Chapter - 6 Bayesian Learning

(Source: DM by Vipin Tan)

↳ Models probabilistic relationship b/w attribute set and class variable

→ Class labels can't be predicted with certainty

X_1	X_2	Y
Diet	Exercise	Heart Attack
Healthy	Regularly →	0
		1

Animal

Legs	Pattern	Y
4	Black & White Stripes	Zebra

Zebra

Cat

Naive Bayes

$P(A, B) = P(4, R) = 2/52$

Joint

Bayes Theorem

$$P(A/B) = \frac{P(A)\,P(B/A)}{P(B)}$$

52

$P(\underset{A}{\text{Fair}}/\underset{B}{\text{Red}}) = 2/26$ → Conditional

Conditional Probability vs. Joint Probability and Marginal Probability

P(A/B)

- **Conditional probability**: p(A|B) is the probability of event A occurring, **given that** event B occurs. For example, given that you drew a red card, what's the probability that it's a four (p(four|red))=2/26=1/13. So out of the 26 red cards (given a red card), there are two fours so 2/26=1/13.

P(A)

- **Marginal probability**: the probability of an event occurring (p(A)) in isolation. It may be thought of as an unconditional probability. It is not conditioned on another event. Example: the probability that a card drawn is red (p(red) = 0.5). Another example: the probability that a card drawn is a 4 (p(four)=1/13).

$P(4) = \frac{4}{52} = \frac{1}{13}$

P(AB)

- **Joint probability**: p(A ∩B). Joint probability is that of event A **and** event B occurring. It is the probability of the intersection of two or more events. The probability of the intersection of A and B may be written p(A ∩ B). Example: the probability that a card is a four and red =p(four and red) = 2/52=1/26. (There are two red fours in a deck of 52, the 4 of hearts and the 4 of diamonds).

$P(4/R)$

Dependent

$P(H/T)$

Independent

$P(\underbrace{\text{Fever}}_{S_2} / \underbrace{\text{Loss of Smell}}_{S_1})$

$P(H/T)$
$P(T/H)$] Same

Independent Events

Events can be "Independent", meaning each event is **not affected** by any other events.

Example: Tossing a coin.

Each toss of a coin is a perfect isolated thing.

What it did in the past will not affect the current toss.

The chance is simply 1-in-2, or 50%, just like ANY toss of the coin.

So each toss is an **Independent Event**.

Conditional Prob: How to handle dependent events

$P(X/Y)$

Dependent Events

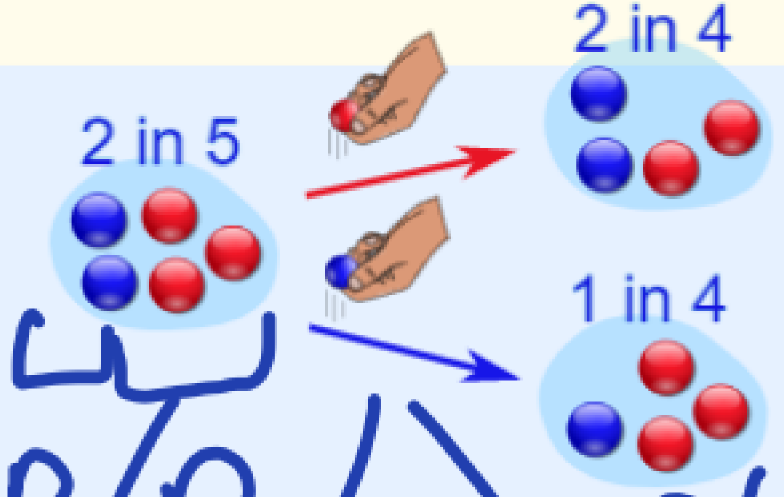
But events can also be "dependent" ... which means they **can be affected by previous events** ...

Example: Marbles in a Bag

2 blue and 3 red marbles are in a bag.

What are the chances of getting a blue marble?

