Data Science and Artificial Intelligence

Machine Learning

Bayesian learning

Lecture No. 3

By- SIDDHARTH SABHARWAL SIR

Recap of Previous Lecture

Topic $P(h)$

Topic $P(h/D)$

Topic $P(D/h)$

Topic $P(D)$

Topic

Which is best hypothesis

ML
$\max P(D|h)$

MAP
$\max P(h/D)$

Bayes optimum classifier

Topics to be Covered

Topic — Bayes classifier

Topic — Naive Bayes classifier

Topic

Topic

Topic


"DREAM IS NOT THAT WHICH YOU SEE WHILE SLEEPING,
IT IS SOMETHING THAT DOES NOT LET YOU SLEEP"
excitement to achieve goal

Bayes optimum Classifier

$(H_1, H_2, H_3 - - -)$ hypothesis, M hypothesis

For any newpoint we find $\sum_{i=1}^{M} P(h_i/D) \cdot P(C_j/h_i)$ → For C_j class

⇒ The class which has this value ⇒ max is choosen as the Result

Bayes optimum Classifier

→ This algo give best results.

→ But this algo has a problem ⇒

Bayes optimum Classifier

- $P(h/D) \rightarrow$

⇒ So to find $P(h|D)$ we have to check the hypothesis on whole training data

N points

	Class
X_1	C
X_2	C
⋮	
X_N	C

⇒ Which need N number of computation

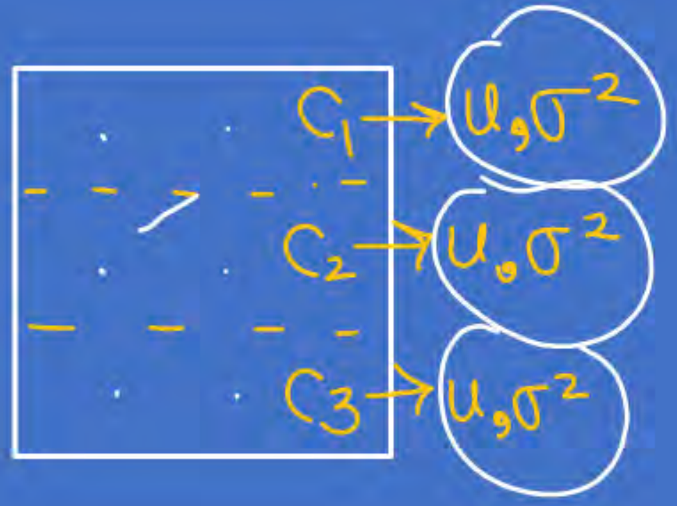
⇒ If we have m number of hypothesis then total computation will be of order of $O(Nm)$

