

Naive Bayes Classification

Review of Classification

When do we want to use
classification?

What is Naive Bayes?

Naive Bayes is a very popular and simple machine learning method used for classification.

It uses Bayes' theorem to predict the class of new data.

Review of Bayes Theorem

Posterior probability of class c given predictor x

Likelihood of predictor x given class c

Prior probability of class c

Prior probability of predictor x

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

The diagram illustrates the components of Bayes' Theorem. Three boxes at the top contain descriptive text: 'Posterior probability of class c given predictor x' (green border), 'Likelihood of predictor x given class c' (blue border), and 'Prior probability of class c' (orange border). A fourth box at the bottom, 'Prior probability of predictor x' (red border), points to the denominator of the formula. Arrows connect these boxes to the corresponding terms in the equation $P(c|x) = \frac{P(x|c)P(c)}{P(x)}$.

Probability Review

Marginal Probability

$$P(A) = \text{count}(A) / \text{total}$$

$$P(B) = \text{count}(B) / \text{total}$$

Joint Probability

$$P(A,B) = \text{count}(A,B) / \text{total}$$

$$P(B,A) = \text{count}(B,A) / \text{total}$$

Conditional Probability

$$P(A | B) = \text{count}(A,B) / \text{count}(B)$$

$$P(B | A) = \text{count}(B,A) / \text{count}(A)$$

Naive Bayes Process

1. Calculate the prior probability for each class
2. Calculate the conditional probability for each category given the class
3. Calculate the posterior probability using Bayes theorem
4. Predict a class based on which class has the higher probability conditional on the predictors.

Naive Bayes Example

Let's say we want to predict whether or not we are going to play tennis with our friend based on the weather. Below is some data we have collected in the past

Weather	Play
Rainy	No
Sunny	Yes
Overcast	Yes
Overcast	No
Sunny	Yes
Rainy	No
Rainy	No
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rainy	No
Rainy	Yes

Today it is overcast.
Will we play tennis?

Naive Bayes Example

Let's say we want to predict whether or not we are going to play tennis with our friend based on the weather. Below is some data we have collected in the past

Weather	Play
Rainy	No
Sunny	Yes
Overcast	Yes
Overcast	No
Sunny	Yes
Rainy	No
Rainy	No
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rainy	No
Rainy	Yes

Today it is overcast.
Will we play tennis?

$$P(\text{Yes} \mid \text{Overcast}) = ??$$

$$P(\text{No} \mid \text{Overcast}) = ??$$

Naive Bayes Example Cont.

Weather	Play
Rainy	No
Sunny	Yes
Overcast	Yes
Overcast	No
Sunny	Yes
Rainy	No
Rainy	No
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rainy	No
Rainy	Yes

$n = 12$

1. Calculate the prior probability of each class.

Weather

$$P(\text{Rainy}) =$$

$$P(\text{Overcast})$$

$$=$$

$$P(\text{Sunny}) =$$

Play

$$P(\text{No}) =$$

$$P(\text{Yes}) =$$

2. Calculate the conditional probability for each category given the class

$$P(\text{Rainy} | \text{No}) =$$

$$P(\text{Overcast} | \text{No})$$

$$=$$

$$P(\text{Sunny} | \text{No}) =$$

$$P(\text{Rainy} | \text{Yes}) =$$

$$P(\text{Overcast} | \text{Yes})$$

$$=$$

$$P(\text{Sunny} | \text{Yes}) =$$

Naive Bayes Example Cont.

Weather	Play
Rainy	No
Sunny	Yes
Overcast	Yes
Overcast	No
Sunny	Yes
Rainy	No
Rainy	No
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rainy	No
Rainy	Yes

$n = 12$

1. Calculate the prior probability of each class.

Weather

$$P(\text{Rainy}) = 5/12 = 0.42$$

$$P(\text{Overcast}) = 4/12 = 0.33$$

$$= 3/12 = 0.25$$

$$P(\text{Sunny}) =$$

Play

$$P(\text{No}) = 5/12 = 0.42$$

$$P(\text{Yes}) = 7/12 = 0.58$$

2. Calculate the conditional probability for each category given the class

$$P(\text{Rainy} | \text{No}) = 4/5 =$$

$$P(\text{Overcast} | \text{No}) = 0.80$$

$$= 1/5 =$$

$$P(\text{Sunny} | \text{No}) = 0.20$$

$$0/5 =$$

$$P(\text{Rainy} | \text{Yes}) = 0.00$$

$$P(\text{Overcast} | \text{Yes})$$

$$= 1/7 =$$

$$P(\text{Sunny} | \text{Yes}) = 0.14$$

Naive Bayes Example Cont.

3. Calculate the posterior probability using Bayes Theorem

$$\begin{aligned}P(No|Overcast) &= P(Overcast|No)P(No)/P(Overcast) \\&= 0.2 * 0.42/0.33 \\&= 0.25\end{aligned}$$

$$\begin{aligned}P(Yes|Overcast) &= P(Overcast|Yes)P(Yes)/P(Overcast) \\&= 0.43 * 0.58/0.33 \\&= 0.76\end{aligned}$$

4. Predict a class based on the probability conditional on the predictors

Will we play tennis if
it is overcast?