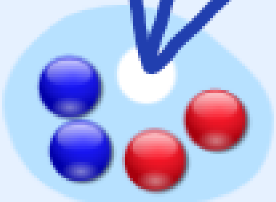
The chance is **2 in 5**



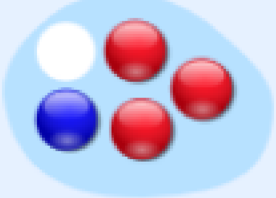
But after taking one out the chances change!

$P(B/R) = 2/4$

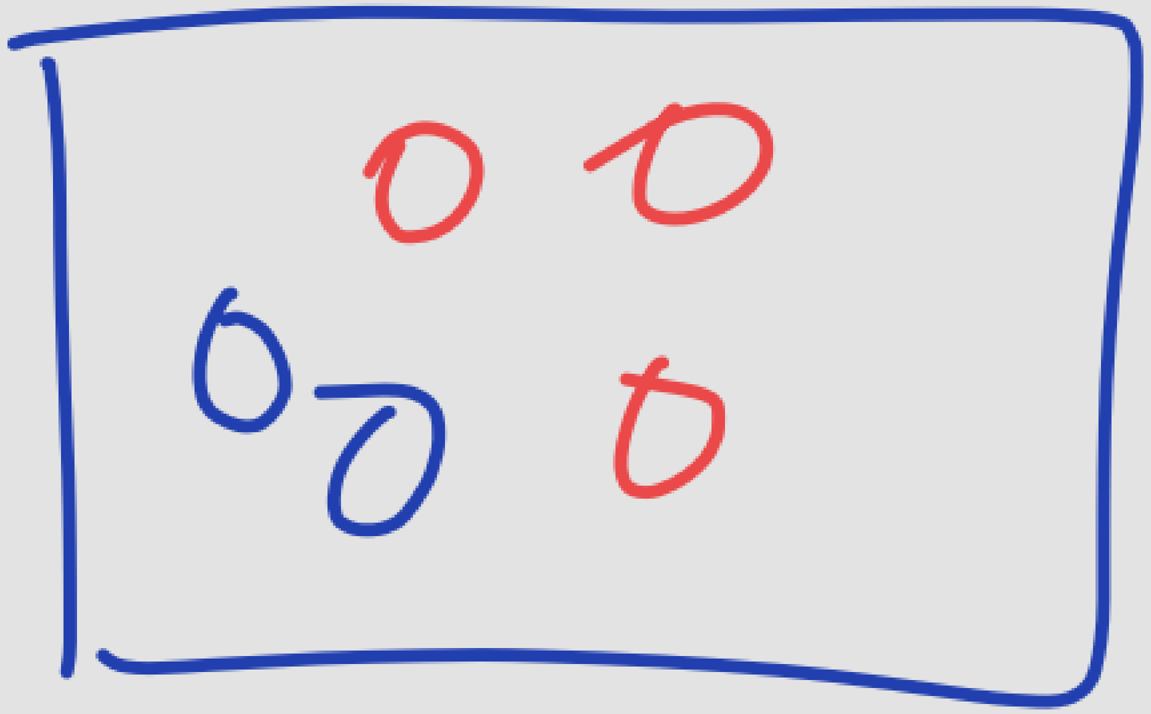
So the next time:



if we got a **red** marble before, then the chance of a blue marble next is **2 in 4**



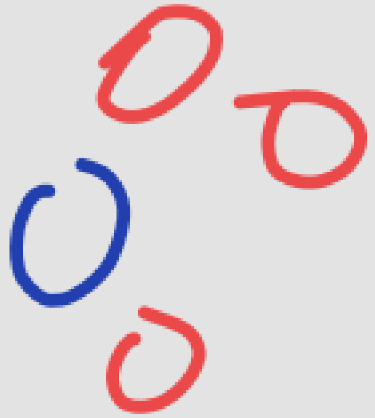
if we got a **blue** marble before, then the chance of a blue marble next is **1 in 4**