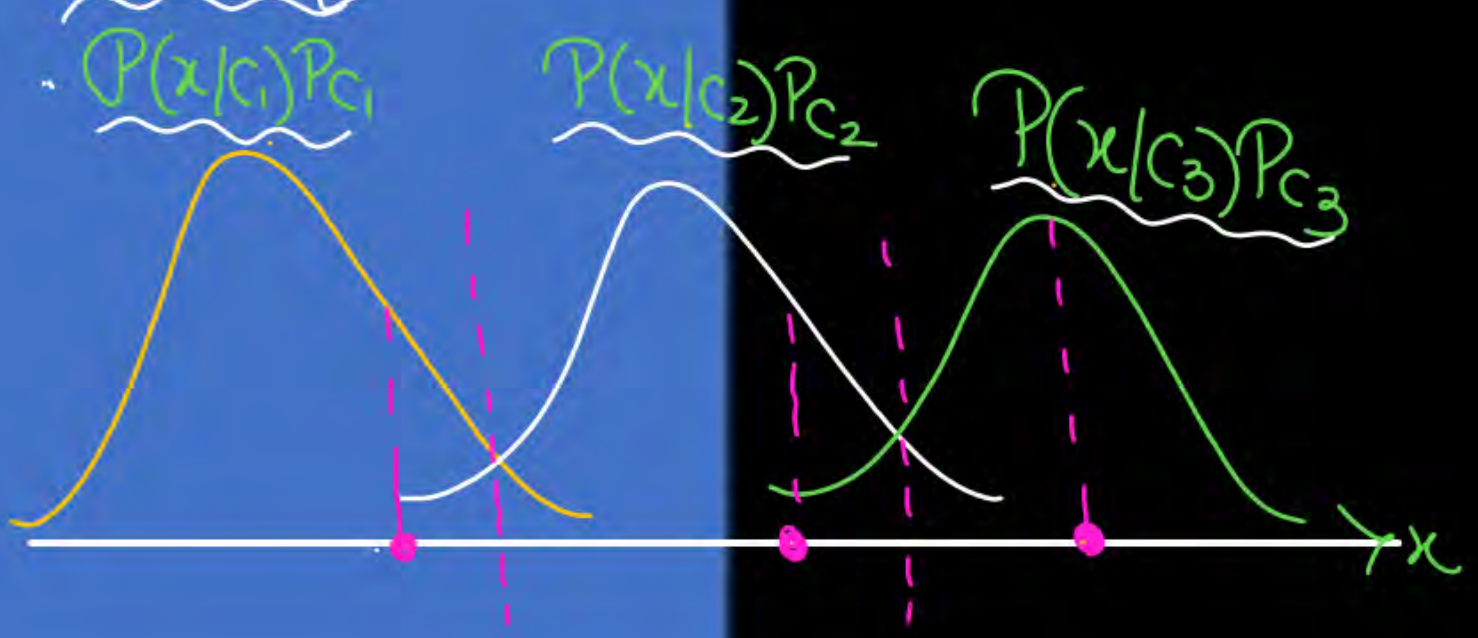
Bayes optimum Classifier

Testing wont take much time

we have P(h/D) we only need to find $P(C_j/h_i)$

a subset of data ⇒





Bayes optimum Classifier

$\Rightarrow$ (So testing is much easier)

Computational complexity of Bayes optimum Classifier

The Complexity is $O(N m)$

N ↓ Number of points

m ↓ No of hypothesis

Bayesian Decision Theory

⇒ (Brute force Bayes classifier)

Bayes Classifier

How we make the decision in finding the exact class of the point.

⇒ Simple Single map Rule Classifier

The MAP rule for decision (we have seen this earlier) we have Sample → directly we find

$P(X_i/C_j)P_{C_j}$

- we are using whole data together

Sample data

N points	$M_1 \to C_1$
	$M_2 \to C_2$
	$M_3 \to C_3$
	$M_4 \to C_4$
	⋮
	$M_j \to C_j$

P_{C_j} = Probability of any class

= Prior Probab of class

$\Rightarrow P_{C_j} = M_j/N$

Bayes Classifier

3 dimensions

$X^1 \Rightarrow (0, 1)$
$X^2 \Rightarrow (0, 1)$
$X^3 \Rightarrow (0, 1)$

Let us have 2 classes

$P(X_i / C_j)$

any point in training data

Show this with help of an example...lets take 3 dimensions then explain it

So we need to find

$P(000/C_1)$
$P(001/C_1)$
$P(010/C_1)$
$P(011/C_1)$
$P(100/C_1)$
$P(101/C_1)$
$P(110/C_1)$
$P(111/C_1)$

Similarly for class 2
(16 parameters)

Bayes classifier ⇒ (all data together)

⇒ (Single hypothesis)

For any point (x_i) the class C_j which has max value of $\left(P(x_i/C_j)P C_j\right)$ is assigned to x_i

- $\underbrace{\text{Bayes optimum Cl.}}_{\text{ensemble learning}}$ and $\underbrace{\text{Bayes Classifier}}_{\text{Single hypothesis}}$ are similar

So here we need

$$P(x_i/C_j)\,P_{C_j} \qquad P_{C_j} \rightarrow \frac{M_j}{N}$$

$$P[x_i^1, x_i^2, x_i^3 - - - x_i^D / C_j]\,P_{C_j} \rightarrow ((10^D) \times m)$$

⇒ D number of dimension

⇒ each dimension is categorical with 10 values

⇒ we have M number of classes

So total parameter needed to create hypothesis

$\Rightarrow \underbrace{a^D \times M}$

$+$

we need $P_{c_j} = M$
No of Parameter

a: No of values each dimension can take
D: No of dimension
M: No of classes

$\Rightarrow$ we need $(a^D \times m + m)$ No of Parameters.

2 dimension

$X^1 \Rightarrow (0,1)$

$X^2 \Rightarrow (0,1)$

$C_1 \Rightarrow$ Class.

