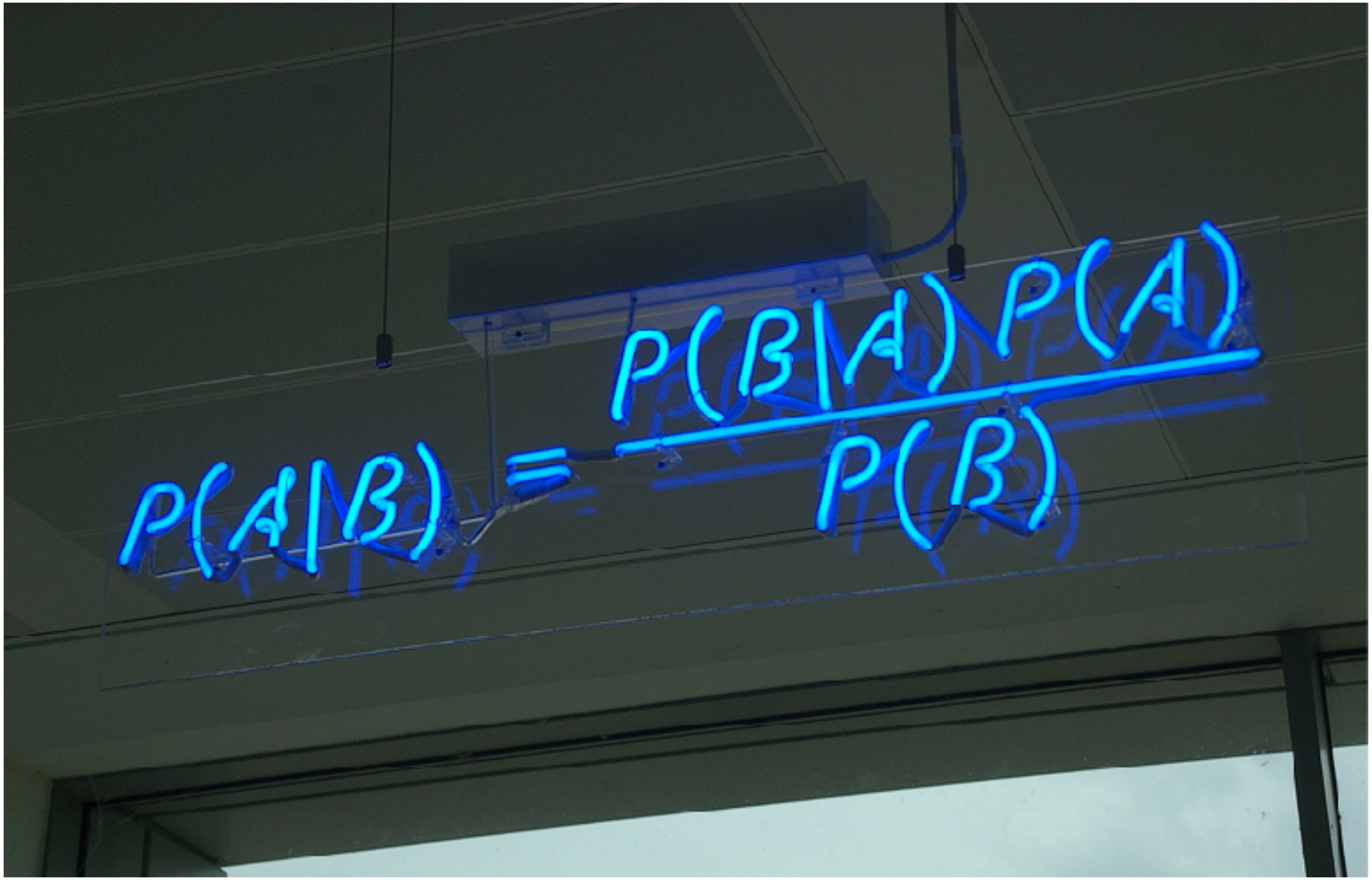

Naïve Bayes - Categorical Independent Variables

Leandro L. Minku

From Previous Lecture…

Training Set

Perso	x_1 (Wash)	x_2 (Pain)	y (Cavity)
P1	no	yes	yes
P2	no	yes	yes
P3	yes	yes	yes
P4	yes	no	no
P5	yes	no	no
P6	no	no	no

$$P(c|a_1,\ldots,a_d) = \alpha\ P(c)\ P(a_1,\ldots,a_d|c)$$

Model

Frequency Table	Cavity = no	Cavity = yes	Total:
Wash=no and Pain=no	1	0	1
Wash=no and Pain=yes	0	2	2
Wash=yes and Pain=no	2	0	2
Wash=yes and Pain=yes	0	1	1
Total:	3	3	6

