

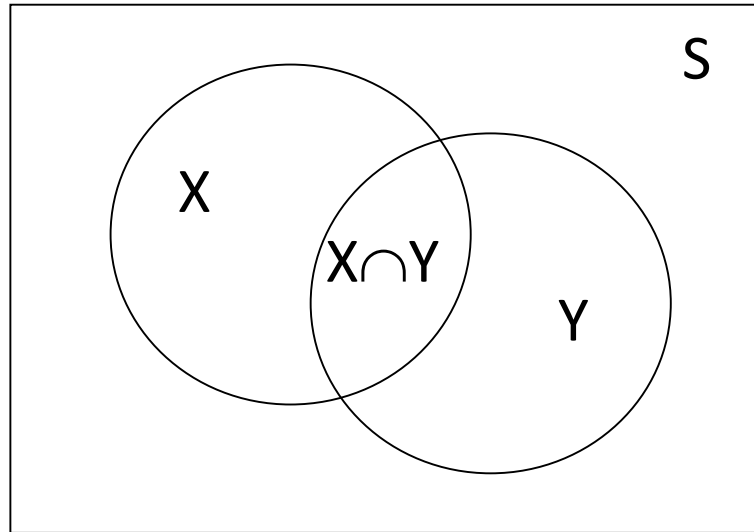
Metode Bayes

Nur Rosyid M

Mengapa Metode Bayes

- Metode Find-S tidak dapat digunakan untuk data yang tidak konsisten dan data yang bias, sehingga untuk bentuk data semacam ini salah satu metode sederhana yang dapat digunakan adalah metode bayes.
- Metode Bayes ini merupakan metode yang baik di dalam mesin pembelajaran berdasarkan data training, dengan menggunakan probabilitas bersyarat sebagai dasarnya.

Probabilitas Bersyarat



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

Probabilitas X di dalam Y adalah probabilitas interseksi X dan Y dari probabilitas Y, atau dengan bahasa lain $P(X|Y)$ adalah prosentase banyaknya X di dalam Y

Probabilitas Bersyarat Dalam Data

#	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
1	Cerah	Normal	Pelan	Ya
2	Cerah	Normal	Pelan	Ya
3	Hujan	Tinggi	Pelan	Tidak
4	Cerah	Normal	Kencang	Ya
5	Hujan	Tinggi	Kencang	Tidak
6	Cerah	Normal	Pelan	Ya

Banyaknya data berolah-raga=ya adalah 4 dari 6 data maka dituliskan

$$P(\text{Olahraga}=\text{Ya}) = 4/6$$

Banyaknya data cuaca=cerah dan berolah-raga=ya adalah 4 dari 6 data maka dituliskan

$$P(\text{cuaca}=\text{cerah dan Olahraga}=\text{Ya}) = 4/6$$

$$P(\text{cuaca} = \text{cerah} \mid \text{olahraga} = \text{ya}) = \frac{4/6}{4/6} = 1$$

Distribusi Bersama dan Distribusi Marginal

Dari 100 orang mahasiswa menunjukkan 20 orang mahasiswa menyukai keduanya, 30 orang mahasiswa menyukai bulu tangkis tapi tidak menyukai bola volley, 40 orang mahasiswa menyukai bola volley tapi tidak menyukai bulu tangkis, dan 10 orang mahasiswa tidak menyukai keduanya. Dari data ini dapat disusun bentuk distribusi bersama sebagai berikut:

Suka bulu tangkis (X)	Suka bola volley (Y)		P(X)
	Ya	Tidak	
Ya	0.2	0.3	0.5
Tidak	0.4	0.1	0.5
P(Y)	0.6	0.4	1



Distribusi Bersama



Distribusi Marginal X dan Y

Probabilitas Bersyarat Dalam Data

#	Cuaca	Temperatur	Berolahraga
1	cerah	normal	ya
2	cerah	tinggi	ya
3	hujan	tinggi	tidak
4	cerah	tinggi	tidak
5	hujan	normal	tidak
6	cerah	normal	ya

