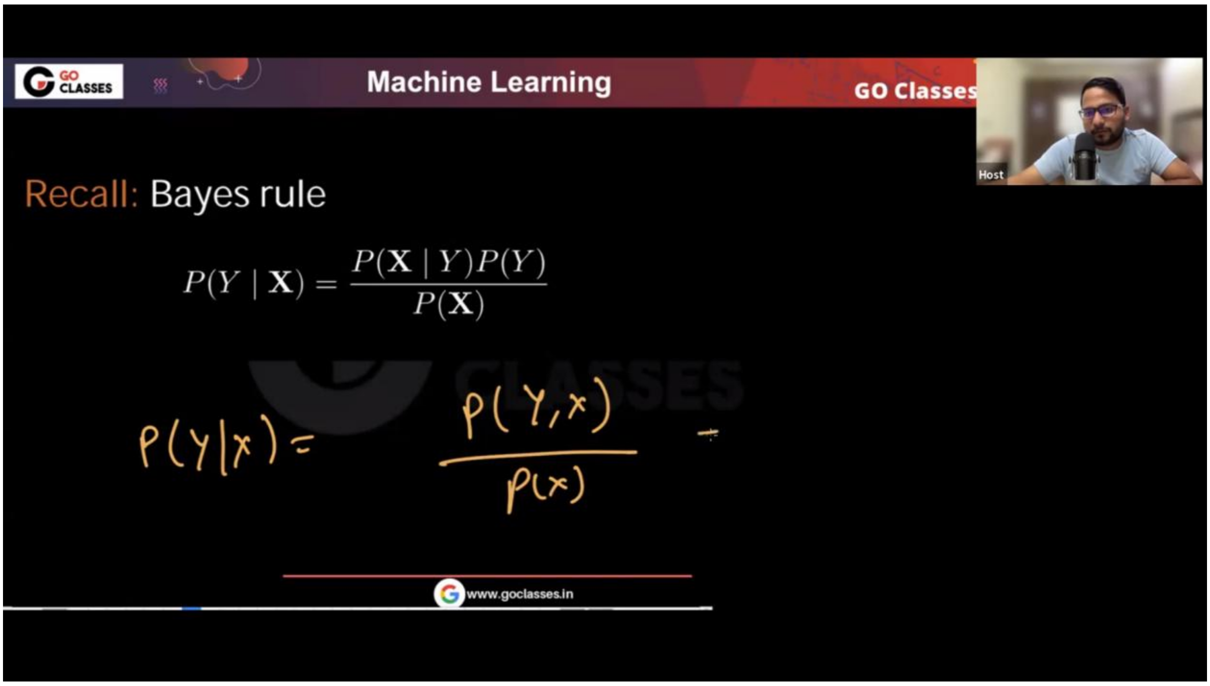
Bayes Optimal Classifier
and
Naïve Bayes