1st Ball = Red

$P(\text{2nd Ball} = R / R) = 2/4 = 1/2$

1st Ball = B



$P(\text{2nd Ball} = R / B) = 1/4$

Bayes Classifier

- **Used for modelling probabilistic relationships between attribute set and the class variable.**
- A probabilistic framework for solving classification problems
- Conditional Probability:

$$P(Y \mid X) = \frac{P(X,Y)}{P(X)}$$

$$P(X \mid Y) = \frac{P(X,Y)}{P(Y)}$$

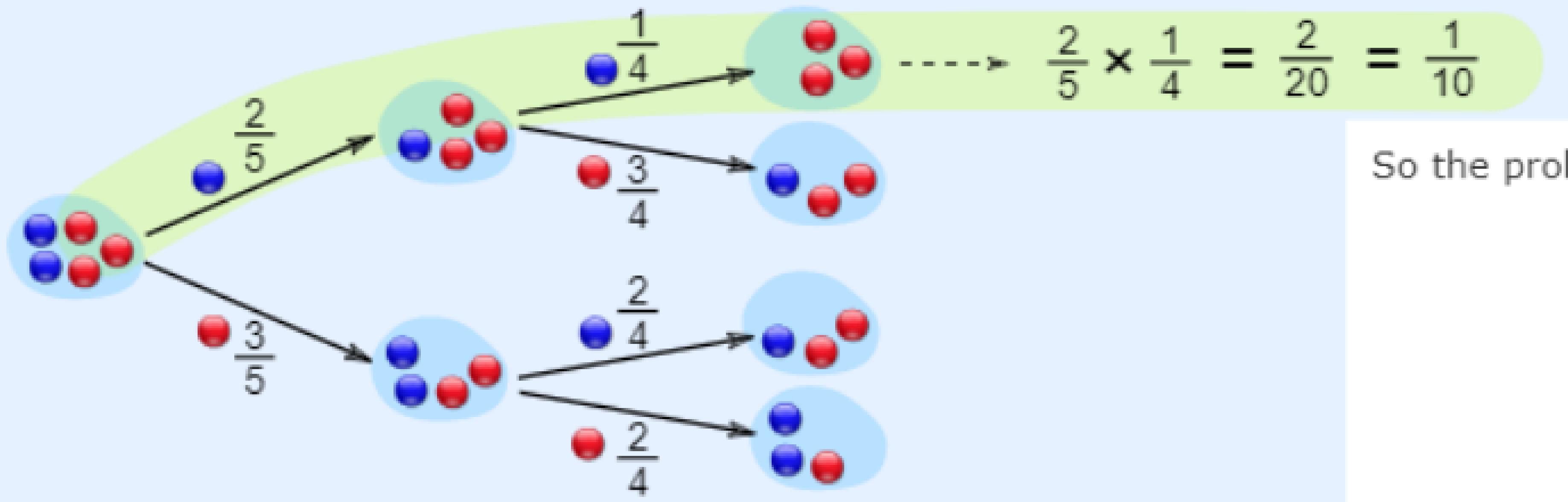
$$P(X,Y) = P(X/Y) \cdot P(Y)$$

- Bayes theorem:

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)}$$

Now we can answer questions like **"What are the chances of drawing 2 blue marbles?"**

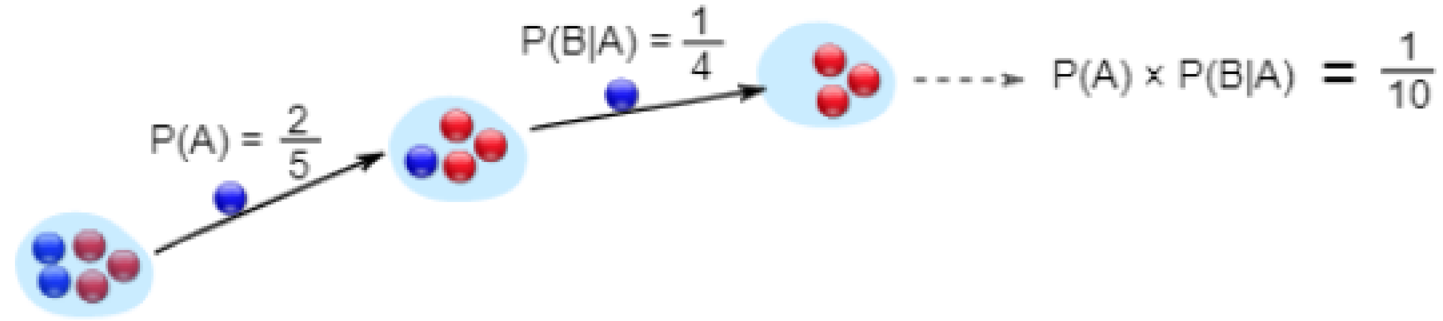
Answer: it is a **2/5 chance** followed by a **1/4 chance**:



