Department of AI&DS

MACHINE LEARNING
22AD2203R

Topic:

PROBABILITY MODELS

Session - 05

Dr. NARENDRA BABU TATINI
Associate Professor
Department of IoT

INTRODUCTION

Basic Terminology

- Prior probability
- Conditional probability
- Posterior probability
- Bayes theorem

PRIOR PROBABILITY

The **prior probability** of an event is its probability calculated from some prior information about the event.

- ex: the probabilities of Apples in a bag

  P(Apples) is 40 %

- **Prior information:**
- No of customers =100
- 20% of customers buys computer.
- 50% customers are age is >35.

So P(buys computer)=0.2

P(age>35)=0.5

CONDITIONAL PROBABILITY

Conditional probabilities expresses the probability that Event-A will occur when you assume (or know) that Event-B has already occurred.

Notation: P(A|B)

Posterior probability:(Inference)

It is a Conditional probability which represents the updated prior probability after taking into account some new piece of information.-P(A|B)

Bayes Theorem

- **P(h)** to denote the initial probability that hypothesis h holds, before we have observed the training data.
- **P(h)** is often called the prior probability of h.
- **P(D)** to denote the prior probability that training data D .The probability of D given no knowledge about which hypothesis holds
- **P(D|h)** to denote the probability of observing data D given some world in which hypothesis h holds.

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

Bayes Theorem (Contd…)

- In machine learning problems we are interested in the probability ***P (h| D)*** that ***h*** holds given the observed training data ***D.***
- ***P (h| D)*** is called the ***posterior probability*** of ***h,*** as it reflects our confidence that ***h*** holds after we have seen the training data ***D.***
- The posterior probability ***P(h|D)*** reflects the influence of the training data ***D,*** in contrast to the prior probability ***P(h)*** , which is independent of D.

Bayes Theorem (Contd…)

Bayesian reasoning is

- Applicable for decision making
- Used for Inferential statistics that deals with probability inference.
- Knowledge of prior events will be used to predict future events.
- Ex. Predicting the color of marbles in a basket

Bayes Theorem (Contd…)

Does patient have cancer or not?

A patient takes a lab test and the result comes back positive. The test returns a correct positive result in only 98% of the cases in which the disease is actually present, and a correct negative result in only 97% of the cases in which the disease is not present. Furthermore, 0.008 of the entire populatio

$$
\begin{aligned}
P(cancer) &= \\
P(\neg cancer) &= \\
P(+|cancer) &= \\
P(-|cancer) &= \\
P(+|\neg cancer) &= \\
P(-|\neg cancer) &= \\
P(cancer|+) &=
\end{aligned}
$$

Choosing Hypotheses

Find most probable hypothesis given training data

Maximum a posteriori hypothesis h_{MAP}:

$$\begin{aligned} h_{MAP} &= \arg\max_{h \in H} P(h|D) \\ &= \arg\max_{h \in H} \frac{P(D|h)P(h)}{P(D)} \\ &= \arg\max_{h \in H} P(D|h)P(h) \end{aligned}$$

Assuming $P(h_i) = P(h_j)$ we can further simplify, and choose the *Maximum likelihood* (ML) hypothesis

$$h_{ML} = \arg\max_{h_i \in H} P(D|h_i)$$

Naive Bayes Classifier

Along with decision trees, neural networks, nearest nbr, one of the most practical learning methods.

