Naïve Bayes

Data Science Dojo

Agenda

- Probability Review
  - Conditional Probability
  - Bayes Theorem
  - Conditional Independence
- Naïve Bayes Classifier

Naïve Bayes Classifier

This is a computationally efficient method that is sometimes very effective.

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

CONDITIONAL PROBABILITY

Conditional Probability

- $P(A|B)$: the conditional probability of event A “given” event B
- i.e. the probability of event A occurring assuming event B has happened/will happen

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

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

Conditional Probability

- Definition: $P(A|B) = P(A \,\&\, B) / P(B)$

Example:

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

$$P(make|close) = P(make \,\&\, close) / P(close) = (3/20) / (5/20)$$

$$= 0.15/0.25 = 0.6$$

Note: This means $P(A|B)*P(B) = P(B|A)*P(A)$

BAYES THEOREM

Bayes Rule

- Conditional Probability:

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

$$P(A|C) = \frac{P(A \& C)}{P(C)}$$

- Bayes Theorem:

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

Example of Bayes Rule

- Givens
  - 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?

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

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
data science for everyone

Independence

- Here the events **are** independent:

$$P(make \ \& \ far) = 9/20 = 0.45$$

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

	far	close	total
make	9	3	12
miss	6	2	8
total	15	5	20

CONDITIONAL INDEPENDENCE

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?

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

NAÏVE BAYES CLASSIFIER

Bayesian Classifiers

- Consider each attribute and class label as random variables
- Given a record with attributes $\{A_1, A_2, \ldots, A_n\}$
  - Want 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?

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

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

  - Need value of C with maximum $P(C \mid A_1, A_2, \ldots, A_n)$
  - Equivalent to choosing value of C that maximizes

$$P(A_1, A_2, \ldots, A_n|C)^* P(C)$$

- How to estimate $P(A_1, A_2, \ldots, A_n \mid C)$?

Naïve Bayes Classifier

- Assume conditional independence among attributes A_i with respect to class:
- $P(A_1, A_2, \ldots, A_n |C) = P(A_1| C_j)\, P(A_2| C_j) \ldots P(A_n| C_j)$

- Estimate $P(A_i| C_j)$ for all A_i and C_j
- For each new record $\{A_1, A_2, \ldots, A_n\}$
  - Calculate $P(C_j |A_1, A_2, \ldots, A_n)$ for each class C_j
  - Assign the class with the largest conditional probability

How to Estimate Probabilities from Data?

- Class: $P(C) = N_c/N$
  - e.g., $P(No) = 6/10$,
    $P(Yes) = 4/10$
- For discrete attributes:
- $P(A_i \mid C_k) = |A_{ik}|/ N_c$

where $|A_{ik}|$ is number of instances which have attribute A_i and belong to class C_k

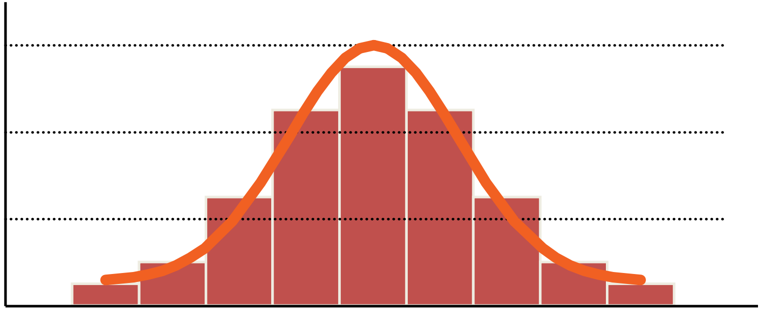
Examples:

$$P(Sex{=}Female|No) = 0$$
$$P(Pclass{=}1|Yes){=}\ 2/4$$

Pid	**Sex**	**Age**	**Pclass**	**Survived**
2	Female	38	1	Yes
3	Female	26	3	Yes
5	Male	35	3	No
7	Male	54	1	No
13	Male	20	3	No
14	Male	39	3	No
21	Male	35	2	No
24	Male	28	1	Yes
34	Male	66	1	No
54	Female	29	2	Yes

How to Estimate Probabilities from Data?

- 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
data science for everyone

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)
  - Sample mean = 37
  - Sample variance = 189

$$P(Age = 29 \mid No) = \frac{1}{\sqrt{2\pi(262)}} e^{-\frac{(29-42)^2}{2(262)}} = 0.0179$$

Pid	**Sex**	**Age**	**Pclass**	**Survived**
2	Female	38	1	Yes
3	Female	26	3	Yes
5	Male	35	3	No
7	Male	54	1	No
13	Male	20	3	No
14	Male	39	3	No
21	Male	35	2	No
24	Male	28	1	Yes
34	Male	66	1	No
54	Female	29	2	Yes

Example of Naïve Bayes Classifier

Test Record: $X = (\text{Sex} = \text{Male}, \text{Age} = 32, \text{Pclass} = 2)$

P(Sex=Male|No) = 6/7
P(Sex=Female|No) = 0
P(Sex=Male|Yes) = 1/7
P(Sex=Female|Yes) = 1
P(Pclass=1|No) = 2/4
P(Pclass=2|No) = 1/2
P(Pclass=3|No) = 1/4
P(Pclass=1|Yes) = 2/4
P(Pclass=2|Yes) = 1/2
P(Pclass=3|Yes) = 3/4
Mean(Age|No) = 41.5
Var(Age|No) = 262
Mean(Age|Yes) = 28
Var(Age|Yes) = 1.6

- $P(X|Class=No) = P(Sex=Male|No) \times P(Pclass=2|No) \times P(Age=32|No)$
  $= 6/7 \times 1/2 \times 0.0204 = 0.0128$

$P(X|No)P(No) = 0.0128 \times 6/10 = 0.00768$

- $P(X|Class=Yes) = P(Sex=Male|Yes) \times P(Pclass=2|Yes) \times P(Age=32|Yes)$
  $= 1/7 \times 1/2 \times 0.0021 = 0.00015$

$P(X|Yes)P(Yes) = 0.00015 \times 4/10 = 6 \times 10^{-5}$

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

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

=> Class = No

datasciencedojo
data science for everyone

Naïve Bayes Classifier

- If one of the conditional probabilities is zero, then the entire expression becomes zero.
- Apply probability correction

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

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

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)