Naïve Bayes classifier

Data Science Dojo

datasciencedojo
unleash the data scientist in you

Naïve Bayes Classifier

This is a simple method that is sometimes quite effective.

- Key concepts to understand are:
  - Conditional probability
  - Bayes theorem
  - Conditional independence

datasciencedojo
unleash the data scientist in you

Conditional Probability – Shooting a Basketball

- The conditional probability of event A given event B is written as $P(A|B)$
- This is the probability of event A if/when/given event B happens

	far	close	total
make	5	3	8
miss	10	2	12
total	15	5	20

- $P(make) = 8/20=0.4$ $P(make|close)=3/5=0.6$

datasciencedojo
unleash the data scientist in you

Conditional Probability – Shooting a Basketball

- You can define the conditional probability $P(A|B)$ as $P(A,B)/P(B)$

Example:

	far	close	total
make	5	3	8
miss	10	2	12
total	15	5	20

$$P(make|close) = P(make,close) / P(close) = (3/20) / (5/20)$$
$$= 0.15/0.25 = 0.6$$

- Bayes Theorem turns this around to solve for $P(B|A)$ if you have $P(A|B)$

$$P(B|A) = P(A,B) / P(A) = P(A|B)\ P(B) / P(A)$$

datasciencedojo
unleash the data scientist in you

Independence

- A and B are independent if $P(A,B) = P(A)*P(B)$

- Here the events are not independent:

  $P(make,far) = 5/20=0.25$

  but $P(make)*P(far) = 8/20*15/20= 0.30$

	far	close	total
make	5	3	8
miss	10	2	12
total	15	5	20

datasciencedojo
unleash the data scientist in you

Independence

- Here the events are independent:

  $P(make,far) = 9/20=0.45$

  which equals

  $P(make)*P(far) = 12/20*15/20=0.45$

datasciencedojo
unleash the data scientist in you

Conditional Independence

- A and B are conditionally independent given C *iff*

$$P(A,B|C) = P(A|C)*P(B|C)$$

- Question:
  - Are height and reading ability independent?
  - What if we take age into account?

datasciencedojo
unleash the data scientist in you

Conditional Independence

- A and B are conditionally independent given C *iff*

$$P(A,B|C) = P(A|C)*P(B|C)$$

- Example: Height and reading ability are not independent but they are conditionally independent given the age level

	all		
	short	tall	total
reads poorly	92	29	121
reads well	18	81	99
total	110	110	220

	young		
	short	tall	total
reads poorly	90	9	99
reads well	10	1	11
total	100	10	110

	old		
	short	tall	total
reads poorly	2	20	22
reads well	8	80	88
total	10	100	110

datasciencedojo
unleash the data scientist in you

Bayes Rule

- Conditional Probability:

$$P(C \mid A) = \frac{P(A,C)}{P(A)}$$

$$P(A \mid C) = \frac{P(A,C)}{P(C)}$$

- Bayes Theorem:

$$P(C \mid A) = \frac{P(A \mid C)P(C)}{P(A)}$$

datasciencedojo
unleash the data scientist in you

Example of Bayes Rule

- Given:
  - A doctor knows that meningitis causes stiff neck 50% of the time
  - Prior probability of any patient having meningitis is 1/50,000
  - Prior probability of any patient having stiff neck is 1/20

- If a patient has stiff neck, what's the probability he/she has meningitis?

$$P(M \mid S) = \frac{P(S \mid M)P(M)}{P(S)} = \frac{0.5 \times 1/50000}{1/20} = 0.0002$$

datasciencedojo
unleash the data scientist in you

Bayesian Classifiers

- Consider each attribute and class label as random variables
- Given a record with attributes $(A_1, A_2, \ldots, A_n)$
  - Goal is to predict class C
  - Specifically, we want to find the value of C that maximizes $P(C| A_1, A_2, \ldots, A_n)$
- Can we estimate $P(C| A_1, A_2, \ldots, A_n)$ directly from data?

datasciencedojo
unleash the data scientist in you

Bayesian Classifiers

- Approach:
  - Compute the posterior probability $P(C \mid A_1, A_2, \ldots, A_n)$ for all values of C using the Bayes theorem

$$\boxed{P(C \mid A_1 A_2 \ldots A_n) = \frac{P(A_1 A_2 \ldots A_n \mid C) P(C)}{P(A_1 A_2 \ldots A_n)}}$$

  - Choose value of C that maximizes $P(C \mid A_1, A_2, \ldots, A_n)$
  - Equivalent to choosing value of C that maximizes $P(A_1, A_2, \ldots, A_n|C) P(C)$
- How to estimate $P(A_1, A_2, \ldots, A_n \mid C)$?

datasciencedojo
unleash the data scientist in you

Naïve Bayes Classifier

- Assume independence among attributes A_i when class is given: (in other words, assume conditional independence between attributes)
- $P(A_1, A_2, \ldots, A_n | C) = P(A_1| C_j) P(A_2| C_j) \ldots P(A_n| C_j)$

Can estimate $P(A_i| C_j)$ for all A_i and C_j

New point is classified to C_j if $P(C_j) \Pi P(A_i| C_j)$ is maximal

datasciencedojo
unleash the data scientist in you

How to Estimate Probabilities from Data?

