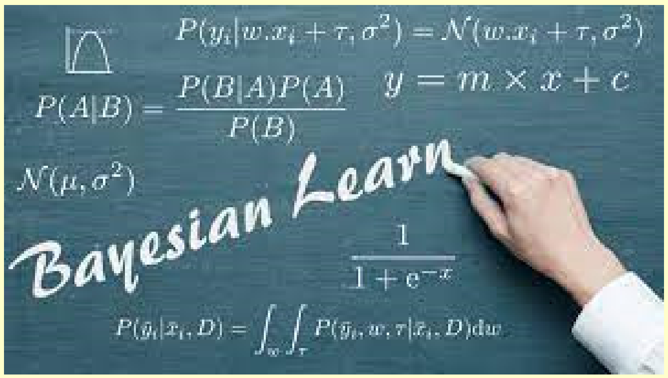


Bayesian Learning

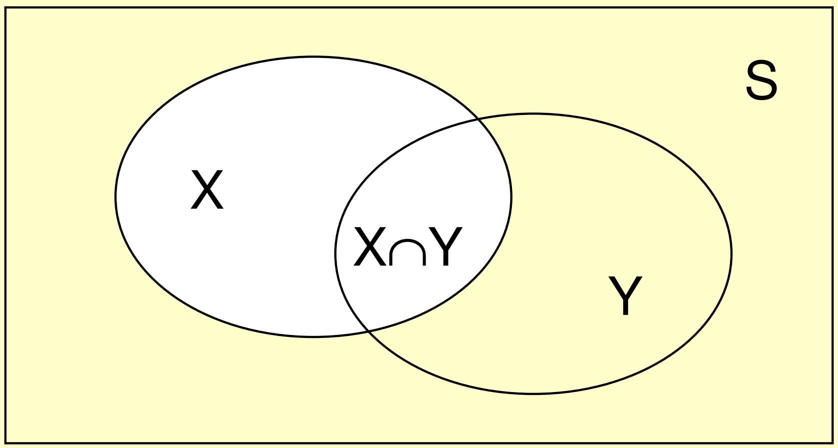
Aliridho Barakbah

Knowledge Engineering Research Group
Department of Information and Computer Engineering
Politeknik Elektronika Negeri Surabaya

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



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

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(Olahraga=Ya) = 4/6

Banyaknya data cuaca=cerah dan berolah-raga=ya adalah 4 dari 6 data maka dituliskan
P(cuaca=cerah dan Olahraga=Ya) = 4/6

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

Probabilitas Bersyarat Dalam Data

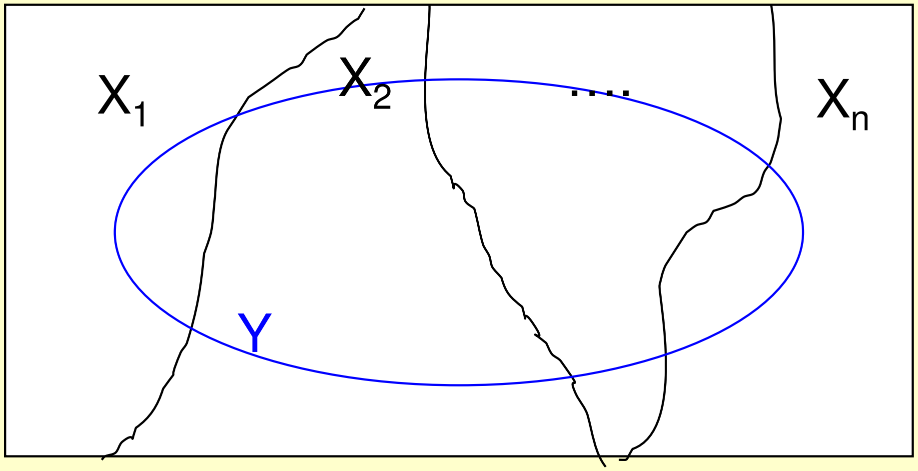
#	Cuaca	Temperatur	Berolahraga
1	cerah	normal	ya
2	cerah	tinggi	ya
3	hujan	tinggi	tidak
4	cerah	tinggi	tidak
5	hujan	normal	tidak
6	cerah	normal	ya

Banyaknya data berolah-raga=ya adalah 3 dari 6 data maka dituliskan
P(Olahraga=Ya) = 3/6

Banyaknya data cuaca=cerah, temperatur=normal dan berolah-raga=ya adalah 4 dari 6 data maka dituliskan
P(cuaca=cerah, temperatur=normal, Olahraga=Ya) = 2/6

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

Bayes Theorem



$$P(X_k | Y) \quad \frac{P(X_k \cap Y)}{P(Y)}$$

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

$$P(X_k \cap Y) = P(Y | X_k)\, P(X_k)$$

$$P(Y) = \sum_i P(Y | X_i)$$

Keadaan Posterior (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)

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

Bayes Theorem in Data

- 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):***

$$\boxed{P(h|D)=P(D|h)P(h)/P(D)}$$

MAP Hyphotesis

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

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

Contoh MAP Hypotheses

Terdapat 0.8% penderita kanker di dunia. Untuk membuktikan seseorang sakit kanker, dilakukan pendekatan dengan alat tes. Dari yang menderita kanker, terdapat 98% menunjukkan hasil tes kanker yang positif. Dari yang tidak menderita kanker, terdapat 97% menunjukkan hasil tes kanker yang negatif. Jika seseorang melakukan tes kanker dan hasil tesnya positif, maka manakah yang lebih besar kemungkinannya dia menderita kanker atau tidak?