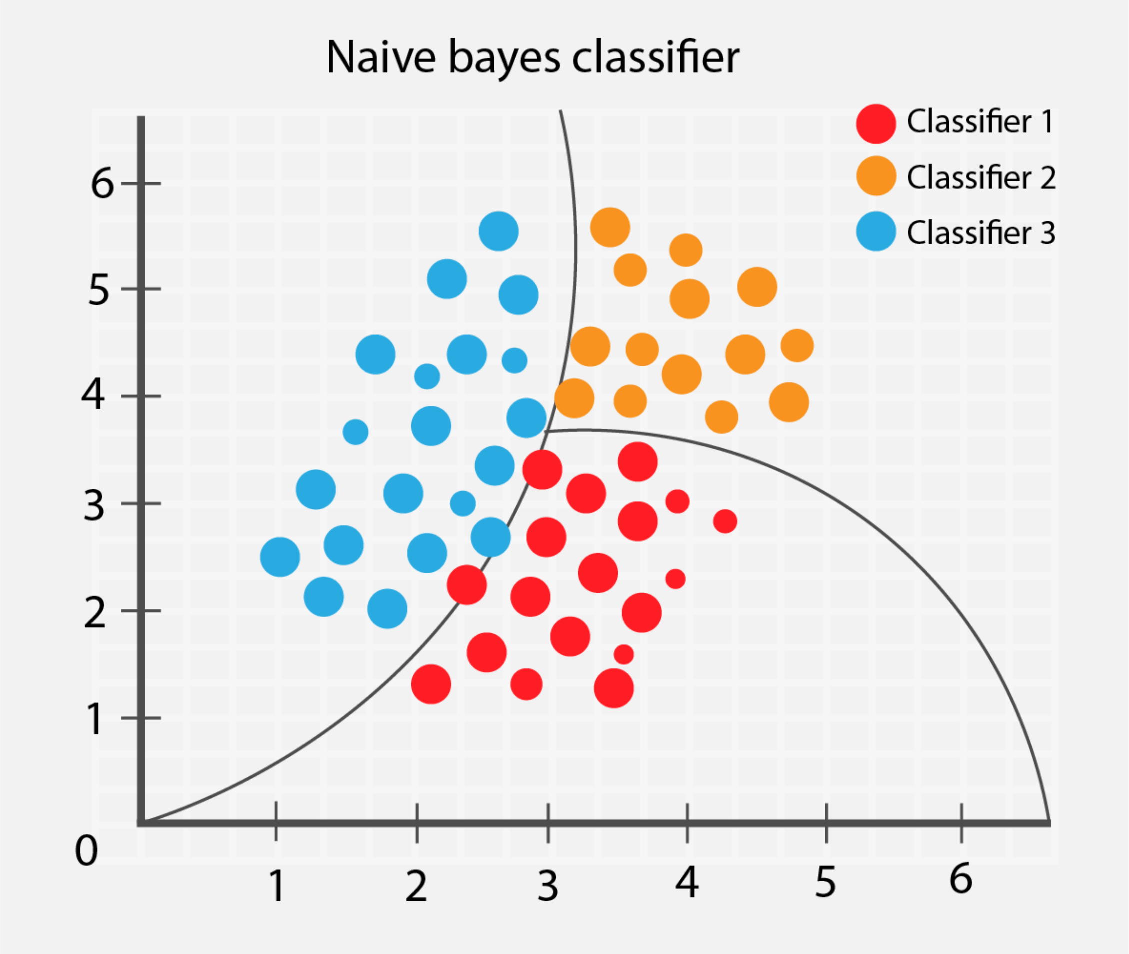
When to use

- Moderate or large training set available
- Attributes that describe instances are conditionally independent given classification

Successful applications:

- Diagnosis
- Classifying text documents

NAIVE BAYES CLASSIFIER



Salient Features

- Supervised Model
- Probabilistic model , based on **Bayes'** theorem
- Fast and Simple
- Few Parameters

Example 1- Single featured Data set

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

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
All	5	9		
	=5/14	=9/14		
	0.36	0.64		

P(Yes | Sunny) = P(Sunny | Yes) * P(Yes) / P (Sunny)

Here we have P (Sunny |Yes) = 3/9 = 0.33, P(Sunny) = 5/14 = 0.36, P(Yes)= 9/14 = 0.64

Now, P (Yes | Sunny) = 0.33 * 0.64 / 0.36 = 0.60, which has higher probability.

and P(No|Sunny)=2/5 = 0.40, which has lower probability.

