Naive Bayes Classification

Review of Classification

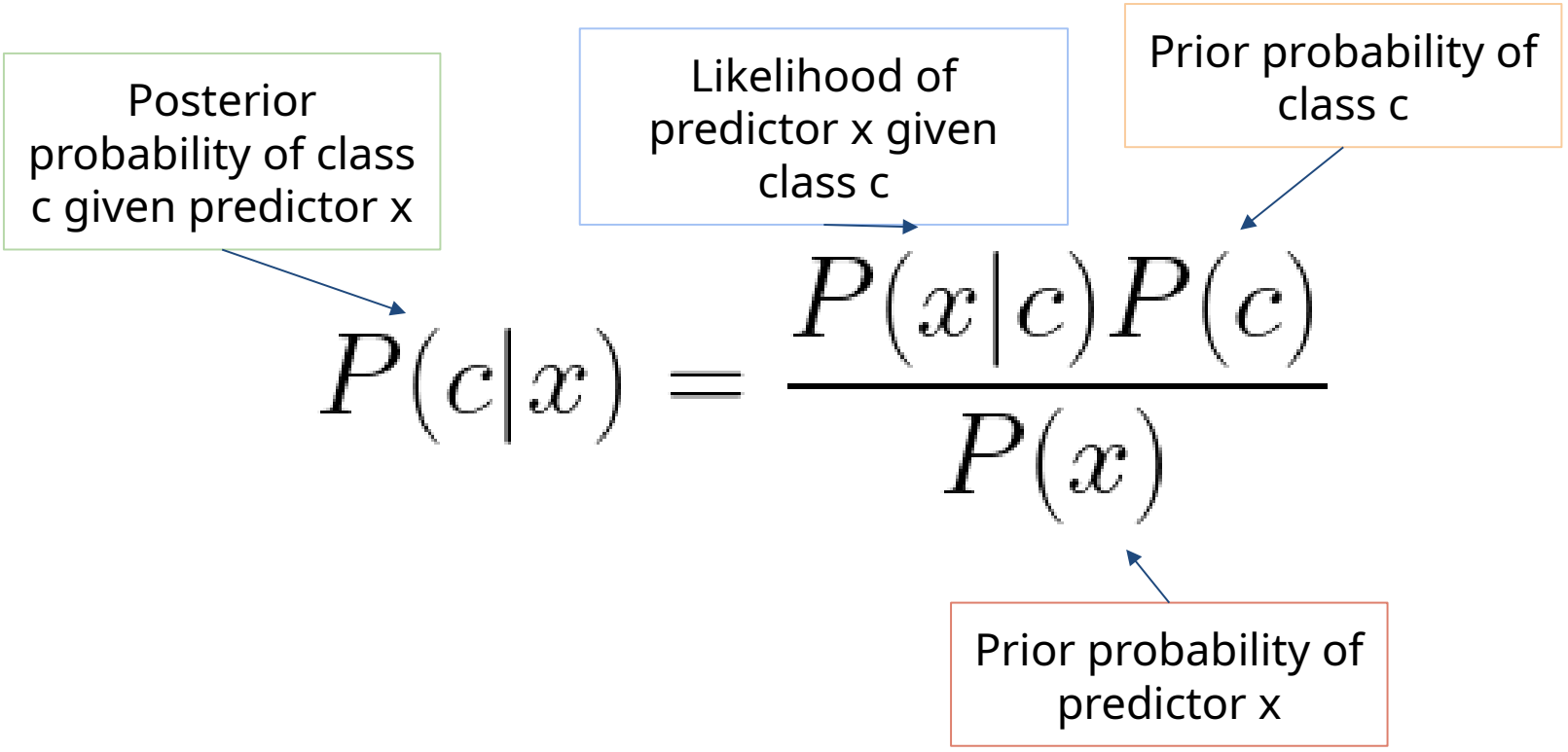
When do we want to use classification?

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



Probability Review

Marginal Probability

P(A) = count(A) / total

P(B) = count(B) / total

Joint Probability

P(A,B) = count(A,B) / total

P(B,A) = count(B,A) / total

Conditional Probability

P(A|B) = count(A,B) / count(B)

P(B|A) = count(B,A) / count(A)

Chapter 1 in Think Bayes

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?

$P(Yes \mid Overcast) = ??$
$P(No \mid Overcast) = ??$

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

$P(Rainy) =$

$P(Overcast) =$

$P(Sunny) =$

Play

$P(No) =$

$P(Yes) =$

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

$P(Rainy|No) =$

$P(Overcast|No) =$

$P(Sunny|No) =$

$P(Rainy|Yes) =$

$P(Overcast|Yes) =$

$P(Sunny|Yes) =$

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

$P(Rainy) = 5/12 = 0.42$

$P(Overcast) = 4/12 = 0.33$

$P(Sunny) = 3/12 = 0.25$

Play

$P(No) = 5/12 = 0.42$

$P(Yes) = 7/12 = 0.58$

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

$P(Rainy|No) = 4/5 = 0.80$

$P(Overcast|No) = 1/5 = 0.20$

$P(Sunny|No) = 0/5 = 0.00$

$P(Rainy|Yes) = 1/7 = 0.14$

$P(Overcast|Yes) = 3/7 = 0.43$

$P(Sunny|Yes) = 3/7 = 0.43$

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?