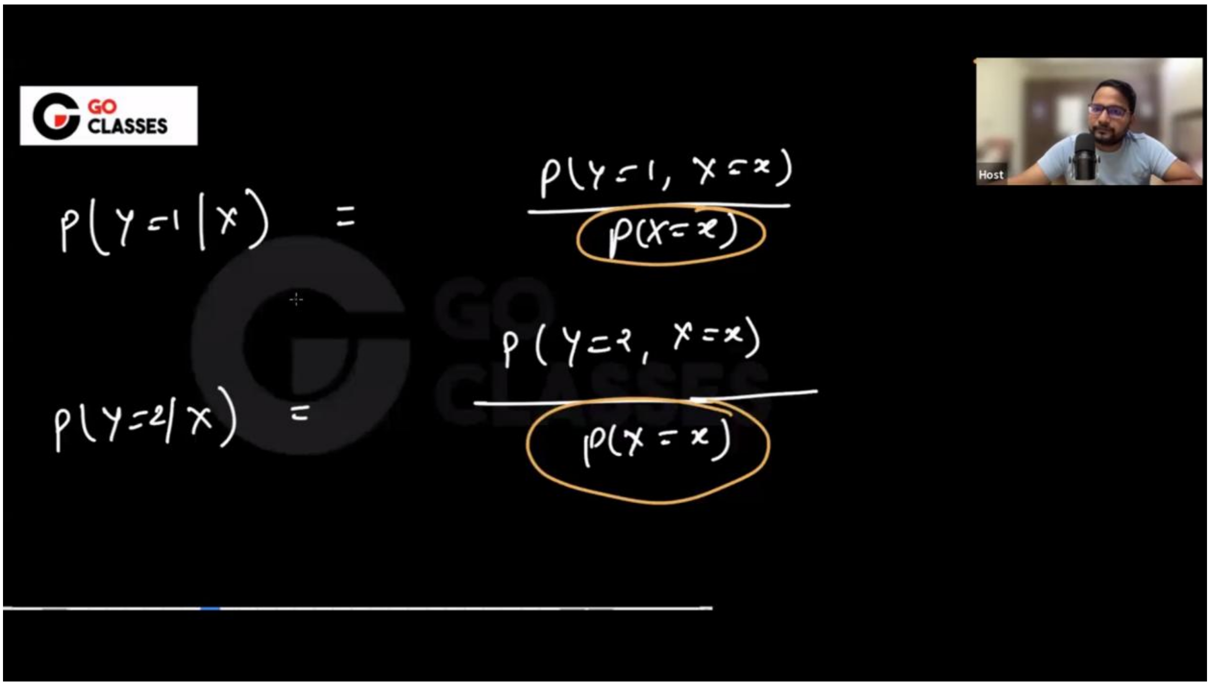
Bayes Optimal Classifier

Suppose you know P(Y | X) exactly, how should you classify?

Bayes optimal classifier:

$$\hat{y} = \arg\max_{y} P(Y = y \mid X = x)$$





To compare which is more

Example: avocados

You have a green-black avocado, and want to know if it is ripe. Based on this data, would you predict that your avocado is ripe or unripe?

Train Data:

color	ripeness
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	unripe
purple-black	ripe
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	ripe
green-black	unripe
purple-black	ripe

Test Point: green-black avocado

$P(\text{ripe}|\text{green-black}) = ?$

$P(\text{unripe}|\text{green-black}) = ?$

10:05 -1:48:06

GO CLASSES Machine Learning GO Classes

$$P(\text{class}|\text{features}) = \frac{P(\text{class})\cdot P(\text{features}|\text{class})}{P(\text{features})}$$

We can ignore the denominator P(features) as this is just normalisation factor:

$$P(class|features) \propto P(class) \cdot P(features|class)$$

10:26 -1:47:45

Machine Learning

GO Classes

color	ripeness
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	unripe
purple-black	ripe
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	ripe
green-black	unripe
purple-black	ripe

- There are 7 out of 11 rows with "ripe" labels:

$$P(\text{ripe}) = \frac{7}{11}$$

- There are 4 out of 11 rows with "unripe" labels:

$$P(\text{unripe}) = \frac{4}{11}$$

$\frac{7}{11}$ ripe

$\frac{4}{11}$ unripe

www.goclasses.in

11:41 -1:46:30

color	ripeness
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	unripe
purple-black	ripe
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	ripe
green-black	unripe
purple-black	ripe

- There are 7 out of 11 rows with "ripe" labels:

$$P(\text{ripe}) = \frac{7}{11}$$

- There are 4 out of 11 rows with "unripe" labels:

$$P(\text{unripe}) = \frac{4}{11}$$

$\frac{7}{11}$ ripe

$\frac{4}{11}$ unripe

$P(Y)$ ✓ Prior

$P(Y|x)$ Posterior Prob.

x

if we dont look at posterior or only look at prior $\Rightarrow$ classification will always be ripe.