Naïve Bayes

- Assumes that each input variable is conditionally independent of all other input variables given the output.
- Conditional independence:
  - X_1 is conditionally independent of X_2 given Y if the following is satisfied:
  - $P(X_1 \mid X_2, Y) = P(X_1 \mid Y)$

If we know the value of Y, we don't need to know the value of X_2 in order to determine the value of X_1.

Naïve Bayes

$$P(c|a_1,\ldots,a_d) = \alpha\, P(c)\, P(a_1,\ldots,a_d|c)$$

$$\downarrow$$

$$P(c|a_1,\ldots,a_d) = \alpha\, P(c)\, P(a_1|c)\, P(a_2|c) \ldots P(a_d|c)$$

$$P(c|a_1,\ldots,a_d) = \alpha\, P(c) \prod_{i=1}^{d} P(a_i|c)$$

where $\alpha = 1 / \beta$ and $\beta = \sum_{c \in Y} \left(P(c) \prod_{i=1}^{d} P(a_i|c) \right)$

Naïve Bayes predicts the class with the maximum $P(c|a_1,\ldots,a_d)$.

Example

Training Set

Person	x_1 (Wash)	x_2 (Pain)	y (Cavity)
P1	no	yes	yes
P2	no	yes	yes
P3	yes	yes	yes
P4	yes	no	no
P5	yes	no	no
P6	no	no	no

Problem: number of possible combinations of input attribute values becomes very large when the number of input attributes and values is large.

Model

Frequency Table	Cavity = no	Cavity = yes	Total:
Wash=no and Pain=no	1	0	1
Wash=no and Pain=yes	0	2	2
Wash=yes and Pain=no	2	0	2
Wash=yes and Pain=yes	0	1	1
Total:	3	3	6

Example Using Naïve Bayes

Training Set

Person	x_1 (Wash)	x_2 (Pain)	y (Cavity)
P1	no	yes	yes
P2	no	yes	yes
P3	yes	yes	yes
P4	yes	no	no
P5	yes	no	no
P6	no	no	no

Number of rows grows linearly with the number of independent variable values.

Model

Frequency Table for Wash	Cavity = no	Cavity = yes	Total:
Wash=no	1	2	3
Wash=yes	2	1	3
Total:	3	3	6

Frequency Table for Pain	Cavity = no	Cavity = yes	Total:
Pain=no	3	0	3
Pain=yes	0	3	3
Total:	3	3	6

Example of Prediction

Based on the frequency tables below, determine the predicted class for the following instance: (Wash=yes, Pain=yes, y = ?)

Frequency Table for Wash	Cavity = no	Cavity = yes	Total:
Wash=no	1	2	3
Wash=yes	2	1	3
Total:	3	3	6

Frequency Table for Pain	Cavity = no	Cavity = yes	Total:
Pain=no	3	0	3
Pain=yes	0	3	3
Total:	3	3	6

$$P(c|a_1,\ldots,a_d) = \alpha\, P(c) \prod_{i=1}^{d} P(a_i|c)$$

$P(\text{Cavity=no} \mid \text{Wash=yes,Pain=yes}) =$

$= \alpha\, P(\text{Cavity=no})\, P(\text{Wash=yes}|\text{Cavity=no}) P(\text{Pain=yes}|\text{Cavity=no})$

$= \alpha\ 3/6 * 2/3 * 0/3 = 0 * \alpha = 0\%$

Example of Prediction

Based on the frequency tables below, determine the predicted class for the following instance: (Wash=yes, Pain=yes, y = ?)