$C_2 \Rightarrow$

$P(x_i/C_j)\,P_{C_j}$

X_1	X_2	Class
0	0	C_1
1	0	C_1
0	1	C_2
1	1	C_2
1	1	C_1
0	1	C_2
1	0	C_1
1	1	C_2
0	1	

$P(00|C_1)$ $P(00|C_2)$

$P(01|C_1)$ $P(01|C_2)$

$P(10|C_1)$ $P(10|C_2)$

$P(11|C_1)$ $P(11|C_2)$

8+2 Parameter

Bayes Classifier

Complexity in this classifer…

- So in Bayes optimum classifier also each hypothesis need these many Parameters

- So the Bayes optimum Cl. and Brute force Bayes Cl. need a large training data and also the large No of Parameters.

$P(A, B/C) \Rightarrow$ If A and B are Independent $\Rightarrow P(A|C)\,P(B|C)$

D: dimension

M: classes

each dimen take a values

$$\underbrace{P(\underbrace{x_i^1, x_i^2 - - x_i^D}_{(a^D \times m) \Rightarrow \text{Parameter}} / C_j)}$$

$$\Rightarrow \overset{a}{\Uparrow}\underbrace{P(x_i^1/C_j)}_{\Downarrow} \overset{a}{\Uparrow}P(x_i^2/C_j) - - \overset{a}{\Uparrow}P(x_i^D/C_j) \Rightarrow Da \text{ for each class}$$

for a single class

$$\underbrace{P(x_i^1/C_j)}_{a \text{ Parameter}} \Rightarrow P(x_i^1 = V_1/C_j) \Rightarrow$$

$$P(x_i^1 = V_2/C_j) \Rightarrow$$

$$\vdots$$

$$P(x_i^1 = V_a/C_j) \Rightarrow$$

$\Rightarrow Dam$ number of Parameters totally.

Naïve Bayes Algorithm

Bayes classifier

$\max \left[P_{C_j} \cdot P(x_i / C_j) \right]$

$\max \left[P_{C_j} \cdot P(x_i^1, x_i^2, \ldots x_i^D / C_j) \right]$

Combo → a^D

$\xrightarrow[\text{Assumption}]{\text{Naive}}$ $\max \left[P_{C_j} \cdot P(x_i^1, x_i^2 \ldots x_i^D / C_j) \right]$

* Each dimension * is independent of each other

$\max \left[P_{C_j} P(x^1 / C_j) P(x^2 / C_j) \ldots P(x_i^D / C_j) \right]$

$(Dam + m)$ no. of parameters.

So in naive Bayes Classification

$$\boxed{\max \left[P_{C_j} \underbrace{P(x_i^1 / C_j) \, P(x_i^2 / C_j)}_{} \cdots \right]}$$

Naïve Bayes Algorithm

The fundamental Naive Bayes assumption is that each feature makes an:

- ❑ **Feature independence:** The features of the data are conditionally independent of each other, given the class label.
- ❑ **Features are equally important:** All features are assumed to contribute equally to the prediction of the class label.

Naïve Bayes Algorithm

We can convert the MAP into...

- So in Naive Bayes we have to find all Parameter

$$\left\{\begin{matrix} P_{C_1} \\ P_{C_2} \Rightarrow \frac{m_j^{\circ}}{N} \\ \vdots \\ P_{C_j} \end{matrix}\right\}$$

$$P(x^1/C_1) \Rightarrow P(x^1 = V_1 | C_1)$$

$$P(x^1 = V_2 | C_1)$$

$$\vdots$$

$$P(x^1 = V_a | C_1)$$

Naïve Bayes Classifier

4 dimension

Outlook (3)	Temp. (3)	Humidity (2)	Wind (2)	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes ✓
Rain	Mild	High	Weak	Yes ✓
Rain	Cool	Normal	Weak	Yes ✓
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Weak	Yes ✓
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes ✓
Rain	Mild	Normal	Strong	Yes ✓
Sunny	Mild	Normal	Strong	Yes ✓
Overcast	Mild	High	Strong	Yes ✓
Overcast	Hot	Normal	Weak	Yes ✓
Rain	Mild	High	Strong	No

We have to calculate ....

