Naive Bayes Classifier

By: Trevor Hamilton & Jonathan Stonebraker

Overview

Origin

Introduction

Common Use Cases, Pros & Cons

Naive Bayes Classifier

Naive Bayes Classifier Example

Laplace Smoothing

Laplace Smoothing Example

Origin

- Based off of Bayes Theorem named after Thomas Bayes
- Came about in 1763 in *An Essay towards solving a Problem in the Doctrine of Chances*
- Focused on the theory of probability
- Use in data mining dates back to the beginning of the field

$$P(A|B) = \frac{P(A)\,P(B|A)}{P(B)},$$

Introduction

- Naive Bayes classifiers family of classifiers based on the use of Bayes theorem. AKA Simple Bayes or Independence Bayes
- Is one of the best classifiers while attributes have strong independence

Class

$$p(C_k|\mathbf{x}) = \frac{p(C_k)\ p(\mathbf{x}|C_k)}{p(\mathbf{x})}.$$

Instance

Use Cases

- Spam filters on email
- Sentence analysis (Subject, Connotation)
- Medical Diagnosis
- Credit approval
- Target Markets and Ads

Pros

- Requires smaller training data size than many other classifiers
- Great for discrete data
- Theoretically one of the best classifiers
- Works well in sentence recognition since all sentences do not have to be of the same length

Cons

- Attributes must be independent*
- Continuous values do not work as well, use Gaussian Bayes Classifier
- If a new value is encountered during testing that was not included in the training set likelihood will move to zero known as “ Zero Frequency”

Naive Bayes - The Data

- Needs less training data than other classifiers
- Attributes must be as independent as possible
    - In real world data as long as there is no strong correlations it still holds up very well
- Is unaware of data type
- More attributes the better

Naive Bayes - Generate Model

- Count the number of occurrences of each distinct value of each attribute for each classification
- Divide that by the current classification count
- The result is a chart where you have the likelihood of each distinct value of each attribute for each classification

Naive Bayes - Generate Classification

- For each class for each attribute select the occurrence value that matches your input attribute
- Get its likelihood for the current classification

Naive Bayes Classifier Example

Outlook	Temp	Humidty	Windy	Play
Rainy	Hot	High	FALSE	No
Rainy	Hot	High	TRUE	No
Overcast	Hot	High	FALSE	Yes
Sunny	Mild	High	FALSE	Yes
Sunny	Cool	Normal	FALSE	Yes
Sunny	Cool	Normal	TRUE	No
Overcast	Cool	Normal	TRUE	Yes
Rainy	Mild	High	FALSE	No
Rainy	Cool	Normal	FALSE	Yes
Sunny	Mild	Normal	FALSE	Yes
Rainy	Mild	Normal	TRUE	Yes
Overcast	Hot	High	TRUE	Yes

Attribute Value	No - 4	Yes - 8
Rainy	3	2
Overcast	0	3
Sunny	1	3
Hot	2	2
Mild	1	4
Cool	1	3
High	3	3
Normal	1	5
False	2	5
True	2	3

Attribute Value	No - 4/12	Yes - 8/12
Rainy	3/4	2/8
Overcast	0/4	3/8
Sunny	1/4	3/8
Hot	2/4	2/8
Mild	1/4	4/8
Cool	1/4	3/8
High	3/4	3/8
Normal	1/4	5/8
False	2/4	5/8
True	2/4	3/8

Attribute Value	No - 4/12	Yes - 8/12
Rainy	3/4	2/8
Overcast	0/4	3/8
Sunny	1/4	3/8
Hot	2/4	2/8
Mild	1/4	4/8
Cool	1/4	3/8
High	3/4	3/8
Normal	1/4	5/8
False	2/4	5/8
True	2/4	3/8

Overcast	Mild	Normal	FALSE	?

Attribute Value	No - 4/12	Yes - 8/12
Rainy	3/4	2/8
Overcast	0/4	3/8
Sunny	1/4	3/8
Hot	2/4	2/8
Mild	1/4	4/8
Cool	1/4	3/8
High	3/4	3/8
Normal	1/4	5/8
False	2/4	5/8
True	2/4	3/8

Overcast	Mild	Normal	FALSE	?
0/4	1/4	1/4	2/4	4/12

Attribute Value	No - 4/12	Yes - 8/12
Rainy	3/4	2/8
Overcast	0/4	3/8
Sunny	1/4	3/8
Hot	2/4	2/8
Mild	1/4	4/8
Cool	1/4	3/8
High	3/4	3/8
Normal	1/4	5/8
False	2/4	5/8
True	2/4	3/8

Overcast	Mild	Normal	FALSE	?
0/4	2/4	1/4	2/4	4/12
3/8	4/8	5/8	5/8	8/12

Attribute Value	No - 4/12	Yes - 8/12
Rainy	3/4	2/8
Overcast	0/4	3/8
Sunny	1/4	3/8
Hot	2/4	2/8
Mild	1/4	4/8
Cool	1/4	3/8
High	3/4	3/8
Normal	1/4	5/8
False	2/4	5/8
True	2/4	3/8

Overcast	Mild	Normal	FALSE	?
				0
				$\boxed{0.0488}$

Attribute Value	No - 4/12	Yes - 8/12
Rainy	3/4	2/8
Overcast	0/4	3/8
Sunny	1/4	3/8
Hot	2/4	2/8
Mild	1/4	4/8
Cool	1/4	3/8
High	3/4	3/8
Normal	1/4	5/8
False	2/4	5/8
True	2/4	3/8

Overcast	Mild	Normal	FALSE	Yes
				0
				0.0488

Laplace Smoothing

- Technique used to smooth categorical data
- Commonly used in Naive Bayes Classifiers
    - Helps to deal with values not in the training set

$$\hat{\theta}_i = \frac{x_i + \alpha}{N + \alpha d} \qquad (i = 1, \ldots, d),$$

