Naïve Bayes

Kookjin Lee

(kookjin.Lee@asu.edu)

Your second generative learning algorithm

- Naïve Bayes classification
  - Bayes rule:

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

Training naïve Bayes

- For each class, calculate probability given features

$$p(y|x) = p(x|y)p(y)$$

- Difficult to work directly on the joint probability:

$$p(y|x) = p(x_1, x_2, \ldots, x_n|y)p(y)$$

  - Expansion:

$$p(y|x) = p(x_1|x_2, \ldots, x_n, y)p(x_2, \ldots, x_n|y)p(y)$$

$$p(y|x) = p(x_1|x_2, \ldots, x_n, y)p(x_2|x_3, \ldots, x_n, y)p(x_3, \ldots, x_n|y)p(y)$$

Training naïve Bayes: the naïve assuption

- Assumption: assume all features independent of each other

$$p(y|x) = p(x_1|y)p(x_2|y)\cdots p(x_n|y)p(y)$$

$$p(y|x) = p(y)\prod_{i=1}^{n} p(x_i|y)$$

- Maximum a posteriori (MAP) rule:

$$\arg\max_{k\in\{1,\ldots,K\}} \quad p(y_k)\prod_{i=1}^{n} p(x_i|y_k)$$

Example 1

-

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Example 1

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$P(Play=Yes) = 9/14$

$P(Play=No) = 5/14$

Example 1 (continued)

- Create probability lookup tables based on training data

P(Play=Yes) = 9/14

P(Play=No) = 5/14

Outlook	Play=Yes	Play=No
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Temperature	Play=Yes	Play=No
Hot	2/9	2/5
Mild	4/9	2/5
Cool	3/9	1/5

Humidity	Play=Yes	Play=No
High	3/9	4/5
Normal	6/9	1/5

Wind	Play=Yes	Play=No
Strong	3/9	3/5
Weak	6/9	2/5

Example 1 (continued)

- Predict the outcome (class) if
  - Outlook = Sunny, Temperature = Cool, Humidity = High, Wind = Strong

$$P(yes|sunny, cool, high, strong) = P(sunny|yes) * P(cool|yes) * P(high|yes) * P(strong|yes) * P(yes)$$

$$P(no|sunny, cool, high, strong) = P(sunny|no) * P(cool|no) * P(high|no) * P(strong|no) * P(no)$$

Example 1 (continued)

- Predict the outcome (class) if
  - Outlook = Sunny, Temperature = Cool, Humidity = High, Wind = Strong

$$P(yes|sunny, cool, high, strong) = P(sunny|yes) * P(cool|yes) * P(high|yes) * P(strong|yes) * P(yes)$$

$$P(no|sunny, cool, high, strong) = P(sunny|no) * P(cool|no) * P(high|no) * P(strong|no) * P(no)$$

Feature	Play=Yes	Play=No
Outlook=Sunny	2/9	3/5
Temperature=Cool	3/9	1/5
Humidity=High	3/9	4/5
Wind=Strong	3/9	3/5
Overall Label	**9/14**	**5/14**
Probability	**0.0053**	**0.0206**

Example 2:

- Spam filtering

$$x = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \begin{array}{l} \text{a} \\ \text{aardvark} \\ \text{aardwolf} \\ \vdots \\ \text{buy} \\ \vdots \\ \text{zygmurgy} \end{array}$$

$$x \in \{0, 1\}^{50000}$$

Example 2:

- Spam filtering

$$p(\text{spam}) = \frac{\sum_{i=1}^{n} \mathbb{1}[y^{(i)} = 1]}{n} \qquad p(\text{not spam}) = 1 - p(\text{spam})$$

$$p(x_j = 1 | y = 1) = \frac{\sum_{i=1}^{n} \mathbb{1}[x_j^{(i)} = 1 \wedge y^{(i)} = 1]}{\sum_{i=1}^{n} \mathbb{1}[y^{(i)} = 1]}$$

$$p(x_j = 1 | y = 0) = \frac{\sum_{i=1}^{n} \mathbb{1}[x_j^{(i)} = 1 \wedge y^{(i)} = 0]}{\sum_{i=1}^{n} \mathbb{1}[y^{(i)} = 0]}$$

Example 2:

- Spam filtering

$$p(\text{spam}) = \frac{\sum_{i=1}^{n} \mathbb{1}[y^{(i)} = 1]}{n} \qquad p(\text{not spam}) = 1 - p(\text{spam})$$

$$p(x_j = 1 | y = 1) = \frac{\sum_{i=1}^{n} \mathbb{1}[x_j^{(i)} = 1 \wedge y^{(i)} = 1]}{\sum_{i=1}^{n} \mathbb{1}[y^{(i)} = 1]} \qquad p(x_j = 1 | y = 0) = \frac{\sum_{i=1}^{n} \mathbb{1}[x_j^{(i)} = 1 \wedge y^{(i)} = 0]}{\sum_{i=1}^{n} \mathbb{1}[y^{(i)} = 0]}$$

$$p(y = 1 | x) = p(y = 1) \prod_{i=1}^{n} p(x_j | y = 1)$$

$$p(y = 0 | x) = p(y = 0) \prod_{i=1}^{n} p(x_j | y = 0)$$