Banyaknya data berolah-raga=ya adalah 3 dari 6 data maka dituliskan

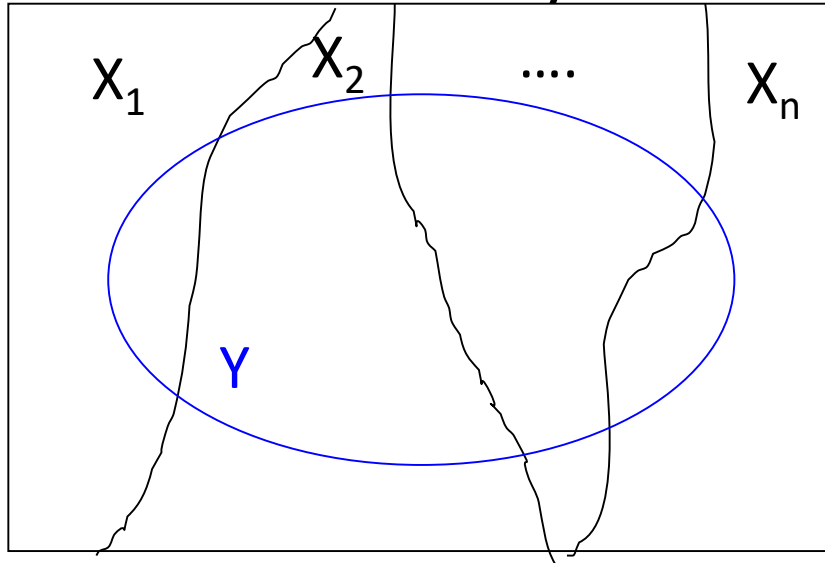
$$P(\text{Olahraga}=\text{Ya}) = 3/6$$

Banyaknya data cuaca=cerah, temperatur=normal dan berolah-raga=ya adalah 2 dari 6 data maka dituliskan

$$P(\text{cuaca}=\text{cerah}, \text{temperatur}=\text{normal}, \text{Olahraga}=\text{Ya}) = 2/6$$

$$P(\text{cuaca} = \text{cerah}, \text{temperatur} = \text{normal} \mid \text{olahraga} = \text{ya}) = \frac{2/6}{3/6} = \frac{2}{3}$$

Metode Bayes



$$P(X_k | Y) = \frac{P(Y | X_k)}{\sum_i P(Y | X_i)}$$

Keadaan Posteriror (Probabilitas X_k di dalam Y) dapat dihitung dari keadaan prior (Probabilitas Y di dalam X_k dibagi dengan jumlah dari semua probabilitas Y di dalam semua X_i)

HMAP

HMAP (*Hypothesis Maximum Appropri Probability*) menyatakan hipotesa yang diambil berdasarkan nilai probabilitas berdasarkan kondisi prior yang diketahui.

$$\begin{aligned} P(S | X) &= \operatorname{argmax}_{x \in X} \frac{P(Y | X) P(X)}{P(X)} \\ &= \operatorname{argmax}_{x \in X} P(Y | X) P(X) \end{aligned}$$

HMAP adalah model penyederhanaan dari metode bayes yang disebut dengan ***Naive Bayes***. HMAP inilah yang digunakan di dalam machine learning sebagai metode untuk mendapatkan hipotesis untuk suatu keputusan.

Contoh HMAP

Diketahui hasil survey yang dilakukan sebuah lembaga kesehatan menyatakan bahwa 30% penduduk di dunia menderita sakit paru-paru. Dari 90% penduduk yang sakit paru-paru ini 60% adalah perokok, dan dari penduduk yang tidak menderita sakit paru-paru 20% perokok.

Fakta ini bisa didefinisikan dengan: X =sakit paru-paru dan Y =perokok.

Maka : $P(X) = 0.9$

$$P(\sim X) = 0.1$$

$$P(Y|X) = 0.6 \rightarrow P(\sim Y|X) = 0.4$$

$$P(Y|\sim X) = 0.2 \rightarrow P(\sim Y|\sim X) = 0.8$$

Dengan metode bayes dapat dihitung:

$$P(\{Y\}|X) = P(Y|X).P(X) = (0.6) . (0.9) = 0.54$$

$$P(\{Y\}|\sim X) = P(Y|\sim X) P(\sim X) = (0.2).(0.1) = 0.02$$

Bila diketahui seseorang merokok, maka dia menderita sakit paru-paru karena $P(\{Y\}|X)$ lebih besar dari $P(\{Y\}|\sim X)$. HMAP diartikan mencari probabilitas terbesar dari semua instance pada atribut target atau semua kemungkinan keputusan. Pada persoalan keputusan adalah sakit paru-paru atau tidak.

HMAP Dari Data Training

#	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
1	Cerah	Normal	Pelan	Ya
2	Cerah	Normal	Pelan	Ya
3	Hujan	Tinggi	Pelan	Tidak
4	Cerah	Normal	Kencang	Ya
5	Hujan	Tinggi	Kencang	Tidak
6	Cerah	Normal	Pelan	Ya

Asumsi:

$Y = \text{berolahraga},$

$X_1 = \text{cuaca},$

$X_2 = \text{temperatur},$

$X_3 = \text{kecepatan angin}.$

Fakta menunjukkan:

$$P(Y=\text{ya}) = 4/6 \rightarrow P(Y=\text{tidak}) = 2/6$$

HMAP Dari Data Training

#	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
1	Cerah	Normal	Pelan	Ya
2	Cerah	Normal	Pelan	Ya
3	Hujan	Tinggi	Pelan	Tidak
4	Cerah	Normal	Kencang	Ya
5	Hujan	Tinggi	Kencang	Tidak
6	Cerah	Normal	Pelan	Ya

Apakah bila cuaca cerah dan kecepatan angin kencang, orang akan berolahraga?

