

Naïve Bayes

Fajar Agung Nugroho, S.Kom, M.CS
Fakultas Ilmu Komputer
Universitas Dian Nuswantoro

Definisi

- Adalah metode classifier yang berdasarkan probabilitas dan Teorema Bayesian dengan asumsi bahwa setiap variabel X bersifat bebas (*independence*)

Bayes's Rule

- Karena asumsi atribut tidak saling terkait (*conditionally independent*), maka:

$$\text{Bayes's rule: } P(H | E) = \frac{P(E | H) \times P(H)}{P(E)}$$

Example

| | outlook | temperature | humidity | windy | play |
|----|----------|-------------|----------|-------|------|
| 1 | sunny | hot | high | false | no |
| 2 | sunny | hot | high | true | no |
| 3 | overcast | hot | high | false | yes |
| 4 | rainy | mild | high | false | yes |
| 5 | rainy | cool | normal | false | yes |
| 6 | rainy | cool | normal | true | no |
| 7 | overcast | cool | normal | true | yes |
| 8 | sunny | mild | high | false | no |
| 9 | sunny | cool | normal | false | yes |
| 10 | rainy | mild | normal | false | yes |
| 11 | sunny | mild | normal | true | yes |
| 12 | overcast | mild | high | true | yes |
| 13 | overcast | hot | normal | false | yes |
| 14 | rainy | mild | high | true | no |

| Outlook | Temp. | Humidity | Windy | Play |
|---------|-------|----------|-------|------|
| Sunny | Cool | High | True | ? |

Posterior Probability

| Outlook | Temp. | Humidity | Windy | Play |
|---------|-------|----------|-------|------|
| Sunny | Cool | High | True | ? |

← *Evidence E*

$$\begin{aligned}\Pr[\text{yes} | E] &= \Pr[\text{Outlook} = \text{Sunny} | \text{yes}] \\ &\quad \times \Pr[\text{Temperature} = \text{Cool} | \text{yes}] \\ &\quad \times \Pr[\text{Humidity} = \text{High} | \text{yes}] \\ &\quad \times \Pr[\text{Windy} = \text{True} | \text{yes}] \\ &\quad \times \frac{\Pr[\text{yes}]}{\Pr[E]}\end{aligned}$$

↑
*Probability of
class "yes"*

$$= \frac{\frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}}{\Pr[E]}$$

Result

Likelihood of the two classes

$$\text{For "yes"} = 2/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14 = 0.0053$$

$$\text{For "no"} = 3/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14 = 0.0206$$

Conversion into a probability by normalization:

$$P(\text{"yes"}) = 0.0053 / (0.0053 + 0.0206) = 0.205$$

$$P(\text{"no"}) = 0.0206 / (0.0053 + 0.0206) = 0.795$$

Exercise

- Given data :

| chills | runny nose | headache | fever | flu? |
|--------|------------|----------|-------|------|
| Y | N | Mild | Y | Y |
| Y | Y | No | N | Y |
| Y | N | Strong | Y | Y |
| N | Y | Mild | Y | Y |
| N | N | No | N | N |
| N | Y | Strong | Y | Y |
| N | Y | Strong | N | N |
| Y | Y | Mild | Y | Y |

- New data

| chills | runny nose | headache | fever | flu? |
|--------|------------|----------|-------|------|
| Y | N | Mild | Y | ? |

Referensi

- **A Tutorial on Naive Bayes Classification,** Choochart Haruechaiyasak
- Mitchell, 1996 and Russel & Norvig, 2003
- **Naïve Bayesian Classifier,** Dr. Taufik Fuadi Abidin, S.Si., M.Tech