Final Decision is Play=Yes

Naive Bayes Classifier

- Given a class variable V and a dependent feature vector x1 through xn, Bayes' theorem states the following relationship
- $P(V|x_1, \dots \dots x_n) = \frac{P(V)P(x_1, \dots\dots x_n|V)}{P(x_1, \dots\dots x_n)}$
- Naive Bayes assumption
- $P(x_1 \dots, x_n|V) = \prod_{i=1}^{n} P(x_i|V)$

Example 2- Multi- featured Data set

Target Attribute V

Attribute a1

Row - represented as X

RID	age	income	student	credit_rating	Class: buys_computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

Naive Bayes Classifier – Continuous features

If A_k is continous-valued

– $P(x_k|C_i)$ is usually computed based on Gaussian distribution with a mean μ and standard deviation σ:

$$g(x,\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\mu=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$\sigma=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2}$$

– and $P(x_k | C_i)$ is

$$P(x_k | C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i})$$

– μ_{Ci} and σ_{Ci}: the mean and standard deviation, respectively, of the values of attribute A_k for training instances of class C_i.

Naive Bayes Classifier (Contd...)

The instance we wish to classify is

X = (age = youth, income = medium, student = yes, credit rating = fair)

Bayesian Classification

We need to maximize $P(X \mid C_i)\, P(C_i)$, for $i = 1, 2$.

$P(C_i)$, the probability of each class, can be computed based on the training data

The probability of each class:

$$P(buys\_computer\ 9/14 = 0.643$$

$$P(buys\_computer\ 5/14 = 0.357$$

The conditional probabilities P(X | Ci) for i = 1, 2:

$$P(age = youth \mid buys\_computer = yes)\ 2/9 = 0.222$$

$$P(age = youth \mid buys\_computer = no)\ 3/5 = 0.600$$

$$P(income = medium \mid buys\_computer\ 4/9 = 0.444$$

$$P(income = medium \mid buys\_computer\ 2/5 = 0.400$$

$$P(student = yes \mid buys\_computer = ye\ 6/9 = 0.667$$

$$P(student = yes \mid buys\_computer = no\ 1/5 = 0.200$$

$$P(credit\_rating = fair \mid buys\_computer\ 6/9 = 0.667$$

$$P(credit\_rating = fair \mid buys\_computer\ 2/5 = 0.400$$

CASE STUDY – IMPLEMENT NAÏVE BAYES CLASSIFICATION FOR THE FOLLOWING DATASET

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

ASSIGNMENT

1. Define Bayes theorem, Conditional probability, Inference.
2. State MAP and ML algorithm
3. Define conditional probability and provide an example to illustrate its usage.
4. Explain the concept of a probability distribution and provide examples of discrete and continuous probability distributions.
5. Describe the concept of Bayes' theorem and its significance in probability theory.

SELF ASSESSMENT QUESTIONS

1. Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as ___________________

(a) Independent Probabilities
(b) Posterior Probabilities
(c) Prior Probabilities
(d) Dependent Probabilities

2. If A and B are two events, then the probability of exactly one of them occurs is given by _____________

(a) $P(A \cap B') + P(A' \cap B)$
(b) $P(A) + P(B) - 2P(A)\,P(B)$
(c) $P(A') + P(B') - 2P(A')\,P(B')$
(d) $P(A) + P(B) - P(A \cap B)$

Self-Assessment Questions

1. The Bayesian classifiers are not appropriate in which of the following applications?

(a) Weather prediction
(b) Fraud Detection
(c) Spam mail Filtering
(d) Target marketing

2. P(color=green ^ shape = round)/P(Orange) is known as

(a) Prior Probability
(b) Posterior Probability
(c) Likely hood

Resources

- https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c
- https://cse.iitkgp.ac.in/~sourangshu/coursefiles/FADML19S/09-ML-classification.pdf
- https://www.youtube.com/watch?v=O2L2Uv9pdDA
- https://www.youtube.com/watch?time_continue=82&v=dYMCwxgl3vk&feature=emb_logo

THANK YOU

OUR TEAM