Fakta: $P(X1=\text{cerah} | Y=\text{ya}) = 1$, $P(X1=\text{cerah} | Y=\text{tidak}) = 0$
 $P(X3=\text{kencang} | Y=\text{ya}) = 1/4$, $P(X3=\text{kencang} | Y=\text{tidak}) = 1/2$

HMAP dari keadaan ini dapat dihitung dengan:

$P(X1=\text{cerah}, X3=\text{kencang} | Y=\text{ya})$

$= \{ P(X1=\text{cerah} | Y=\text{ya}) \cdot P(X3=\text{kencang} | Y=\text{ya}) \} \cdot P(Y=\text{ya})$

$= \{ (1) \cdot (1/4) \} \cdot (4/6) = 1/6$

$P(X1=\text{cerah}, X3=\text{kencang} | Y=\text{tidak})$

$= \{ P(X1=\text{cerah} | Y=\text{tidak}) \cdot P(X3=\text{kencang} | Y=\text{tidak}) \} \cdot P(Y=\text{tidak})$

$= \{ (0) \cdot (1/2) \} \cdot (2/6) = 0$

KEPUTUSAN ADALAH BEROLAHRAGA = YA

Kelemahan Metode Bayes

- Metode Bayes hanya bisa digunakan untuk persoalan klasifikasi dengan *supervised learning* dan data-data kategorikal.
- Metode Bayes memerlukan pengetahuan awal untuk dapat mengambil suatu keputusan. Tingkat keberhasilan metode ini sangat tergantung pada pengetahuan awal yang diberikan.

Beberapa Aplikasi Metode Bayes

- Menentukan diagnosa suatu penyakit berdasarkan data-data gejala (sebagai contoh hipertensi atau sakit jantung).
- Mengenali buah berdasarkan fitur-fitur buah seperti warna, bentuk, rasa dan lain-lain
- Mengenali warna berdasarkan fitur indeks warna RGB
- Mendeteksi warna kulit (*skin detection*) berdasarkan fitur warna chrominant
- Menentukan keputusan aksi (olahraga, art, psikologi) berdasarkan keadaan.
- Menentukan jenis pakaian yang cocok untuk keadaan-keadaan tertentu (seperti cuaca, musim, temperatur, acara, waktu, tempat dan lain-lain)

Latihan Soal

- 1. Diketahui hasil survey yang dilakukan sebuah lembaga kesehatan menyatakan bahwa 30% penduduk di dunia menderita sakit paru-paru. Dari 90% penduduk yang sakit paru-paru ini 60% adalah perokok, dan dari penduduk yang tidak menderita sakit paru-paru 20% perokok. Dengan menggunakan HMAP (Hypothesis Maximum Appropri Probability), hitunglah kemungkinan seseorang perokok mengidap penyakit paru-paru, dan kemungkinan orang tersebut tidak sakit paru-paru!

2. Misalkan populasi penduduk di suatu desa dibagi menurut jenis kelamin dan status pekerjaan sebagai berikut:

Jenis kelamin	Bekerja	Menganggur	Jumlah
Laki-laki	460	40	500
Wanita	140	260	400
Jumlah	600	300	900

Bila diambil seorang dari mereka secara acak maka berapakah peluang yang terpilih adalah status yang bekerja dan berjenis kelamin laki-laki. Serta berapa peluang yang terpilih adalah status yang bekerja dan berjenis kelamin perempuan!

3. Dibawah ini adalah data Ikan Lele dan Ikan Gurami .

Panjang	Lebar	Berat	Jenis Ikan
Panjang	kecil	ringan	Ikan Lele
Panjang	sedang	Berat	Ikan Lele
Panjang	Lebar	Berat	Gurami
Sedang	kecil	Berat	Ikan Lele
Sedang	sedang	ringan	Ikan Lele
Sedang	Lebar	Berat	Gurami
Kecil	kecil	ringan	Gurami
Kecil	sedang	ringan	Gurami
Kecil	Lebar	Berat	Gurami

Jika diketahui ikan “Panjang Lebar Berat” maka termasuk ikan apakah?
hitunglah menggunakan Naïve Bayes!

4. Diketahui data orang yang membeli Laptop sebagai berikut

Umur	Gaji	Status	Hutang	Beli Laptop
<=30	tinggi	single	punya	tidak
<=30	tinggi	single	tidak	tidak
30...40	tinggi	single	punya	iya
>40	sedang	single	punya	iya
>40	rendah	menikah	punya	iya
>40	rendah	menikah	tidak	tidak
31...40	rendah	menikah	tidak	iya
<=30	sedang	single	punya	tidak
<=30	rendah	menikah	punya	iya
>40	sedang	menikah	punya	iya
<=30	sedang	menikah	tidak	iya
31...40	sedang	single	tidak	iya
31...40	tinggi	menikah	punya	iya
>40	sedang	single	tidak	tidak

- Jika ada orang dengan data $X = (\text{Umur} \leq 30, \text{Gaji} = \text{Sedang}, \text{Status} = \text{menikah}, \text{Hutang} = \text{punya})$ hitunglah kemungkinan dia beli laptop dan kemungkinan dia tidak beli laptop!