Yes or No

$\Rightarrow P_{Yes} = 9/14$

$\Rightarrow P_{No} = 5/14$

Bayesian Decision Theory

P(Sunny|Yes) P(Sunny|No)
P(overcast|Yes) P(overcast|No)

Naïve Bayes Classifier

P(Rain|Yes) P(Rain|No)

→ Sunny/overcast/Rain

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny S	Hot	High	Weak	No
Sunny S	Hot	High	Strong	No
Overcast ✓ O	Hot	High	Weak	Yes
Rain ✓ R	Mild	High	Weak	Yes
Rain ✓ R	Cool	Normal	Weak	Yes
Rain R	Cool	Normal	Strong	No
Overcast ✓ O	Cool	Normal	Weak	Yes
Sunny S	Mild	High	Weak	No
Sunny ✓ S	Cool	Normal	Weak	Yes
Rain ✓ R	Mild	Normal	Strong	Yes
Sunny ✓ S	Mild	Normal	Strong	Yes
Overcast ✓ O	Mild	High	Strong	Yes
Overcast ✓ O	Hot	Normal	Weak	Yes
Rain R	Mild	High	Strong	No

Outlook	P(O/Yes)	P(O/No)
Sunny	$P(S/Y) = 2/9$	$P(S|No) = 3/5$
Overcast	$P(overcast/Y) = 4/9$	$P(O|No) = 0$
Rain	$P(R/Y) = 3/9$	$P(R/No) = 2/5$

Bayesian Decision Theory

P(HOT|Y) P(HOT|N)
P(mild|Y) P(mild|No)

Naïve Bayes Classifier

P(cold|Y) P(cold|No)

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot H	High	Weak	No
Sunny	Hot H	High	Strong	No
Overcast	Hot H ✓	High	Weak	Yes
Rain	Mild M ✓	High	Weak	Yes
Rain	Cool C ✓	Normal	Weak	Yes
Rain	Cool C	Normal	Strong	No
Overcast	Cool C ✓	Normal	Weak	Yes
Sunny	Mild M	High	Weak	No
Sunny	Cool C ✓	Normal	Weak	Yes
Rain	Mild M ✓	Normal	Strong	Yes
Sunny	Mild M ✓	Normal	Strong	Yes
Overcast	Mild M ✓	High	Strong	Yes
Overcast	Hot H ✓	Normal	Weak	Yes
Rain	Mild M	High	Strong	No

Temperature	P(T/Yes)	P(T/No)
Hot	$P(H/Y) = 2/9$	$P(H/N) = 2/5$
Mild	$P(M/Y) = 4/9$	$P(M/N) = 2/5$
Cold	$P(C/Y) = 3/9$	$P(C/N) = 1/5$

Bayesian Decision Theory

P(High/Y) P(Normal|Yes)
P(High|No) P(Normal|No)

Naïve Bayes Classifier

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High H	Weak	No
Sunny	Hot	High H	Strong	No
Overcast	Hot	High ✓H	Weak	Yes
Rain	Mild	High ✓H	Weak	Yes
Rain	Cool	Normal ✓	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal ✓	Weak	Yes
Sunny	Mild	High H	Weak	No
Sunny	Cool	Normal ✓	Weak	Yes
Rain	Mild	Normal ✓	Strong	Yes
Sunny	Mild	Normal ✓	Strong	Yes
Overcast	Mild	High ✓H	Strong	Yes
Overcast	Hot	Normal ✓	Weak	Yes
Rain	Mild	High H	Strong	No

Humidity	P(H/Yes)	P(H/No)
High	$P(High\|Y)=3/9$	$P(High\|No)=4/5$
Normal	$P(Normal\|Y)=\frac{6}{9}$	$P(normal/No)=\frac{1}{5}$

Bayesian Decision Theory

Naïve Bayes Classifier

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak ✓	Yes
Rain	Mild	High	Weak ✓	Yes
Rain	Cool	Normal	Weak ✓	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Weak ✓	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak ✓	Yes
Rain	Mild	Normal	Strong ✓	Yes
Sunny	Mild	Normal	Strong ✓	Yes
Overcast	Mild	High	Strong ✓	Yes
Overcast	Hot	Normal	Weak ✓	Yes
Rain	Mild	High	Strong	No

Wind	P(W/Yes)	P(W/No)
Weak	P(weak/Y) = 6/9	P(weak/No) = 2/5
Strong	P(strong/Y) = 3/9	P(strong/No) = 3/5

Naïve Bayes Algorithm

No

Outlook	Temperature	Humidity	Wind
Sunny	Cool	High	Strong

For each class $\left(P C_j \cdot P(x_i^1/C_j) P(x_i^2/C_j) \ldots\right)$

$P_{yes} P(S/Y) P(C/Y) P(H/Y) P(S/Y) \Rightarrow \frac{9}{14} \times \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \Rightarrow 5.29 \times 10^{-3}$

$P_{No} P(S/No) P(C/N) P(H/N) P(S/N) \Rightarrow \frac{5}{14} \times \frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times 3/5 \Rightarrow .0205$ ✓

Naïve Bayes Algorithm

Complexity in naïve bayes classfier…

Naïve Bayes Algorithm

Consider the following dataset showing the result whether a person has passed or failed the exam based on various factors. Suppose the factors are independent to each other. We want to classify a new instance with Confident=Yes, Studied=Yes, and Sick=No.