Machine Learning

GO Classes

color	ripeness
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	unripe
purple-black	ripe
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	ripe
green-black	unripe
purple-black	ripe

- There are 7 out of 11 rows with "ripe" labels:

$$P(\text{ripe}) = \frac{7}{11}$$

- There are 4 out of 11 rows with "unripe" labels:

$$P(\text{unripe}) = \frac{4}{11}$$

- Out of 7 rows with "ripe" labels, 3 have "green-black":

$$P(\text{green-black} \mid \text{ripe}) = \frac{3}{7}$$

- Out of 4 rows with "unripe" labels, 2 have "green-black":

$$P(\text{green-black} \mid \text{unripe}) = \frac{2}{4} = \frac{1}{2}$$

19:16 -1:38:55

Machine Learning

GO Classes

color	ripeness
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	unripe
purple-black	ripe
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	ripe
green-black	unripe
purple-black	ripe

$$P(\text{ripe} \mid \text{green-black}) \propto P(\text{ripe}) \cdot P(\text{green-black} \mid \text{ripe})$$
$$= \frac{7}{11} \cdot \frac{3}{7} = \frac{3}{11}$$

$$P(\text{unripe} \mid \text{green-black}) \propto P(\text{unripe}) \cdot P(\text{green-black} \mid \text{unripe})$$
$$= \frac{4}{11} \cdot \frac{1}{2} = \frac{2}{11}$$

Since:

$$P(\text{ripe} \mid \text{green-black}) = \frac{3}{11} > P(\text{unripe} \mid \text{green-black}) = \frac{2}{11}$$

Prediction: Given that the avocado is green-black, it is more likely to be **ripe**.

Machine Learning

GO Classes

Question:

Suppose that you are trying to solve a binary classification problem, and your data set has 4 attributes. Each attribute can take 3 possible values.
If you modeled the full joint distribution of the attributes and the class label, how many parameters would you need?

$$P(\underset{2}{\underset{\downarrow}{Y}}\ \underset{3}{\underset{\downarrow}{X_1}}\ \underset{3}{\underset{\downarrow}{X_2}}\ \underset{3}{\underset{\downarrow}{X_3}}\ \underset{3}{\underset{\downarrow}{X_4}}) = 2 \times 3 \times 3 \times 3 \times 3 - 1$$

https://www.cs.cmu.edu/~epxing/Class/10701/exams/midterm2005sp-solution.pdf

Machine Learning

GO Classes

Solution

We would have the joint distribution:

$$P(X_1, X_2, X_3, X_4, Y)$$

where X_1, X_2, X_3, X_4 are the attributes and Y is the class label.
The total number of entries in this table would be:

$$3^4 \times 2 = 162 \text{ entries}$$

Since probabilities must sum to 1, the number of independent parameters is:

$$3^4 \times 2 - 1 = 161 \text{ independent parameters.}$$

Naive Bayes

Naïve Bayes Assumption

Naïve Bayes assumes $y=0$

$$P(X_1, X_2, \ldots, X_d \mid Y) = P(X_1 \mid Y)\, P(X_2 \mid Y) \cdots P(X_d \mid Y)$$

i.e., that X_i and X_j are **conditionally independent** given Y, for all $i \neq j$

56:43 -1:01:28

GO CLASSES

$$P(X_1, X_2, \ldots, X_d \mid Y) = P(X_1 \mid Y)\, P(X_2 \mid Y) \cdots P(X_d \mid Y) \checkmark$$

Naive Bayes assumption

$$\Downarrow \times$$

$$P(X_1, X_2, \ldots, X_d) = P(X_1) \cdot P(X_2) \cdot \cdots P(X_d) \times$$

Not a Naive Bayes assumption

GO CLASSES — Machine Learning — GO Classes

Naïve Bayes classifier

$$\begin{aligned}
\hat{y} &= argmax_y\, P(Y = y \mid X = x) \\
&= argmax_y P(Y = y) P(X = x \mid Y = y) \longrightarrow \text{ignore denominator} \\
&= argmax_y P(Y = y) \prod_{i=1}^{d} P(X_i = x_i \mid Y = y)
\end{aligned}$$

