

Outline

- Background
- Probability Basics
- Probabilistic Classification
- Naïve Bayes
- Example: Play Tennis
- Relevant Issues
- Conclusions

The Bayes Classifier

- Use Bayes Rule!

$$P(Y|X_1, \dots, X_n) = \frac{\overset{\text{Likelihood}}{P(X_1, \dots, X_n|Y)} \overset{\text{Prior}}{P(Y)}}{\underset{\text{Normalization Constant}}{P(X_1, \dots, X_n)}}$$

- $P(Y|X)$ is Posterior probability: Probability of hypothesis Y on the observed event X .
- $P(X|Y)$ is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.
- $P(Y)$ is Prior Probability: Probability of hypothesis before observing the evidence.
- $P(X)$ is Marginal Probability: Probability of Evidence

The Bayes Classifier

- A good strategy is to predict:

$$\arg \max_Y P(Y | X_1, \dots, X_n)$$

- So ... How do we compute that?

The Naïve Bayes Model

- The *Naïve Bayes Assumption*: Assume that all features are independent **given the class label Y**
- Equationally speaking:

$$P(y|x_1, \dots, x_n) = \frac{P(x_1|y)P(x_2|y)\dots P(x_n|y)P(y)}{P(x_1)P(x_2)\dots P(x_n)}$$

$$P(X_1, \dots, X_n|Y) = \prod_{i=1}^n P(X_i|Y)$$

- (We will discuss the validity of this assumption later)

Example

- Example: Play Tennis

PlayTennis: training examples

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

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

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

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

Wind	Play=Yes	Play=No
<i>Strong</i>	3/9	3/5

Learning Phase

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

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

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

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

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

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

Learning Phase

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

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

$\mathbf{x}' = (\text{Outlook}=\textit{Sunny},$
 $\text{Temperature}=\textit{Cool},$
 $\text{Humidity}=\textit{High}, \text{Wind}=\textit{Strong})$

Example

- Test Phase

- Given a new instance,

$\mathbf{x}' = (\text{Outlook}=\textit{Sunny}, \text{Temperature}=\textit{Cool}, \text{Humidity}=\textit{High}, \text{Wind}=\textit{Strong})$

- Look up tables

$$P(\text{Outlook}=\textit{Sunny} \mid \text{Play}=\textit{Yes}) = 2/9$$

$$P(\text{Temperature}=\textit{Cool} \mid \text{Play}=\textit{Yes}) = 3/9$$

$$P(\text{Humidity}=\textit{High} \mid \text{Play}=\textit{Yes}) = 3/9$$

$$P(\text{Wind}=\textit{Strong} \mid \text{Play}=\textit{Yes}) = 3/9$$

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

$$P(\text{Outlook}=\textit{Sunny} \mid \text{Play}=\textit{No}) = 3/5$$

$$P(\text{Temperature}=\textit{Cool} \mid \text{Play}=\textit{No}) = 1/5$$

$$P(\text{Humidity}=\textit{High} \mid \text{Play}=\textit{No}) = 4/5$$

$$P(\text{Wind}=\textit{Strong} \mid \text{Play}=\textit{No}) = 3/5$$

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

- MAP rule

$$P(\text{Yes} \mid \mathbf{x}'): [P(\textit{Sunny} \mid \textit{Yes})P(\textit{Cool} \mid \textit{Yes})P(\textit{High} \mid \textit{Yes})P(\textit{Strong} \mid \textit{Yes})]P(\text{Play}=\textit{Yes}) = 0.0053$$

$$P(\text{No} \mid \mathbf{x}'): [P(\textit{Sunny} \mid \textit{No})P(\textit{Cool} \mid \textit{No})P(\textit{High} \mid \textit{No})P(\textit{Strong} \mid \textit{No})]P(\text{Play}=\textit{No}) = 0.0206$$

Given the fact $P(\text{Yes} \mid \mathbf{x}') < P(\text{No} \mid \mathbf{x}')$, we label \mathbf{x}' to be “No”.

Relevant Issues

- Violation of Independence Assumption
 - For many real world tasks,
 - Nevertheless, naïve Bayes works surprisingly well anyway!
- Zero conditional probability Problem
 - If no example contains the attribute value
 - In this circumstance, during test
 - For a remedy, conditional probabilities estimated with **Laplace smoothing**:

Relevant Issues

- Continuous-valued Input Attributes
 - Numberless values for an attribute
 - Conditional probability modeled with the normal distribution
- Learning Phase:
Output: normal distributions and
- Test Phase:
 - Calculate conditional probabilities with all the normal distributions
 - Apply the MAP rule to make a decision