Laplace Smoothing Example

Attribute Value	No - 4/12	Yes - 8/12
Rainy	3/4	2/8
Overcast	0/4	3/8
Sunny	1/4	3/8
Hot	2/4	2/8
Mild	1/4	4/8
Cool	1/4	3/8
High	3/4	3/8
Normal	1/4	5/8
False	2/4	5/8
True	2/4	3/8

$$\frac{\text{Numerator} + K}{\text{Denominator} + \text{ParentClass.possibleValues} * K}$$

Attribute Value	No - (k+4)/(12+2)	Yes - (k+8)/(12+2)
Rainy	(k+3))/(4+3*k)	(k+2)/(8+3*k)
Overcast	(k+0)/(4+3*k)	(k+3)/(8+3*k)
Sunny	(k+1)/(4+3*k)	(k+3)/(8+3*k)
Hot	(k+2)/(4+3*k)	(k+2)/(8+3*k)
Mild	(k+1)/(4+3*k)	(k+4)/(8+3*k)
Cool	(k+1)/(4+3*k)	(k+3)/(8+3*k)
High	(k+3)/(4+2*k)	(k+3)/(8+2*k)
Normal	(k+1)/(4+2*k)	(k+5)/(8+2*k)
False	(k+2)/(4+2*k)	(k+5)/(8+2*k)
True	(k+2)/(4+2*k)	(k+3)/(8+2*k)

Attribute Value	No - (k+4)/(12+2)	Yes - (k+8)/(12+2)
Rainy	(k+3))/(4+3*k)	(k+2)/(8+3*k)
Overcast	(k+0)/(4+3*k)	(k+3)/(8+3*k)
Sunny	(k+1)/(4+3*k)	(k+3)/(8+3*k)
Hot	(k+2)/(4+3*k)	(k+2)/(8+3*k)
Mild	(k+1)/(4+3*k)	(k+4)/(8+3*k)
Cool	(k+1)/(4+3*k)	(k+3)/(8+3*k)
High	(k+3)/(4+2*k)	(k+3)/(8+2*k)
Normal	(k+1)/(4+2*k)	(k+5)/(8+2*k)
False	(k+2)/(4+2*k)	(k+5)/(8+2*k)
True	(k+2)/(4+2*k)	(k+3)/(8+2*k)

Overcast	Mild	None	TRUE	Yes

Attribute Value	No - (k+4)/(12+2)	Yes - (k+8)/(12+2)
Rainy	(k+3))/(4+3*k)	(k+2)/(8+3*k)
Overcast	(k+0)/(4+3*k)	(k+3)/(8+3*k)
Sunny	(k+1)/(4+3*k)	(k+3)/(8+3*k)
Hot	(k+2)/(4+3*k)	(k+2)/(8+3*k)
Mild	(k+1)/(4+3*k)	(k+4)/(8+3*k)
Cool	(k+1)/(4+3*k)	(k+3)/(8+3*k)
High	(k+3)/(4+2*k)	(k+3)/(8+2*k)
Normal	(k+1)/(4+2*k)	(k+5)/(8+2*k)
False	(k+2)/(4+2*k)	(k+5)/(8+2*k)
True	(k+2)/(4+2*k)	(k+3)/(8+2*k)

Overcast	Mild	None	TRUE	?
1/7	2/7	1/6	3/6	5/14

Attribute Value	No - (k+4)/(12+2)	Yes - (k+8)/(12+2)
Rainy	(k+3))/(4+3*k)	(k+2)/(8+3*k)
Overcast	(k+0)/(4+3*k)	(k+3)/(8+3*k)
Sunny	(k+1)/(4+3*k)	(k+3)/(8+3*k)
Hot	(k+2)/(4+3*k)	(k+2)/(8+3*k)
Mild	(k+1)/(4+3*k)	(k+4)/(8+3*k)
Cool	(k+1)/(4+3*k)	(k+3)/(8+3*k)
High	(k+3)/(4+2*k)	(k+3)/(8+2*k)
Normal	(k+1)/(4+2*k)	(k+5)/(8+2*k)
False	(k+2)/(4+2*k)	(k+5)/(8+2*k)
True	(k+2)/(4+2*k)	(k+3)/(8+2*k)

Overcast	Mild	None	TRUE	?
1/7	2/7	1/6	3/6	5/14
4/11	5/11	1/10	4/10	9/14

Attribute Value	No - (k+4)/(12+2)	Yes - (k+8)/(12+2)
Rainy	(k+3))/(4+3*k)	(k+2)/(8+3*k)
Overcast	(k+0)/(4+3*k)	(k+3)/(8+3*k)
Sunny	(k+1)/(4+3*k)	(k+3)/(8+3*k)
Hot	(k+2)/(4+3*k)	(k+2)/(8+3*k)
Mild	(k+1)/(4+3*k)	(k+4)/(8+3*k)
Cool	(k+1)/(4+3*k)	(k+3)/(8+3*k)
High	(k+3)/(4+2*k)	(k+3)/(8+2*k)
Normal	(k+1)/(4+2*k)	(k+5)/(8+2*k)
False	(k+2)/(4+2*k)	(k+5)/(8+2*k)
True	(k+2)/(4+2*k)	(k+3)/(8+2*k)

Overcast	Mild	None	TRUE	?
				0.0012
				$\boxed{0.0425}$

Live Demo
& Results

Questions?
Code Samples?

Contact Info: hamilton.tj@outlook.com stonebrakerjon@gmail.com

Source Repo:

https://github.com/hamiltontj/NaiveBayes