Bayes rule

+ Conditional independence assumption

www.goclasses.in

Machine Learning

GO Classes

age	income	student	credit	C_i: buy
youth	high	no	fair	C_2: no
youth	high	no	excellent	C_2: no
middle-aged	high	no	fair	C_1: yes
senior	medium	no	fair	C_1: yes
senior	low	yes	fair	C_1: yes
senior	low	yes	excellent	C_2: no
middle-aged	low	yes	excellent	C_1: yes
youth	medium	no	fair	C_2: no
youth	low	yes	fair	C_1: yes
senior	medium	yes	fair	C_1: yes
youth	medium	yes	excellent	C_1: yes
middle-aged	medium	no	excellent	C_1: yes
middle-aged	high	yes	fair	C_1: yes
senior	medium	no	excellent	C_2: no

Test Point:

The sample we wish to classify is

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

$2^5 - 1 = \underline{\underline{31}}$

Machine Learning

GO Classes

The sample we wish to classify is

$\mathbf{X} = (age = youth, income = medium, student = yes, credit = fair)$

We need to maximize $P(\mathbf{X}|C_i)P(C_i)$, for $i = 1, 2$. $P(C_i)$, the a priori probability of each class, can be estimated based on the training samples:

$$P(buy = yes) = \frac{9}{14}$$

$$P(buy = no) = \frac{5}{14}$$

To compute $P(\mathbf{X}|C_i)$, for $i = 1, 2$, we compute the following conditional probabilities:

$$P(age = youth|buy = yes) = \frac{2}{9}$$

$$P(age = youth|buy = no) = \frac{3}{5}$$

$$P(income = medium|buy = yes) = \frac{4}{9}$$

$$P(Y|X) = \frac{P(Y) \cdot P(X|Y)}{P(X)}$$

1:08:16 -49:55

Machine Learning

GO Classes

The sample we wish to classify is

$$\mathbf{X} = (age = youth, income = medium, student = yes, credit = fair)$$

We need to maximize $P(\mathbf{X}|C_i)P(C_i)$, for $i = 1, 2$. $P(C_i)$, the a priori probability of each class, can be estimated based on the training samples:

$$P(buy = yes) = \frac{9}{14}$$

$$P(buy = no) = \frac{5}{14}$$

Prior Probablities

$$P(Y|X) = \frac{P(Y) \cdot P(X|Y)}{P(X)}$$

ignore

To compute $P(\mathbf{X}|C_i)$, for $i = 1, 2$, we compute the following conditional probabilities:

$$P(age = youth|buy = yes) = \frac{2}{9}$$

$$P(age = youth|buy = no) = \frac{3}{5}$$

$$P(income = medium|buy = yes) = \frac{4}{9}$$

www.goclasses.in

1:09:17 -48:54

Machine Learning

GO Classes

The sample we wish to classify is

$$\mathbf{X} = (age = youth, income = medium, student = yes, credit = fair)$$

We need to maximize $P(\mathbf{X}|C_i)P(C_i)$, for $i = 1, 2$. $P(C_i)$, the a priori probability of each class, can be estimated based on the training samples:

$$\left.\begin{aligned} P(buy = yes) &= \frac{9}{14} \\ P(buy = no) &= \frac{5}{14} \end{aligned}\right]$$ Prior Probablities