Frequency Table for Wash	Cavity = no	Cavity = yes	Total:
Wash=no	1	2	3
Wash=yes	2	1	3
Total:	3	3	6

Frequency Table for Pain	Cavity = no	Cavity = yes	Total:
Pain=no	3	0	3
Pain=yes	0	3	3
Total:	3	3	6

$$P(c|a_1,\ldots,a_d) = \alpha \, P(c) \prod_{i=1}^{d} P(a_i|c)$$

$$P(\text{Cavity=yes} \mid \text{Wash=yes,Pain=yes}) =$$

$$= \alpha \, P(\text{Cavity=yes}) \, P(\text{Wash=yes}|\text{Cavity=yes}) \, P(\text{Pain=yes}|\text{Cavity=yes})$$

$$= \alpha \; 3/6 * 1/3 * 3/3 = 1/6 * \alpha$$

Example of Prediction

P(Cavity=no | Wash=yes,Pain=yes) =

= α P(Cavity=no) P(Wash=yes|Cavity=no)P(Pain=yes|Cavity=no)

= α 3/6 *2/3 * 0/3 = 0 *α = 0%

P(Cavity=yes | Wash=yes,Pain=yes) =

= α P(Cavity=yes) P(Wash=yes|Cavity=yes) P(Pain=yes|Cavity=yes)

= α 3/6 * 1/3 * 3/3 = 1/6*α = $\boxed{100\%}$ Predicted class = yes

$$\alpha = \frac{1}{3/6 * 2/3 * 0/3 + 3/6 * 1/3 * 3/3}$$

Problem: because there were no examples of Cavity=no with Pain=yes in the training set, no matter the value for Wash, P(Cavity=no | Wash=?, Pain=yes) = 0%

Example of Prediction

P(Cavity=no | Wash=yes,Pain=yes) =

= α P(Cavity=no) P(Wash=yes|Cavity=no)P(Pain=yes|Cavity=no)

= α 3/6 *2/3 * 0/3 = 0 *α = 0%

P(Cavity=yes | Wash=yes,Pain=yes) =

= α P(Cavity=yes) P(Wash=yes|Cavity=yes) P(Pain=yes|Cavity=yes)

= α 3/6 * 1/3 * 3/3 = 1/6*α = $\boxed{100\%}$ Predicted class = yes

$$\alpha = \frac{1}{3/6 * 2/3 * 0/3 \;+\; 3/6 * 1/3 * 3/3}$$

However, clearly there exist cases where a person has pain, but no cavity!

Laplace Smoothing

Training Set

Perso	x_1 (Wash)	x_2 (Pain)	y (Cavity)
P1	no	yes	yes
P2	no	yes	yes
P3	yes	yes	yes
P4	yes	no	no
P5	yes	no	no
P6	no	no	no

Model

Frequency Table for Wash	Cavity = no	Cavity = yes	Total:
Wash=no	1+1	2+1	3+2
Wash=yes	2+1	1+1	3+2
Total:	3+2	3+2	6+4

Frequency Table for Pain	Cavity = no	Cavity = yes	Total:
Pain=no	3+1	0+1	3+2
Pain=yes	0+1	3+1	3+2
Total:	3+2	3+2	6+4

To fix this problem, we can add 1 to each of the frequency cells and use that when calculating $P(a_i|c)$.

We calculate P(c) using the original frequencies.

Naïve Bayes Approach

- Naïve Bayes Learning Algorithm:
  - Create frequency tables: for each independent variable and its possible values, count number of training examples of each class with each possible independent variable value.
  - Apply Laplace Smoothing.
- Naïve Bayes Model:
  - Frequency tables (with and without Laplace Smoothing), associated with the Bayes Theorem using the conditional independence assumption for predictions.
- Naïve Bayes prediction for an instance (**x**,?):
  - Use Bayes Theorem with conditional independence assumption.

Quiz

- Consider the Naïve Bayes model from slide 11.
- Compute the following probabilities (using Laplace Smoothing):
    - P(Cavity=no | Wash=yes,Pain=no)
    - P(Cavity=yes | Wash=yes,Pain=no)
- Based on these probabilities, what class should be predicted?

Further Reading

- Essential:
  - Leandro Minku’s notes on “Naïve Bayes for Categorical Independent Variables.
- Recommended:
  - Bishop’s “Machine Learning and Pattern Recognition”, pages 45-46 (Combining Models).
- Background:
  - Russell and Norvig’s “Artificial Intelligence: A Modern Approach”
    - Section 13.2 (Basic Probability Notation) up to the end of section 13.6(The Wumpus World Revisited).