- Class: $P(C) = N_c/N$
  - e.g., $P(No) = 7/10$,
    $P(Yes) = 3/10$
- For discrete attributes:
- $P(A_i | C_k) = |A_{ik}|/ N_c$

where $|A_{ik}|$ is number of instances having attribute A_i and belongs to class C_k

Examples:

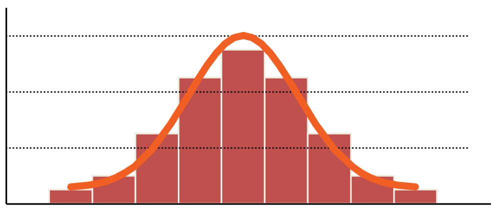
$P(Status=Married|No) = 4/7$
$P(Refund=Yes|Yes)=0$

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

datasciencedojo
unleash the data scientist in you

How to Estimate Probabilities from Data?

- For continuous attributes:
  - Assume attribute follows a normal distribution
  - Use data to estimate parameters of distribution (e.g., mean and standard deviation)
  - Once probability distribution is known, can use it to estimate the conditional probability $P(A_i|c)$



datasciencedojo
unleash the data scientist in you

How to Estimate Probabilities from Data?

- Normal distribution:

$$P(A_i \mid c_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(A_i-\mu_{ij})^2}{2\sigma_{ij}^2}}$$

  - One for each (A_i,c_i) pair

- For (Income, Class=No):
  - If Class=No
    - Sample mean = 110
    - Sample variance = 2975

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

$$P(Income = 120 \mid No) = \frac{1}{\sqrt{2\pi}(54.54)} e^{-\frac{(120-110)^2}{2(2975)}} = 0.0072$$

datasciencedojo
unleash the data scientist in you

Example of Naïve Bayes Classifier

Given a Test Record:

naive Bayes Classifier:

P(Refund=Yes|No) = 3/7
P(Refund=No|No) = 4/7
P(Refund=Yes|Yes) = 0
P(Refund=No|Yes) = 1
P(Marital Status=Single|No) = 2/7
P(Marital Status=Divorced|No)=1/7
P(Marital Status=Married|No) = 4/7
P(Marital Status=Single|Yes) = 2/7
P(Marital Status=Divorced|Yes)=1/7
P(Marital Status=Married|Yes) = 0

For taxable income:
If class=No: sample mean=110
sample variance=2975
If class=Yes: sample mean=90
sample variance=25

$$X = (\text{Refund} = \text{No}, \text{Married}, \text{Income} = 120\text{K})$$

- $P(X|Class=No) = P(Refund=No|Class=No) \times P(Married|\ Class=No) \times P(Income=120K|\ Class=No)$
  $= 4/7 \times 4/7 \times 0.0072 = 0.0024$

- $P(X|Class=Yes) = P(Refund=No|\ Class=Yes) \times P(Married|\ Class=Yes) \times P(Income=120K|\ Class=Yes)$
  $= 1 \times 0 \times 1.2 \times 10^{-9} = 0$

Since $P(X|No)P(No) > P(X|Yes)P(Yes)$

Therefore $P(No|X) > P(Yes|X)$

=> Class = No

datasciencedojo
unleash the data scientist in you

Naïve Bayes Classifier

- If one of the conditional probabilities is zero, then the entire expression becomes zero.
- Probability estimation:

$$\text{Original}: P(A_i \mid C) = \frac{N_{ic}}{N_c}$$

$$\text{Laplace}: P(A_i \mid C) = \frac{N_{ic} + 1}{N_c + c}$$

datasciencedojo
unleash the data scientist in you

Example of Naïve Bayes Classifier

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

A: attributes M: mammals N: non-mammals

$$P(A\,|\,M) = \frac{6}{7} \times \frac{6}{7} \times \frac{2}{7} \times \frac{2}{7} = 0.06$$

$$P(A\,|\,N) = \frac{1}{13} \times \frac{10}{13} \times \frac{3}{13} \times \frac{4}{13} = 0.0042$$

$$P(A\,|\,M)P(M) = 0.06 \times \frac{7}{20} = 0.021$$

$$P(A\,|\,N)P(N) = 0.004 \times \frac{13}{20} = 0.0027$$

$$P(A|M)P(M) > P(A|N)P(N)$$

=> Mammals

datasciencedojo
unleash the data scientist in you

Naïve Bayes (Summary)

- Robust to isolated noise points and any irrelevant attributes
- Handle missing values by ignoring the instance during probability estimate calculations
- Shown to work well on text classification related problems
- Independence assumption may not hold for some attributes
  - Use other techniques such as Bayesian Belief Networks (BBN)

datasciencedojo
unleash the data scientist in you