$P(Y|X)$

$= \dfrac{P(Y) \cdot P(X|Y)}{P(X)}$

ignore

To compute $P(\mathbf{X}|C_i)$, for $i = 1, 2$, we compute the following conditional probabilities:

$$P(age = youth|buy = yes) = \frac{2}{9}$$

$$P(age = youth|buy = no) = \frac{3}{5}$$

$$P(income = medium|buy = yes) = \frac{4}{9}$$

www.goclasses.in

1:11:04 -47:07

Machine Learning

GO Classes

$$P(income = medium|buy = no) = \frac{2}{5}$$

$$P(student = yes|buy = yes) = \frac{6}{9}$$

$$P(student = yes|buy = no) = \frac{1}{5}$$

$$P(credit = fair|buy = yes) = \frac{6}{9}$$

$$P(credit = fair|buy = no) = \frac{2}{5}$$

Using the above probabilities, we obtain

$$\begin{aligned} P(\mathbf{X}|buy = yes) &= P(age = youth|buy = yes) \\ &\quad P(income = medium|buy = yes) \\ &\quad P(student = yes|buy = yes) \\ &\quad P(credit = fair|buy = yes) \\ &= \frac{2}{9}\frac{4}{9}\frac{6}{9}\frac{6}{9} = 0.044. \end{aligned}$$

1:11:56 -46:15

Similarly,

$$P(\mathbf{X}|buy = no) = \frac{3}{5}\frac{2}{5}\frac{1}{5}\frac{2}{5} = 0.019$$

To find the class that maximizes $P(\mathbf{X}|C_i)P(C_i)$, we compute

$$P(\mathbf{X}|buy = yes)P(buy = yes) = 0.028$$

$$P(\mathbf{X}|buy = no)P(buy = no) = 0.007$$

Thus the naive Bayesian classifier predicts $buy = yes$ for sample $\mathbf{X}$.

Suppose we are given the following dataset, where A, B, C are input binary random variables, and y is a binary output whose value we want to predict.

A	B	C	y
0	0	1	0
0	1	0	0
1	1	0	0
0	0	1	1
1	1	1	1
1	0	0	1
1	1	0	1

$$\overset{2/3}{P(A=0|y=0)} \cdot \overset{1/3}{P(B=0|y=0)} \cdot \overset{2/3}{P(C=1|y=0)}$$

Naive assump.

$$P(A=0, B=0, C=1|y=0)$$

How would a **naive** Bayes classifier predict y given this input:

$A = 0, B = 0, C = 1$. Assume that in case of a tie the classifier always prefers to predict 0 for y.

$$P(y=0 | A=0, B=0, C=1) = \frac{\overset{\frac{3}{7}}{P(y=0)} \cdot}{P(A=0, B=0, C=1)}$$

$$\frac{3}{7} \times \left(\frac{2}{3} \cdot \frac{1}{3} \cdot \frac{1}{3}\right) \qquad \frac{4}{7}\left(\frac{1}{4} \cdot \frac{2}{4} \cdot \quad \right)$$

1:18:36 -39:35

Answer: The classifier will predict 1

$P(y=0) = 3/7; P(y=1) = 4/7$
$P(A=0|y=0) = 2/3; P(B=0|y=0) = 1/3; P(C=1|y=0) = 1/3$
$P(A=0|y=1) = 1/4; P(B=0|y=1) = 1/2; P(C=1|y=1) = 1/2$

Predicted y maximizes $P(A=0|y)P(B=0|y)P(C=1|y)P(y)$
$P(A=0|y=0)P(B=0|y=0)P(C=1|y=0)P(y=0) = 0.0317$
$P(A=0|y=1)P(B=0|y=1)P(C=1|y=1)P(y=1) = 0.0357$
Hence, the predicted y is 1.

1:18:51 -39:20

Question:

A_1	A_2	Class Label Y
True	True	+
True	True	+
True	False	−
False	False	+
False	True	−
False	True	−
False	False	−
True	False	+

$$\frac{P(y=+)\cdot P(A_1=T, A_2=T \mid y=+)}{P(A_1=T, A_2=T)}$$

What is the probability $P(Y=+ \mid A_1 = \text{True}, A_2 = \text{True})$ by applying the Naïve Bayes classifier (show each step in computing the probability)?