P(kanker | +)
P(~kanker | +) ?

P(kanker) = 0.008 → *P(~kanker) = 0.992*
P(+ | kanker) = 0.98 → *P(- | kanker) = 0.02*
P(+ | ~kanker) = 0.03 → *P(- | ~kanker) = 0.97*

P(kanker | +) = P(+ | kanker).P(kanker) = 0.98 x 0.008 = 0.0078
P(~kanker | +) = P(+ | ~ kanker) P(~ kanker) = 0.03 x 0.992 = 0.0298

h_{MAP} = *P(~kanker | +)*

Contoh MAP Hypotheses

Diketahui hasil survey yang dilakukan sebuah lembaga kesehatan menyatakan bahwa 0.9% penduduk menderita penyakit paru-paru. Dari yang menderita penyakit paru-paru, 98% adalah perokok. Dari yang tidak menderita paru-paru, 99.4% adalah bukan perokok. Jadi, lebih besar mana kemungkinannya, seseorang yang perokok itu menderita sakit paru-paru atau tidak?

P(paru | perokok)
P(~paru | perokok) ?

P(paru) = 0.009 → *P(~paru) = 0.991*
P(perokok | paru) = 0.98 → *P(~perokok | paru) = 0.02*
P(~perokok | paru) = 0.006 → *P(~perokok | ~paru) = 0.994*

P(paru | perokok) = P(perokok | paru) . P(paru) = 0.98 x 0.009 = 0.0088
P(~paru | perokok) = P(perokok | ~paru) . P(~ paru) = 0.006 x 0.991 = 0.005946

h_{MAP} = *P(paru | perokok)*

Contoh MAP Hypotheses

Pada tahun 2020 terdapat 0.7% penderita Covid di dunia. Untuk membuktikan seseorang terpapar Covid, dilakukan pendekatan dengan Rapid-Test. Dari yang menderita Covid, terdapat 98.9% menunjukkan hasil tes yang reaktif (R). Dari yang tidak menderita Covid, terdapat 99.6% menunjukkan hasil tes yang non-reaktif (NR). Jika seseorang melakukan Rapid-Test dan hasil tesnya reaktif, maka manakah yang lebih besar kemungkinannya dia menderita Covid atau tidak?

P(covid | R)
P(~covid | R) ?

P(covid) = 0.007 → *P(~covid) = 0.993*
P(R | covid) = 0.989 → *P(NR | covid) = 0.011*
P(R | ~covid) = 0.004 → *P(NR | ~covid) = 0.996*

P(covid | R) = P(R | covid).P(covid) = 0.989 x 0.007 = 0.006923
P(~covid | R) = P(R | ~covid) P(~covid) = 0.004 x 0.993 = 0.003972

h_{MAP} = *P(covid | R)*

Maximum A Posteriori (MAP) Hypothesis and Maximum Likelihood in Data

- **Goal:** To find the most probable hypothesis *h* from a set of candidate hypotheses *H* given the observed data *D*.

- ***MAP Hypothesis, hMAP*** $= \textbf{argmax}_{h \in H} \, P(h|D)$

$$= \textbf{argmax}_{h \in H} \, P(D|h)P(h)/P(D)$$

$$= \textbf{argmax}_{h \in H} \, P(D|h)P(h)$$

- If every hypothesis in *H* is equally probable a priori, we onlyneed to consider the likelihood of the data *D* given *h*, ***P(D|h).*** Then, h_{MAP} becomes the ***Maximum Likelihood***,

$$h_{ML} = \textbf{argmax}_{h \in H} \, P(D|h)$$

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:

$$\boldsymbol{argmax_{vj \in V} h \sum_{hi \in H} P(vh|hi)P(hi|D)}$$

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 <$a_1, a_2, .., a_n$> and let *f(x)*, the target function, be such that $f(x) \in V$, a finite set.

- **Bayesian Approach:**

$$vMAP = \operatorname{argmax}_{v_j \in V} P(v_j | a_1, a_2, .., a_n)$$
$$= \operatorname{argmax}_{v_j \in V} [P(a_1, a_2, .., a_n | v_j) P(v_j) / P(a_1, a_2, .., a_n)]$$
$$= \operatorname{argmax}_{v_j \in V} [P(a_1, a_2, .., a_n | v_j) P(v_j)$$

- **Naïve Bayesian Approach:** We assume that the attribute values are conditionally independent so that $P(a1,a2,..,an|vj) = \prod_i P(a_i|v_j)$ [and not too large a data set is required.]

- ***Naïve Bayes Classifier:***

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

Naïve Bayes Classifier

#	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=cerah|Y=ya) = 1, P(X1=cerah|Y=tidak) = 0
P(X3=kencang|Y=ya) = 1/4 , P(X3=kencang|Y=tidak) = 1/2

P(X1=cerah,X3=kencang | Y=ya)
= { P(X1=cerah|Y=ya).P(X3=kencang|Y=ya) } . P(Y=ya)
= { (1) . (1/4) } . (4/6) = 1/6

P(X1=cerah,X3=kencang | Y=tidak)
= { P(X1=cerah|Y=tidak).P(X3=kencang|Y=tidak) } . P(Y=tidak)
= { (0) . (1/2) } . (2/6) = 0

KEPUTUSAN ADALAH BEROLAHRAG A = YA

Kelemahan Metode Bayes

- Metode Bayes hanya bisa digunakan untuk persoalan klasifikasi dengan *supervised learning*.
- 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*) berdarkan 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)
- Menentukan ekspresi (sedih, gembira, dll) dari kalimat yang diucapkan