II Module 3 Bayesian Decision Theory.

probability theory is used.

Building a model with data leads to a random process. process is deterministic but do not have access to complete knowledge about it, we modeled it as random ~~process~~ probability theory is used to analyze it.

Ex: Tossing a coin is a random process.

Result of tossing a coin is $\in$ {Head, Tails}

Random var $X \in \{1, 0\}$

Bernoulli principle : $p(X=1) = p_0^{x}(1-p_0)^{1-x}$

$p(X=1) = p_0$

$p(X=0) = 1 - p(X=1) = 1 - p_0$

Dr. Sunanda
ISE

prediction is heads $p_0 > 0.5$

Error $= 1 - p_0 = 1 - 0.5 = \underline{0.5}$. it is minimum.

If this is fair coin predicting $p_0 = 0.5$, $\therefore$ we can choose heads all the time.

Sample X, where coin tossing example outcomes of the past N tosses. we can estimate p_0 (Refers to distribution)

$$\therefore p_0 = \frac{\#\{\text{tosses with outcome heads}\}}{\#\text{ tosses}} \quad (1)$$

$x^t = 1$, if outcome is head

$\therefore$ Heads if $p_0 > 0.5$
Tails otherwise

Ex: Sample {heads, heads, heads, tail, heads, tail, tail, heads, heads}, we have

$X = \{1, 1, 1, 0, 1, 0, 0, 1, 1\}$ and the estimate is

$$p_0 = \frac{\sum_{t=1}^{N} x^t}{N} = \frac{6}{9}.$$

3.2 Classification

Credit scoring: inputs are income and savings. output is low risk v/s high risk.

X_1 is income

X_2 is saving.

using Bernoulli random variable 'C' conditioned on the observables $X = [X_1, X_2]^T$

where $C = 1 \Rightarrow$ indicates high risk customer

$C = 0 \Rightarrow$ indicates low risk customer

If we know $P(C \mid X_1, X_2)$ when new application arrives with $X_1 = x_1$, $X_2 = x_2$,

$$\text{Choose } \begin{cases} C = 1 & \text{if } P(C=1 \mid x_1, x_2) > 0.5 \\ C = 0 & \text{otherwise} \end{cases}$$

or equivalently

$$\text{Choose } \begin{cases} C = 1 & \text{if } P(C=1 \mid x_1, x_2) > P(C=0 \mid x_1, x_2) \\ C = 0 & \text{otherwise.} \end{cases}$$

The probability of error is $1 - \max(P(C=1 \mid x_1, x_2), P(C=0 \mid x_1, x_2))$

Using Bayes Rule,

$$p(C/x) = \frac{p(c)\,p(x|c)}{p(x)}$$

prior probability :

$p(c=1)$ .. customer has high risk base.

it is the knowledge what we have,

$p(c=0) + p(c=1) = 1$.

~~Lik~~ Likelihood :

$p(x/c)$ is called class likelihood, & is the conditional probability, that an event belong to 'c' as associated observation value x.

$p(x_1, x_2 | c=1)$ => probability that high risk customer $X_1 = x_1$, $X_2 = x_2$, where data tells us.

$p(x)$ is the 'evidence', is the marginal probability that an observation x is seen, regardless of whether it is a +ve or -ve example.

$$p(x) = \sum_{c} p(x,c) = p(x|c=1)\,p(c=1) + p(x|c=0)\,p(c=0)$$

Combining priori and data tells us, calculate posterior probability. $p(c/x)$. after seeing observation x

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

$p(c=0/x) + p(c=1/x) = 1$

Dr. Sunanda Dixit
ISE, DSCE

Baye's Rule

Bayes' theorem is a way to figure out Conditional probability

Conditional probability is the probability of an event happening, given that it has some relationship to one or more other events

Ex: Car parking — time of day you park
— where you park
— What conventions are going on at any time

It (Bayes) gives you the actually probability of an event, given information about the test

p(Event 1) : prior probability
p(Event 2) : Evidence
p(Event 2 / Event 1) : Likelihood
p(Event 1 / Event 2) : posterior probability

$$\text{posterior probability} = \frac{\text{prior probability} \times \text{likelihood}}{\text{Evidence}}$$

$$p(\text{Event } 1 / \text{Event } 2) = \frac{p(\text{Event } 1) \times p(\text{Event } 2 / \text{Event } 1)}{p(\text{Event } 2)}$$

Bayes Theorem.

probability previously learned.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

1 $$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Prob 1 What is the probability of two girls given at least one girl?

