

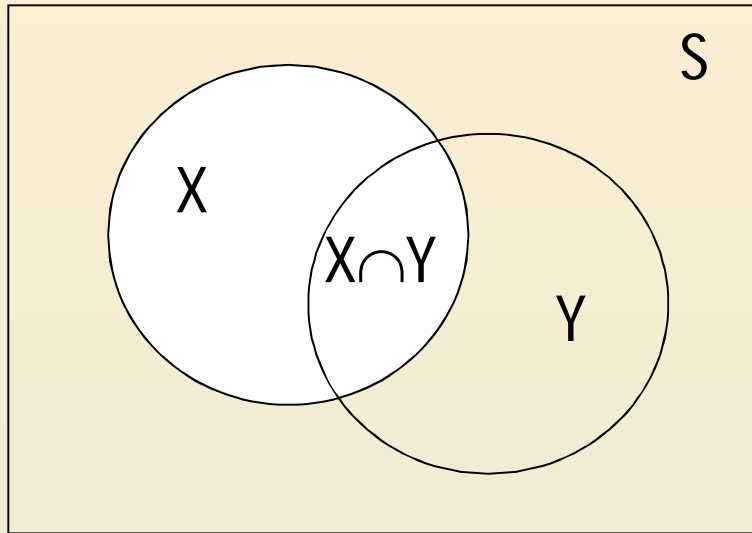
Bayesian Learning

(Based on Chapter 6 of Tom M.
Mitchell, Machine Learning, 1997)

Mengapa Bayesian Learning

- 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.
- Bayesian Learning 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

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

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

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

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

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

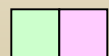
Distribusi Bersama dan Distribusi Marginal

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

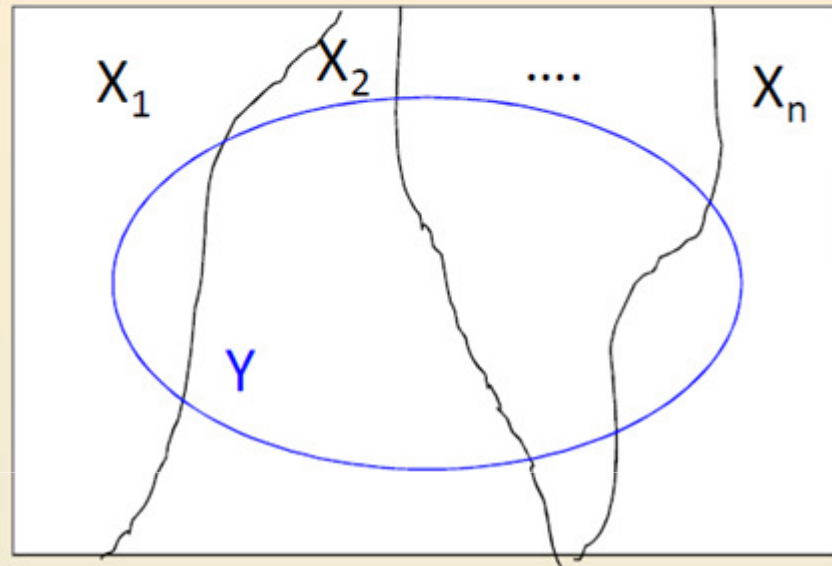


Distribusi Bersama



Distribusi Marginal X dan Y

Metode Bayes



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

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)

MAP Hypothesis

MAP (*Maximum A propri Probability*) Hypothesis menyatakan hipotesa yang diambil berdasarkan nilai probabilitas berdasarkan kondisi prior yang diketahui.

$$\begin{aligned} h_{\text{MAP}} &= \operatorname{argmax}_{h \in H} P(h \mid D) \\ &= \operatorname{argmax}_{h \in H} \frac{P(D \mid h) P(h)}{P(D)} \\ &= \operatorname{argmax}_{h \in H} P(D \mid h) P(h) \end{aligned}$$

Contoh MAP Hypotheses

Dalam suatu populasi terdapat 0.008 penduduk yang menderita sakit kanker. Dari tes untuk positif kanker, yang terkena sakit kanker sebesar 98%. Dari tes untuk bebas kanker, yang tidak sakit kanker sebesar 97%.

