Naïve Bayes

A study material for the students of GLS University
Compiled by Dr. Krupa Mehta

Naïve Bayes

Weather Condition	Wins in last 3 matches	Humidity	Win toss	Won match?
Rainy	3 wins	High	FALSE	No
Rainy	3 wins	High	TRUE	No
OverCast	3 wins	High	FALSE	Yes
Sunny	2 wins	High	FALSE	Yes
Sunny	1 win	Normal	FALSE	Yes
Sunny	1 win	Normal	TRUE	No
OverCast	1 win	Normal	TRUE	Yes
Rainy	2 wins	High	FALSE	No
Rainy	1 win	Normal	FALSE	Yes
Sunny	2 wins	Normal	FALSE	Yes
Rainy	2 wins	Normal	TRUE	Yes
OverCast	2 wins	High	TRUE	Yes
OverCast	3 wins	Normal	FALSE	Yes
Sunny	2 wins	High	TRUE	No

Weather Conditions = Rainy (a1),
won two of the last three matches (a2),
Humidity = Normal (a3),
won the toss in the particular match (a4)

Naïve Bayes

- Step 1: Construct a frequency table

	Won Match			Won Match	
Weather condition	Yes	No	**Humidity**	Yes	No
Sunny	3	2	High	3	4
OverCast	4	0	Normal	6	1
Rainy	2	3			
Total	9	5	Total	9	5

	Won Match			Won Match	
Wins in last 3 matches	Yes	No	Win toss	Yes	No
3 wins	2	2	FALSE	6	2
1 win	4	2	TRUE	3	3
2 wins	3	1			
Total	9	5	Total	9	5

Naïve Bayes

- Step 2: Identify the cumulative probability
- P(Win match|a1 ∩ a2 ∩ a3 ∩ a4)

$$= \frac{P(a_1|\text{Win match})P(a_2|\text{Win match})P(a_3|\text{Win match})P(a_4|\text{Win match})P(\text{Win match})}{P(a_1)P(a_2)P(a_3)P(a_4)}$$

= 2/9 * 3/9 * 6/9 * 9/14

= 0.22 * 0.33 * 0.67 * 0.64

= **0.0311**

- P(!Win match|a1 ∩ a2 ∩ a3 ∩ a4)

$$= \frac{P(a_1|!\text{Win match})P(a_2|!\text{Win match})P(a_3|!\text{Win match})P(a_4|!\text{Win match})P(!\text{Win match})}{P(a_1)P(a_2)P(a_3)P(a_4)}$$

= 3/5 * 1/5 * 1/5 * 5/14

= 0.6 * 0.2 * 0.2 * 0.36

= **0.00864**

Naïve Bayes

- Step 3: Calculate probability through normalization by applying the formula

$$P(\text{Yes}) = \frac{P(\text{Yes})}{P(\text{Yes}) + P(\text{No})}$$

$$P(\text{No}) = \frac{P(\text{No})}{P(\text{Yes}) + P(\text{No})}$$

$$P(\text{Win match}) = \frac{P(\text{Win match})}{P(\text{Win match}) + P(\text{!Win match})}$$

= 0.0311 / (0.0311+0.00864)

= 0.0311 / 0.0397

= 0.78

$$P(\text{!Win match}) = \frac{P(\text{!Win match})}{P(\text{Win match}) + P(\text{!Win match})}$$

= 0.00864 / (0.0311+0.00864)

= 0.00864 / 0.0397

= 0.22

0.78 + 0.22 = 1.00

78% chances to win the match

Naïve Bayes

Strengths	Weakness
Simple and fast in calculation but yet effective in result	The basis assumption of equal importance and independence often does not hold true
In situations where there are noisy and missing data, it performs well	If the target dataset contains large numbers of numeric features, then the reliability of the outcome becomes limited
Works equally well when smaller number of data is present for training as well as very large number of training data is available	Though the predicted classes have a high reliability, estimated probabilities have relatively lower reliability
Easy and straightforward way to obtain the estimated probability of a prediction	

Naïve Bayes

- Applications of Naïve Bayes classifier
  - Text classification
  - Spam filtering
  - Hybrid Recommender System
  - Online Sentiment Analysis

Reference

- Machine Learning by Saikat Dutt, Subramanian Chandramouli, Amit Kumar Das published by Pearson