$P(2G \mid \text{at least } 1G)$

GG . GB . BG . BB

$$= \frac{P(1G/2G) \cdot P(2G)}{P(1G)}$$

$$= \frac{1 \cdot 1/4}{3/4} = \boxed{1/3}$$ conditional probability.

Prob 2

3.3 Losses and Risks

Ex: low risk applicant increases profit
high risk applicant decreases loss

C_i - class decision to assign the input
α_i - decision action
λ_{ik} - loss incurred for taking action α_i
C_k - input actually belongs to C_k

$$R(\alpha_i|x) = \sum_{k=1}^{K} \lambda_{ik} P(C_k|x) \quad \text{—} (1)$$

we choose action with minimum risk

Choose α_i if $R(\alpha_i|x) = \min_k R(\alpha_k|x)$ —(2)

K actions $\alpha_i \, i = 1, 2 \ldots k$,
Where α_i is the action of assigning x to C_i
Special case of 0/1 loss where

$$\lambda_{ik} = \begin{cases} 0 & \text{if } i = k \\ 1 & \text{if } i \neq k \end{cases} \quad \text{—} (3)$$

all correct decisions have no loss
all errors are equally costly.
The risk of taking action α_i is

$$R(\alpha_i|x) = \sum_{k=1}^{K} \lambda_{ik} P(C_k|x) \quad \text{—} (4)$$

$$= \sum_{k \neq i} P(C_k|x) \quad \text{—} (5)$$

$$= 1 - P(C_i|x) \quad \text{—} (6)$$

$\because \sum_k P(C_k|x) = 1$. Thus to minimise risk choose probable class.

Define an additional action reject or doubt.

α_{K+1}, with α_i $i = 1 \ldots K$

possible loss function is

$$\lambda_{ik} = \begin{cases} 0 & \text{if } i = k \\ \lambda & \text{if } i = K+1 \\ 1 & \text{otherwise} \end{cases} \quad — (7)$$

where $0 < \lambda < 1$ is the loss incurred for choosing the $(K+1)^{st}$ action of reject.

Then risk of reject is,

$$R(\alpha_{K+1} | x) = \sum_{K=1}^{K} \lambda P(C_K | x) = \lambda \quad — (8)$$

The risk of choosing class C_i is

$$R(\alpha_i | x) = \sum_{K \neq i} P(C_K | x) = 1 - P(C_i | x) \quad — (9)$$

The optimum decision rule is to

Choose C_i if $R(\alpha_i | x) < R(\alpha_K | x)$ for all $K \neq i$ and $R(\alpha_i | x) < R(\alpha_{K+1} | x)$

reject if $R(\alpha_{K+1} | x) < R(\alpha_i | x)$, $i = 1, 2 \ldots K$

from eq. (7).

Choose C_i if $P(C_i | x) > P(C_K | x)$ for all $K \neq i$ and $P(C_i | x) > 1 - \lambda$

reject otherwise.

3.4 Discriminant Functions

A set of discriminant functions can be implemented classification

$g_i(x), \; i = 1 \ldots K$

Choose C_i if $g_i(x) = \max_K g_K(x)$

Represent using Bayes Classifier

$g_i(x) = -R(\alpha_i|x).$

<u>Max</u> discriminant function corresponds to minimum conditional risk.

