

# Naïve Bayes

Kookjin Lee

([kookjin.Lee@asu.edu](mailto:kookjin.Lee@asu.edu))

The contents of this course, including lectures and other instructional materials, are copyrighted materials. Students may not share outside the class, including uploading, selling or distributing course content or notes taken during the conduct of the course. Any recording of class sessions is authorized only for the use of students enrolled in this course during their enrollment in this course. Recordings and excerpts of recordings may not be distributed to others. (see ACD 304 – 06 , “ Commercial Note Taking Services ” and ABOR Policy 5 - 308 F.14 for more information).

# 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:

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

- Expansion:

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

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

# Training naïve Bayes: the naïve assumption

- 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:

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

# 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

$$P(\text{Play}=\text{Yes}) = 9/14$$

$$P(\text{Play}=\text{No}) = 5/14$$

# Example 1 (continued)

- Create probability lookup tables based on training data

$$P(\text{Play}=\text{Yes}) = 9/14$$

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

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

$$P(\text{Play}=\text{No}) = 5/14$$

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

<b>Wind</b>	<b>Play=Yes</b>	<b>Play=No</b>
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

$$P(\text{yes}|\text{sunny, cool, high, strong}) = P(\text{sunny}|\text{yes}) * P(\text{cool}|\text{yes}) * \\ P(\text{high}|\text{yes}) * P(\text{strong}|\text{yes}) * P(\text{yes})$$

$$P(\text{no}|\text{sunny, cool, high, strong}) = P(\text{sunny}|\text{no}) * P(\text{cool}|\text{no}) * \\ P(\text{high}|\text{no}) * P(\text{strong}|\text{no}) * P(\text{no})$$



# Example 1 (continued)

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

$$P(\text{yes}|\text{sunny, cool, high, strong}) = P(\text{sunny}|\text{yes}) * P(\text{cool}|\text{yes}) * \\ P(\text{high}|\text{yes}) * P(\text{strong}|\text{yes}) * P(\text{yes})$$

$$P(\text{no}|\text{sunny, cool, high, strong}) = P(\text{sunny}|\text{no}) * P(\text{cool}|\text{no}) * \\ P(\text{high}|\text{no}) * P(\text{strong}|\text{no}) * P(\text{no})$$

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
<b>Overall Label</b>	<b>9/14</b>	<b>5/14</b>
<b>Probability</b>	<b>0.0053</b>	<b>0.0206</b>

## 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

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

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

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

## Example 2:

- Spam filtering

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

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

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

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