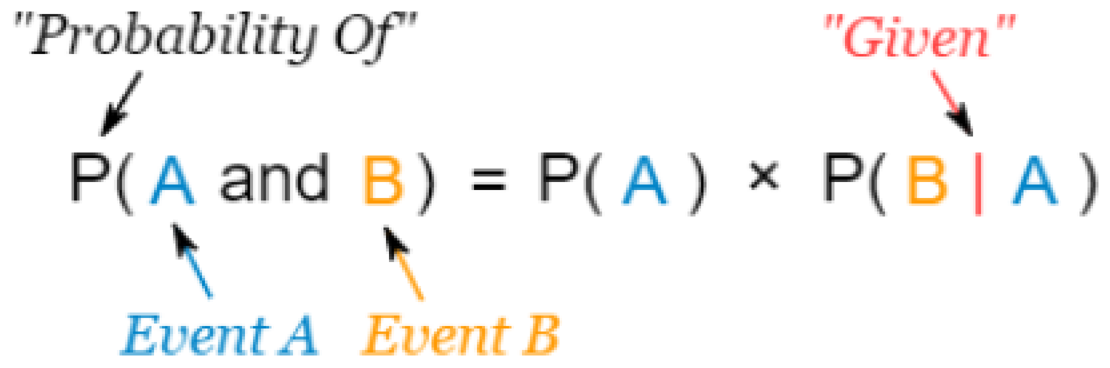
Did you see how we multiplied the chances? And got 1/10 as a result.

The chances of drawing 2 blue marbles is 1/10

So the probability of getting **2 blue marbles** is:



And we write it as



$$P(R,B) = P(R) * P(B/R)$$

$$P(R,B) = P(B) * P(R/B)$$

$$P(Y/X) = \frac{P(X/Y)\,P(Y)}{P(X)}$$

Posterior Prob.

Class Prior Probability

Class Conditional Prob.

Predictor prior probability

(Bayes theorem is used to find posterior prob. from prior prob.)

Train D.S.

X	Y
Sunny	Yes
Windy	NO
Sun	yes
Sunny	Yes

$P(Y) = P(Y = yes) = 3/4$

$P(Y = no) = 1/4$

$P(X = Sunny) = 3/4$

$P(Y = Yes / X = Windy) =$

Example of Bayes Theorem

- 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?

Y = Yes : Patient has meningitis

Y = No : ——— doesn't have ———

X → Stiff neck

P(Y = Yes / X = Stiff Neck)

$$P(M/S) = \frac{P(S/M) * P(M)}{P(S)} = 0.0002$$

$P(S/M) = 0.5$, $P(M) = 1/50,000$, $P(S) = 1/20$

Consider a football game between two rival teams, say team A and team B. Suppose team A wins 65% of the time and team B wins the remaining matches. Among the games won by team A, only 30% of them comes from playing at team B's football field. On the other hand, 75% of the victories for team B are obtained while playing at home. If team B is to host the next match between the two teams, what is the probability that it will emerge as the winner?