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	Yes	Pass

A. Pass
B. Fail

Confident	Studied	Sick	Result
Y	N	N	F
Y	N	Y	P
N	Y	Y	F
N	Y	N	P
Y	Y	Y	P

Confident	Studied	Sick
$P(Y/F) = 1/2$	$P(Y\|F) = 1/2$	$P(Y/F) \rightarrow 1/2$
$P(N\|F) = 1/2$	$P(N\|F) = 1/2$	$P(N/F) \rightarrow 1/2$
$P(Y\|P) = 2/3$	$P(Y\|P) = 2/3$	$P(Y\|P) \rightarrow 2/3$
$P(N\|P) = 1/3$	$P(N\|P) = 1/3$	$P(N\|P) \rightarrow 1/3$

$P_F = 2/5$

$P_P = 3/5$

Naïve Bayes Algorithm

Consider the following dataset showing the result whether a person has passed or failed the exam based on various factors. Suppose the factors are independent to each other. We want to classify a new instance with Confident=Yes, Studied=Yes, and Sick=No.

Pass

Confident	Studied	Sick	Result
Y	N	N	F
Y	N	Y	P
N	Y	Y	F
N	Y	N	P
Y	Y	Y	P

$P_F = 2/5$

$P_P = 3/5$

Confident	Studied	Sick
$P(Y/F) = 1/2$	$P(Y\|F) = 1/2$	$P(Y/F) \rightarrow 1/2$
$P(N/F) = 1/2$	$P(N\|F) = 1/2$	$P(N/F) \rightarrow 1/2$
$P(Y\|P) = 2/3$	$P(Y\|P) = 2/3$	$P(Y\|P) \rightarrow 2/3$
$P(N/P) = 1/3$	$P(N\|P) = 1/3$	$P(N/P) \rightarrow 1/3$

Pass ⇒

$P_P \, P(Y/P) \, P(Y/P) \, P(N/P) \Rightarrow$

$3/5 \times 2/3 \times 2/3 \times 1/3 \Rightarrow .088$

$P_F \, P(Y/F) \, P(Y/F) \, P(N/F) \Rightarrow$

$\frac{2}{5} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \Rightarrow .05$

Bayesian Decision Theory

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Weak	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Strong	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Outlook	P(O/Yes)	P(O/No)
Sunny		
Overcas		

Temperature	P(T/Yes)	P(T/No)
Hot		
Mild		

Humidity	P(H/Yes)	P(H/No)
High		
Normal		

Wind	P(W/Yes)	P(W/No)
Weak		
Strong		

Naïve Bayes Algorithm

Additive smoothing technique

Solving the zero-probability problem…

Naïve Bayes Algorithm

What if the dimension are continuous in nature

The numeric weather data with summary statistics

outlook			temperature		humidity		windy			play	
	yes	no	yes	no	yes	no		yes	no	yes	no
sunny	2	3	83	85	86	85	false	6	2	9	5
overcast	4	0	70	80	96	90	true	3	3		
rainy	3	2	68	65	80	70					
			64	72	65	95					
			69	71	70	91					
			75		80						
			75		70						
			72		90						
			81		75						

Bayesian Decision Theory

The numeric weather data with summary statistics												
outlook			temperature		humidity		windy			play	
	yes	no	yes	no	yes	no		yes	no	yes	no
sunny	2	3	83	85	86	85	false	6	2	9	5
overcast	4	0	70	80	96	90	true	3	3		
rainy	3	2	68	65	80	70					
			64	72	65	95					
			69	71	70	91					
			75		80						
			75		70						
			72		90						
			81		75						

Q: Consider a classification problem with 10 classes $y \in \{1,2,\ldots,10\}$, and two binary features $x1, x2 \in \{0,1\}$.

Suppose:

$p(Y=y)=1/10$,

$p(x1=1|Y=y)=y/10$,

$p(x2=1|Y=y)=y/540$

Which class will naïve Bayes classifier produce on a test item with $(x1=0, x2=1)$?

A. 1

B. 3

C. 5

D. 8

E. 10

Multinomial Naïve Bayes Algorithm

THANK - YOU