$$P(kanker) = 0.008$$

$$P(+ | kanker) = 0.98$$

$$P(+ | \neg kanker) = 0.03$$

$$P(\neg kanker) = 0.992$$

$$P(- | kanker) = 0.02$$

$$P(- | \neg kanker) = 0.97$$

$$P(+ | kanker) P(kanker) = 0.98 \cdot 0.008 = 0.0078$$

$$P(+ | \neg kanker) P(\neg kanker) = 0.03 \cdot 0.992 = 0.0298$$

$$hMAP = \neg kanker$$

$$P(kanker / +) = 0.0078 / (0.0078 + 0.0298) = 0.21$$

Bayes Theorem

- **Goal:** To determine the most probable hypothesis, given the data D plus any initial knowledge about the prior probabilities of the various hypotheses in H .
- ***Prior probability of h , $P(h)$:*** it reflects any background knowledge we have about the chance that h is a correct hypothesis (before having observed the data).
- ***Prior probability of D , $P(D)$:*** it reflects the probability that training data D will be observed given no knowledge about which hypothesis h holds.
- ***Conditional Probability of observation D , $P(D|h)$:*** it denotes the probability of observing data D given some world in which hypothesis h holds.

Bayes Theorem

- **Posterior probability of h , $P(h/D)$:** it represents the probability that h holds given the observed training data D . It reflects our confidence that h holds after we have seen the training data D and it is the quantity that Machine Learning researchers are interested in.
- **Bayes Theorem** allows us to compute $P(h/D)$:

$$P(h/D) = P(D/h)P(h)/P(D)$$

Maximum A Posteriori (MAP) Hypothesis and Maximum Likelihood

- **Goal:** To find the most probable hypothesis h from a set of candidate hypotheses H given the observed data D .
- **MAP Hypothesis, $h_{MAP} = \operatorname{argmax}_{h \in H} P(h|D)$**
$$= \operatorname{argmax}_{h \in H} P(D|h)P(h)/P(D)$$
$$= \operatorname{argmax}_{h \in H} P(D|h)P(h)$$
- If every hypothesis in H is equally probable a priori, we only need to consider the likelihood of the data D given h , $P(D|h)$. Then, h_{MAP} becomes the **Maximum Likelihood**,
$$h_{ML} = \operatorname{argmax}_{h \in H} P(D|h)$$

Bayes Optimal Classifier

- One great advantage of Bayesian Decision Theory is that it gives us a lower bound on the classification error that can be obtained for a given problem.
- **Bayes Optimal Classification:** The most probable classification of a new instance is obtained by combining the predictions of all hypotheses, weighted by their posterior probabilities:

$$\operatorname{argmax}_{v_j \in V} \sum_{h_i \in H} P(v_h / h_i) P(h_i / D)$$

where V is the set of all the values a classification can take and v_j is one possible such classification.

- Unfortunately, Bayes Optimal Classifier is usually too costly to apply! ==> ***Naïve Bayes Classifier***

Naïve Bayes Classifier

- Let each instance x of a training set D be described by a conjunction of n attribute values $\langle a_1, a_2, \dots, a_n \rangle$ and let $f(x)$, the target function, be such that $f(x) \in V$, a finite set.

- Bayesian Approach:**

$$\begin{aligned} v_{MAP} &= \operatorname{argmax}_{v_j \in V} P(v_j | a_1, a_2, \dots, a_n) \\ &= \operatorname{argmax}_{v_j \in V} [P(a_1, a_2, \dots, a_n | v_j) P(v_j) / P(a_1, a_2, \dots, a_n)] \\ &= \operatorname{argmax}_{v_j \in V} [P(a_1, a_2, \dots, a_n | v_j) P(v_j)] \end{aligned}$$

- Naïve Bayesian Approach:** We assume that the attribute values are conditionally independent so that $P(a_1, a_2, \dots, a_n | v_j) = \prod_i P(a_i | v_j)$ [and not too large a data set is required.]

- Naïve Bayes Classifier:**

$$v_{NB} = \operatorname{argmax}_{v_j \in V} P(v_j) \prod_i P(a_i | v_j)$$

Klasifikasi dengan Naïve Bayes

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

Hasil perhitungan Naive Bayes Classifier :

$$\begin{aligned} &P(X1=\text{cerah}, X3=\text{kencang} | Y=\text{ya}) \\ &= \{ P(X1=\text{cerah} | Y=\text{ya}) \cdot P(X3=\text{kencang} | Y=\text{ya}) \} \cdot P(Y=\text{ya}) \\ &= \{ (1) \cdot (1/4) \} \cdot (4/6) = 1/6 \\ &P(X1=\text{cerah}, X3=\text{kencang} | Y=\text{tidak}) \\ &= \{ P(X1=\text{cerah} | Y=\text{tidak}) \cdot P(X3=\text{kencang} | Y=\text{tidak}) \} \cdot P(Y=\text{tidak}) \\ &= \{ (0) \cdot (1/2) \} \cdot (2/6) = 0 \end{aligned}$$

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)

Referensi

- Tom M. Mitchell, Machine Learning, McGraw-Hill, 1997.
- Achmad Basuki, Metode Bayes, PENS, 2006.