$P(Y=B / X=B)$

X: Team who will host the match

Y: Team who will win the match

$P(Y=A) = 0.65$

$P(Y=B) = 1 - P(Y=A) = 0.35$

$P(X=B / Y=A) = 0.3$

$P(X=B / Y=B) = 0.75$

$$P(X/Y) = \frac{P(X,Y)}{P(Y)}$$

$$P(Y=B/X=B) = \frac{P(X=B/Y=B) * P(Y=B)}{P(X=B)}$$

$$= \frac{0.75 * 0.35}{P(X=B, Y=A) + P(X=B, Y=B)}$$

$$P(Y=A)\,P(X=B/Y=A) + P(Y=B)\,P(X=B/Y=B)$$

card

$$P(X_1 = 4) = 4/52$$

$$P(X_1 = 4,\ X_2 = \text{Red}) + P(X_1 = 4,\ X_2 = \text{Bl}$$

$$\frac{2}{52} + \frac{2}{52} = \frac{4}{52}$$

5.3.2 Using the Bayes Theorem for Classification

Before describing how the Bayes theorem can be used for classification, let us formalize the classification problem from a statistical perspective. Let $\mathbf{X}$ denote the attribute set and Y denote the class variable. If the class variable has a non-deterministic relationship with the attributes, then we can treat $\mathbf{X}$ and Y as random variables and capture their relationship probabilistically using $P(Y|\mathbf{X})$. This conditional probability is also known as the **posterior probability** for Y, as opposed to its **prior probability**, $P(Y)$.

Using Bayes Theorem for Classification

- Consider each attribute and class label as random variables
- Given a record with attributes (X_1, X_2,…, X_d)
  - Goal is to predict class Y
  - Specifically, we want to find the value of Y that maximizes $P(Y|X_1, X_2,\ldots, X_d)$
- Can we estimate $P(Y|X_1, X_2,\ldots, X_d)$ directly from data?

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

Home owner

Annual Income

Y

Default Borrower (Defaulter or not)

Pay back the loan

$P(Y=\text{Yes}/X_1, X_2, X_3)$

$P(Y=\text{No}/X_1, X_2, X_3)$

Loan Default Problem

$(X_1 X_2 X_3) \rightarrow Y$

Using Bayes Theorem for Classification

- Approach:
  - compute posterior probability $P(Y \mid X_1, X_2, \ldots, X_d)$ using the Bayes theorem

$$\boxed{P(Y \mid X_1 X_2 \ldots X_n) = \frac{P(X_1 X_2 \ldots X_d \mid Y) P(Y)}{P(X_1 X_2 \ldots X_d)}}$$

  - *Maximum a-posteriori*: Choose Y that maximizes $P(Y \mid X_1, X_2, \ldots, X_d)$

$P(Y/X)$

  - Equivalent to choosing value of Y that maximizes $P(X_1, X_2, \ldots, X_d \mid Y)\, P(Y)$

- How to estimate $P(X_1, X_2, \ldots, X_d \mid Y)$?

Home Owner

Classify → Yes / No

Given a Test Record:

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

MS

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

- Can we estimate
  P(Evade = Yes | X) and P(Evade = No | X)?

  In the following we will replace
  Evade = Yes by Yes, and
  Evade = No by No

$$0.3 = P(Yes/X) = \frac{P(X/yes)\,P(Yes)}{P(X)}$$

$$0.42 = P(No/X) = \frac{P(X/No)\,P(No)}{P(X)}$$

$P(\text{Yes}) = 3/10$

$P(\text{No}) = 7/10$

$P(X/\text{Yes})$

$P(X/\text{No})$

Conditional

Prob

Class

$P(R = \text{No}, \text{Divorced}, 120K)$

NO.

$$P(X_1, X_2) = P(X_1) \cdot P(X_2)$$

H , T

$$P(X_1, X_2 / y) = P(X_1 / y) \cdot P(X_2 / y)$$

Problem: How to estimate class conditional probability?

$P(X/Yes)$ and $P(X/No)$

Answer: Naive Bayes

Class Conditional independence

$$P(X_1, X_2 / Y) = P(X_1 / Y) \cdot P(X_2 / Y)$$

Arm length

Reading Skills

Age

- It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors.
- • A Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.
- **Naive Bayesian** classification is **called** naive because it assumes class conditional independence.