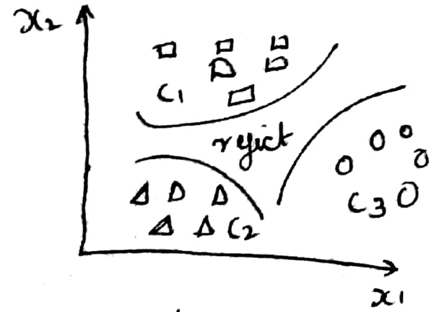
Using 0/1 loss function

$g_i(x) = P(C_i|x)$

or ignoring the common normalise term $p(x)$,

$g_i(x) = p(x|C_i)\, p(C_i)$

This divides the feature space into 'K' decision regions $R_1, R_2 \ldots R_K$ where $R_i = \{x \mid g_i(x) = \max_K g_K(x)\}$



<u>ex</u> decision regions & boundaries.

When there are 2 classes, we can define a single discriminant

$g(x) = g_1(x) - g_2(x)$

& we choose $\begin{cases} C_1 & \text{if } g(x) > 0 \\ C_2 & \text{otherwise.} \end{cases}$

+ve <u>ex</u> C_1

-ve <u>ex</u> C_2

3.5 Association rules

An association rule is the implication of the form $X \rightarrow Y$.

where X is antecedent,
Y is Consequent of the rule.

Ex Basket analysis

In association 3 measures are calculated,

Support Support of the association rule $X \rightarrow Y$:

$$\text{Support}(X, Y) \equiv P(X, Y) = \frac{\#\{\text{Customer who bought } X \,\&\, Y\}}{\#\{\text{Customers}\}}$$

Confidence Confidence of the association rule $X \rightarrow Y$:

$$\text{Confidence}(X \rightarrow Y) \equiv P(Y|X) = \frac{P(X, Y)}{P(X)}$$

$$= \frac{\#\{\text{Customers who bought } X \,\&\, Y\}}{\#\{\text{Customers who bought } X\}}$$

Lift or interest

Lift is the $X \rightarrow Y$ $= \dfrac{P(X, Y)}{P(X)\,P(Y)} = \dfrac{P(Y|X)}{P(Y)}$

Apriori Algorithm

2 steps
(1) finding frequent items, which have enough support.
(2) converting them to rule with enough confidence by splitting the items into two, antecedent, consequen

Apriori algorithm for
→ (1) $\{X, Y, Z\}$ to be frequent. enough support all its subsets $\{X, Y\}$ $\{X, Z\}$ and $\{Y, Z\}$

→ (2) Split k items into two – antecedent
– Consequent

Hidden variables. Ex: "Baby at home" ⇒ baby food, milk, diapers

Finding of frequent item set.

2 transanctions I_1 & I_2

$I_1 = \{1, 2, 3\}$ - item sets

$I_2 = \{2, 3\}$

$\{2, 3\}$ repeat are known as frequent item sets.

Apriori } used for frequent item set.
FP growth }

Apripri — Candidate Generation

FP growth } only
FP tree. } Support.

Support — Support Count data set repeatation

Confidence. — i.e. also meant in apriori

Ex: Transaction item sets

Transaction	item sets
I_1	A, B, C.
I_2	A, C
I_3	A, D
I_4	B, E, F.

Dr. Sunanda Dixit
ISE

minimum Support 50%.
minimum Confidence 50%.

Number is required Support Count $= \frac{50}{100} \times 4 = 2$ — 4 transaction

Generate Candidate.

C_1 =

items	Support
{A}	3
{B}	2
{C}	2
{D}	1
{E}	1
{F}	1

Set 'A' how many times items is present That Count is Support Count.

Eliminate D, E, F not equal to Support Count 2

L_1 =

items	Support
{A}	3
{B}	2
{C}	2

C_2 :

items	Support
{A, B}	1
{B, C}	1
{A, C}	2.

how many times A, B, B, C, A, C is occured.

L_2 =

items	Support.
{A,C}	2

C_1, Single Set, C_2 = 2 Set C_3 = 3 Sets.

Draw rule.

Association ~~rule~~	Support	Confidence	Confidence %
A → C	2	2/3 = 0.66	66 %
C → A.	2	2/2 = 1	100 %.

