5 percobaan (HTTTT). Hal ini dapat di tuliskan dengan P(p|x) atau distribusi parameter p given observasi x. Kemudian kita tahu bahwa data diperoleh dari distribusi Binomial dengan parameter p yang tidak diketahui, atau biasa disebut likelihood function P(x|p). Dan kita memiliki asumsi awal tanpa melihat data parameter p dari distribusi Uniform, dituliskan P(p). J

Jika kita lihat P(p|x) tidak lain merupakan perkalian probabilitas dari P(p) terhadap likelihood P(x|p). Ingat proses simulasi diatas? pada teori probabilitas, perkalian antar dua probabiliatas berarti mencari nilai probabilitas dimana keduanya bernilai benar, misal probabilitas 2 dadu bernilai 1 adalah 1/6 * 1/6. Dan voila, berikut hasil rumusan kita yang tidak lain adalah teorema bayes.



Pada *bayesian statistics*, P(p) biasa di sebut dengan *prior distribution* atau *prior believe*, bagaimana kepercayaan awal kita terhadap parameter yang tidak kita ketahui. Pada kasus ini, karena kita asumsikan tidak mengetahui sama sekali kemungkinan nilai p, maka kita berikan p pada distribusi *Uniform(0,1)* yang artinya nilai p bisa berapa saja di antara 0 sampai 1.

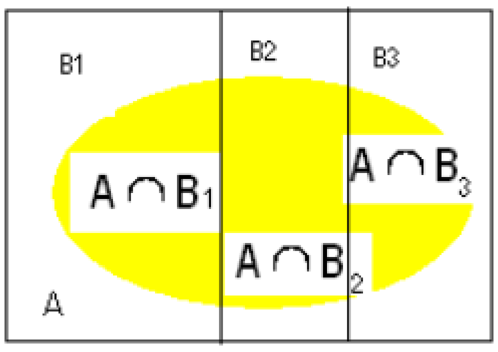
Kemudian, P(x|p) biasa disebut sebagai *likelihood function* yang merupakan probabilitas data dengan parameter yang tidak diketahui, pada hal ini dimodelkan dengan distribusi *Binomial(n, p).* Terakhir, kita seakan memperbaharui kepercayaan kita melalui proses diatas, bagaimanakah kemungkinan distribusi p sebenarnya setelah melihat data, atau P(p|x) yang biasa disebut sebagai *posterior distribution.* Dan inilah inti dari *bayesian statistics.*

Jika difikirkan ulang, konsep ini terasa sangat natural dengan bagaimana proses berfikir manusia. Misal, kenapa kita meyakini bahwa benda yang dilempar akan jatuh ke bawah? Karena data yang kita lihat melalui mata setiap hari mengatakan semua yang kita lempar akan jatuh ke bawah. Atau jika kita meminjamkan mobil pada teman yang sudah sering dipinjamkan mobil tentu kita cukup yakin bahwa mobil kita tidak akan kenapa-kenapa.

Sementara jika teman kantor mencoba meminjam dimana kita tidak pernah melihat dia membawa mobil sendiri , tentu kita akan skeptis dikarenakan ketidakpastian mobil akan selamat atau tidak cukup tinggi. Tapi jika ada seorang yang dipercaya mengatakan pada kita bahwa dia sangat mahir membawa mobil, maka kita punya *prior believe* bahwa mobil tidak akan kenapa-kenapa, meski data yang kita miliki minim. Secara tidak langsung inilah kerangka berfikir bayesian, dimana kepercayaan akhir kita merupakan perkalian antara *prior believe* (yang bisa merupakan opini dari ahli) dengan observasi (data). Dapat kita simpulkan, pada *bayesian statistics* data itu mutlak, karena itu merupakan hasil observasi, sementara parameter lah yang memiliki ketidakpastian.

Konsep Formula Teorema Bayes

Misalkan peristiwa {B1,B2,....,BN} merupakan suatu sekatan (partisi) dari ruang sampel S dengan P(Bn) ≠ 0 untuk n = 1,2,...,N. Dan misalkan A suatu kejadian sembarang dalam S dengan P(A) ≠ 0 untuk **N = 3**



$$P(A)=\sum_{n=1}^{N}P(B_n\cap A)=\sum_{n=1}^{N}P(B_n)P(A\,|\,B_n) \quad (1)$$

Berdasar teorema Probabilitas Bersyarat :
Probabilitas bersyarat suatu peristiwa A, dengan syarat peristiwa B didefinisikan sebagai:

P(A|B) = P(A ∩ B) / P(B) ; P(B) > 0

Atau

P(B|A) = P(B ∩ A) / P(A) ; P(A) > 0

Dimana berdasar teori himpunan kita ketahui :

P(A ∩ B) = P(B ∩ A)

Sehingga dari dua persamaan diatas didapatkan :
P(A∩B) = P(A|B) P(B) = P(B|A) P(A)

Maka :
P(A|B) P(B) = P(B|A) P(A)
P(B|A) = P(A|B) P(B) / P(A)

Berdasarkan hubungan probabilitas A dengan probabilitas kejadian bersyarat sebagaimana ditunjukkan persamaan awal, yaitu

$$P(A)=\sum_{n}^{N}P(A\,|\,B_n)P(B_n)$$

Sehingga Probabilitas suatu kejadian yang dibatasi oleh n buah kejadian sebagai syaratnya akan kita peroleh dari penurunan rumus sebagai berikut :

$$P(B_n\,|\,A)=\frac{P(A\cap B_n)}{P(A)}=\frac{P(B_n)P(A\,|\,B_n)}{\sum_{n=1}^{N}P(B_n)P(A\,|\,B_n)};n=1,2,..N$$

$$P(B_n\,|\,A)=\frac{P(A\,|\,B_n)P(B_n)}{P(A\,|\,B_1)P(B_1)+P(A\,|\,B_2)P(B_2)+\ldots+P(A\,|\,B_n)P(B_n)}$$

Secara umum formula teorema Bayes adalah sebagai berikut:

Peluang Likelihood

Peluang Prior

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

Peluang Posterior

Peluang Evidence

- Peluang Posterior adalah prediksi peluang munculnya satu kejadian berdasarkan informasi dari kejadian yang lain
- Peluang Prior adalah peluang munculnya suatu kejadian yang sudah kita yakini sebelumnya dan bisa jadi kejadian ini dipengaruhi kejadian yang lain
- Peluang Likelihood adalah peluang yang menyatakan derajat kemungkinan pengaruh suatu informasi kejadian terhadap kejadian yang lain

yang lain
- Peluang Evidence adalah sebuah ukuran pembanding konstan berdasarkan peluang suatu informasi kejadian

Informative Prior
Salah satu hal menarik dari kerangka berfikir bayesian, kita bisa masukkan kepercayaan awal kita terhadap model. Dalam kasus lempar koin diatas, tentu kita tahu sebelumnya bahwa koin yang dilempar memiliki biasanya probabilitas 50:50. Artinya, kita ingin coba memasukkan informasi ini kedalam model dengan menggunakan *prior believe.* Caranya? kita ubah prior distribution yang digunakan, dari yang sebelumnya P~Uniform(0,1). Pertanyaannya, distribusi apa yang cocok untuk ini?

Kita tahu bahwa kemungkinan nilai p hanyalah dari 0 sampai 1. Tentu kita tidak bisa menggunakan distribusi normal karena rentang nilainya tidak cocok dengan model kita. Pada kasus ini, dapat kita coba gunakan distribusi Beta, dikarenakan rentang kemungkinan nilai nya diantara 0 dan 1 Misal, gunakan P(p) ~ Beta(5,5) dapat dibaca lebih lanjut mengenai distribusi ini pada link wikipedia.

Mengapa Beta (5,5)? Kita ingin probabilitas tertinggi pada nilai 0.5 dan kita cukup yakin soal itu, tapi kita ingin memberi ruang juga bahwa mungkin saja nilai p tidak pada 0.5. Yang paling penting dari distribusi prior adalah bentuknya, tentu kita bisa pakai Beta (100,100) dimana bentuk lonceng akan lebih kurus yang artinya kita lebih yakin bahwa koin ini cukup adil, atau Beta(2,2) dimana bentuk lonceng akan lebih gendut yang artinya memberi rentang besar pada kemungkinan nilai p. Mari kita bandingkan model kita sekarang dengan sebelumnya.

Sebelumnya
p~Uniform(0,1)
data~Binomial(5,p)

Sekarang
p~Beta(5,5)
data~Binomial(5,p)

Penggunaan Beta (5,5) sebagai *prior believe* dalam istilah *bayesian statistics* dikenal sebagai *informative prior***.** Artinya prior yang digunakan memberikan dampak pada *posterior distribution.* Artinya kita memberikan opini awal pada proses inferensi. Sebaliknya, pada penggunaan distribusi Uniform di awal, biasa disebut sebagai *non-informative prior* sehingga inferensi yang dila

Quiz