1:22:25 -35:46

What is the probability $P(Y=+\mid A_1=\text{True}, A_2=\text{True})$ by applying the Naïve Bayes classifier (show each step in computing the probability)?

$$P(Y=+)=0.5$$
$$P(A_1=\text{True}|Y=+)=0.75$$
$$P(A_1=\text{True}|Y=-)=0.25$$
$$P(A_2=\text{True}|Y=+)=0.5$$
$$P(A_2=\text{True}|Y=-)=0.5$$

$$P(Y=+\mid A_1=\text{True}, A_2=\text{True})$$
$$=\frac{P(Y=+)P(A_1=\text{True}|Y=+)P(A_2=\text{True}|Y=+)}{P(Y=+)P(A_1=\text{True}|Y=+)P(A_2=\text{True}|Y=+)+P(Y=-)P(A_1=\text{True}|Y=-)P(A_2=\text{True}|Y=-)}$$
$$=\frac{0.5*0.75*0.5}{0.5*0.75*0.5+0.5*0.25*0.5}=0.75$$

What label should you predict for input $(A_1=\text{True}, A_2=\text{True})$ based on your result above ?

$$Y=+$$

1:23:55 -34:16

Consider a classification problem with 10 classes $y \in \{1,2,\ldots,10\}$, and two binary features $x_1, x_2 \in \{0,1\}$. Suppose $p(Y=y)=1/10$, $p(x_1=1\mid Y=y)=y/10$, $p(x_2=1\mid Y=y)=y/540$. Which class will naive Bayes classifier produce on a test item with $(x_1=0, x_2=1)$?

(A) 1 (B) 3 (C) 5 (D) 8 (E) 10

C

$$\begin{aligned}
&\operatorname{argmax}_y P(y\mid x_1=0,x_2=1) && (1)\\
&=\operatorname{argmax}_y P(x_1=0,x_2=1\mid y)p(y) && (2)\\
&=\operatorname{argmax}_y P(x_1=0\mid y)P(x_2=1\mid y)p(y) && (3)\\
&=\operatorname{argmax}_y (1-P(x_1=1\mid y))P(x_2=1\mid y)p(y) && (4)\\
&=\operatorname{argmax}_y (1-y/10)y/540 && (5)\\
&=\operatorname{argmax}_y (10-y)y && (6)
\end{aligned}$$

Now you could enumerate y to find the maximum. Or pretend y is continuous, taking derivative and set to zero: 10-2y=0 leads to y=5.

Machine Learning

GO Classes

Question:

Consider a classification problem with two binary features, $x_1, x_2 \in \{0,1\}$. Suppose $P(Y = y) = 1/32$, $P(x_1 = 1 | Y = y) = y/46$, $P(x_2 = 1 | Y = y) = y/62$. Which class will naive Bayes classifier produce on a test item with $x_1 = 1$ and $x_2 = 0$?

- A 16
- B 26
- C 31
- D 32

$$P(y = y \mid x_1 = 1, x_2 = 0)$$

$$= \underbrace{P(y = y)}_{\frac{1}{32}} \cdot \underbrace{P(x_1 = 1 | y)}_{\frac{y}{46}} \cdot \underbrace{P(x_2 = 0 | y)}_{\left(1 - \frac{y}{62}\right)}$$

https://pages.cs.wisc.edu/~sharonli/courses/cs540_spring2021/documents/perceptron-quiz.pdf

www.goclasses.in

1:37:46 -20:25

Machine Learning

GO Classes

GATE 2024 -- Data Science and Artificial Intelligence (DA) -- Sample Question Paper

Q.28	Given a K-class dataset containing N points, where sample points are described using D discrete features with each feature capable of taking V values, how many parameters need to be estimated for Naïve Bayes Classifier?
(A)	$V^D K$
(B)	K^{V^D}
(C)	$VDK + K$
(D)	$K(V + D)$

www.goclasses.in

1:49:02 -09:09