$A \to C = \frac{\text{Support}}{\text{occurance of A}}$

occurance of item actions table. A & C.

$\text{Confidence} = \frac{\text{Support}}{\text{occ. of A.}}$

how my times the item is exis ting in the table

50% > 66%.
100%.

final Rule : A → C
C → A

7) Given the following data of transactions at a shop. Calculate the support and confidence values of milk → bananas, bananas → milk, milk → chocolate, and chocklate → milk.

Soln:

Transaction	Items in basket
1	milk, banana's, chocolate
2	milk, chocklate
3	milk, bananas
4	chocolate
5	chocolate
6	milk, chocolate.

Solution:

milk → bananas : Support = 2/6 = 33%, Confidence = 2/4 = 50%

bananas → milk : Support = 2/6 = 33%, Confidence = 2/2

milk → chocolate : Support = 3/6 = 50%, confidence = 3/4

Chocolate → milk : Support = 3/6 = 50%, Confidence = 3/5

① Support $x \to y = p(x, y) = \frac{\# \{\text{Customers who bought } x \,\&\, y\}}{\# \text{Customers}}$

② Confidence $(x \to y) \equiv p(y/x) = \frac{\# \text{ customers who bought } x \,\&\, y}{\# \{\text{customers who bought } x\}}$

Inference: Though only half of the people who buy milk buy banana's too, anyone who buys banana's also buys milk.

FP growth Algorithm

Pattern Frame growth.

\ Han.

Tid	item Sets.
1	f, a, c, d, g, m, p
2	a, b, c, f, l, m, o
3	b, f, h, o.
4	b, k, c, p
5	a, f, c, l, p, m, n.

minimum Support 3.

[a, b, c, d, f, g, k, l, m, n, o, p]

First → create a table.

Item	Support
a	3
b	3
c	4
d	1
f	4
g	1
k	1
l	2
m	3
n	1
o	2
p	3

greater Count first

create a pattern most occuring first

item	Support
f	4
c	4
a	3
b	3
m	3
p	3

pattern : f, c, a, b, m, p

Same pattern used.

eliminate ← less than Support 3

next order item.

Tid	item sets	ordered items
1	f, a, c, ~~d~~, ~~g~~, m, p	f, c, a, m, p
2	a, b, c, f, ~~l~~, m, ~~o~~	f, c, a, b, m
3	b, f, k, ~~o~~	f, b
4	b, K, c, P	c, b, p.
5	a, f, c, ~~l~~, p, m, ~~n~~	f, c, a, m, p

min Support : 3

[a. b. c. d. f. g. k. l, m, n, o, p]

f, c, a, b, m, p → new pattern

Calculate 'f' how my times occured in table

f : 4.

c : 4

a : 3

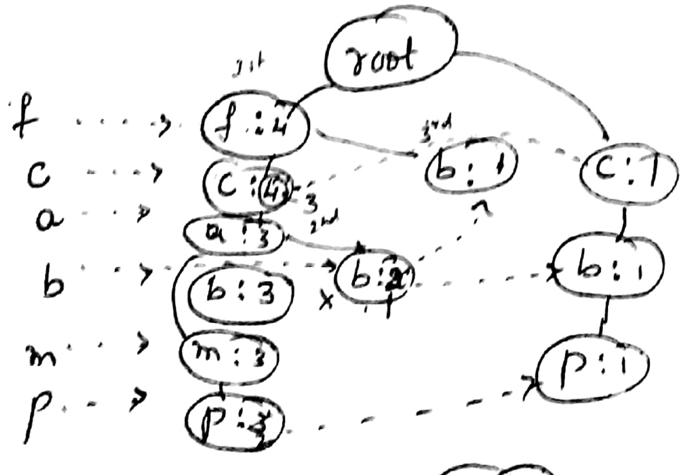
b : 2

m : 3

p : 3

This used for tree draw FP tree. using ordered item

root node.



ordered item

f, c, a . m , p .

f, c, a, b, m

f , b

c , b, p.

f, c, a, m, p.