$P(X/Y)$

Eg: $P(X_1, X_2, X_3 / Y)$

Color Shape Size Fruit

Red Round 3in Apple

Independent Predictor

$$P(X_1, X_2, X_3 / y)$$

$$= P(X_1/y) * P(X_2/y) * P(X_3/y)$$

Naive Bayes Classifier

- Assume independence among attributes X_i when class is given:
  - $P(X_1, X_2, \ldots, X_d | Y_j) = P(X_1 | Y_j)\, P(X_2 | Y_j) \ldots P(X_d | Y_j)$
  - Now we can estimate $P(X_i | Y_j)$ for all X_i and Y_j combinations from the training data
  - $\boxed{\text{New point is classified to } Y_j \text{ if } P(Y_j)\, \Pi\, P(X_i | Y_j) \text{ is maximal.}}$

$$P(Y=0/X) = \frac{\Pi_i P(X_i/Y=0) * P(Y=0)}{P(X)}$$

Naive Bayes Classifier is based upon two important principles: the assumption that all features are independent (the "Naive" part of the classifier) and Bayes' theorem

$$P(Y=1/X) = \frac{\Pi_i P(X_i/Y=1) * P(Y=1)}{P(X)}$$

Given a Test Record:

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

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

- Can we estimate
  P(Evade = Yes | X) and P(Evade = No | X)?

  In the following we will replace
  - Evade = Yes by Yes, and
  - Evade = No by No

$= P(R=No/Yes) * P(D/Y) * P(120/Y)$

$3/10$

$$P(Yes/X) = P(X/Yes) * P(Yes)$$

$$P(No/X) = P(X/No) * P(No)$$

$7/10$

Naïve Bayes on Example Data

Given a Test Record:

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

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(X | Yes) =

P(Refund = No | Yes) x → $\frac{3}{3} = 1$

P(Divorced | Yes) x → 1/3

P(Income = 120K | Yes)

P(X | No) =

P(Refund = No | No) x → 4/7

P(Divorced | No) x → 1/7

P(Income = 120K | No)

Estimate Probabilities from Data

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

- $P(y)$ = fraction of instances of class y
  - e.g., $P(No) = 7/10$,
    $P(Yes) = 3/10$
- For categorical attributes:

  $$P(X_i = c | y) = n_c / n$$

  - where $|X_i = c|$ is number of instances having attribute value $X_i = c$ and belonging to class y
  - Examples:

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

Example of Naïve Bayes Classifier

Given a Test Record:

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

Naïve 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) = 2/3
P(Marital Status = Divorced | Yes) = 1/3
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

- P(X | No) = P(Refund=No | No)
  × P(Divorced | No)
  × P(Income=120K | No)
  = 4/7 × 1/7 × 0.0072 = 0.0006
- P(X | Yes) = P(Refund=No | Yes)
  × P(Divorced | Yes)
  × P(Income=120K | Yes)
  = 1 × 1/3 × 1.2 × 10^{-9} = 4 × 10^{-10}

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

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

=> Class = No

P(No/X)

* 7/10

* 3/10

P(Yes/X)

P(No/X) = P(X/No) * P(No)

Estimate Probabilities from Data

[60, 220]
4 bins

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

- Normal distribution:

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

  - One for each (X_i, Y_j) pair
- For (Income, Class=No):
  - If Class=No
    - sample mean = 110
    - sample variance = 2975

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

Slide Included only for additional Reading.

Categorical Discretize qcut (4) bins

Estimate Probabilities from Data

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

- Normal distribution:

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

  – One for each (X_i, Y_i) pair

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

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

[60, 220]

[60, 100]

[101, 140]

[141, 180]

[181, 220]

120

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Compute class prior prob

1) $P(Y=Yes)$

$P(Y=No)$

Compute all conditional prob used while computing posterior prob:

$P(Outlook=Rainy / PG=Yes)$

Likelihood Table

		Play Golf	
		Yes	No
Outlook	Sunny	3/9	2/5
	Overcast	4/9	0/5
	Rainy	2/9	3/5

$P(S/Y)$

$P(X/Y)$

$P(Rainy/Yes)$ $P(Rainy/No)$

Link: https://wv

Frequency Table

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3

Likelihood Table

		Play Golf	
		Yes	No
Outlook	Sunny	3/9	2/5
	Overcast	4/9	0/5
	Rainy	2/9	3/5

		Play Golf	
		Yes	No
Humidity	High	3/9	4/5
	Normal	6/9	1/5

		Play Golf	
		Yes	No
Temp.	Hot	2/9	2/5
	Mild	4/9	2/5
	Cool	3/9	1/5

		Play Golf	
		Yes	No
Windy	False	6/9	2/5
	True	3/9	3/5

Class-Conditional Prob.

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Train Data

Link: https://www.saedsayad.com/modeling.htm

Ques Determine class Label for following test record

2

X = <Rainy, Cool, High, True>

P(Yes/X)

P(No/X)

H/W

In this example we have 4 inputs (predictors). The final posterior probabilities can be standardized between 0 and 1.

Outlook	Temp	Humidity	Windy	Play
Rainy	Cool	High	True	?

$$P(Yes \mid X) = P(Rainy \mid Yes) \times P(Cool \mid Yes) \times P(High \mid Yes) \times P(True \mid Yes) \times P(Yes)$$

$$P(Yes \mid X) = 2/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14 = 0.00529 \rightarrow 0.2 = \frac{0.00529}{0.02057 + 0.00529}$$

$$P(No \mid X) = P(Rainy \mid No) \times P(Cool \mid No) \times P(High \mid No) \times P(True \mid No) \times P(No)$$

$$P(No \mid X) = 3/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14 = 0.02057 \rightarrow 0.8 = \frac{0.02057}{0.02057 + 0.00529}$$

Now that we have all the values, let's plug them in the equations

P(X | Class = "Yes)

= P({Refund = "No" , Marital Status = "Married" , Taxable Income = 120K} | Class = "Yes")

= P(Refund = "No" | Class = "Yes") * P(Marital Status = "Married" | Class = "Yes") * P(Taxable Income = 120K | Class = "Yes")

= (3/3) * (0/3) * (1.2 * 10^-9)

= 0

Zero Frequency problem

P(X | Class = "No")

= P({Refund = "No" , Marital Status = "Married" , Taxable Income = 120K} | Class = "No")

= P(Refund = "No" | Class = "No") * P(Marital Status = "Married" | Class = "No") * P(Taxable Income = 120K | Class = "No")

= (4/7) * (4/7) * (0.0072)

= 0.0023

The final step

P(Class = "Yes" | X)
= P(X | Class = "Yes") * P(Class = "Yes")
= 0 * (3/10)
= 0

P(Class = "No" | X)
= P(X | Class = "No") * P(Class = "No")
= 0.0023 * (7/10)
= 0.00161

As we can see the probability of Class = "No" is more, hence the unknown data will be assigned the Class = "No".

Issues with Naïve Bayes Classifier

- If one of the conditional probabilities is zero, then the entire expression becomes zero
- Need to use other estimates of conditional probabilities than simple fractions
- Probability estimation:

original: $P(X_i = c|y) = \frac{n_c}{n}$

Laplace Estimate: $P(X_i = c|y) = \frac{n_c + 1}{n + v}$

$\text{m} - \text{estimate: } P(X_i = c|y) = \frac{n_c + mp}{n + m}$

n: number of training instances belonging to class y

n_c: number of instances with $X_i = c$ and $Y = y$

v: total number of attribute values that X_i can take

p: initial estimate of $(P(X_i = c|y))$ known apriori → $\frac{1}{m}$ (generally

m: hyper-parameter for our confidence in p

↳ No. of unique values

Practice Question

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

A: attributes

M: mammals

N: non-mammals

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

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

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

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

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

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

=> Mammals

Naïve Bayes (Summary)

- Robust to isolated noise points
- Handle missing values by ignoring the instance during probability estimate calculations
- Robust to irrelevant attributes
- Redundant and correlated attributes will violate class conditional assumption
  - Use other techniques such as Bayesian Belief Networks (BBN)