Data Mining :: Unit-3

(Classification – Naïve Bayes Classifier)

Er. Dinesh Baniya Kshatri
(Lecturer)

Department of Electronics and Computer Engineering
Institute of Engineering, Thapathali Campus

Thomas Bayes

Reverend Thomas Bayes (1701-1761),
studied logic and theology as an undergraduate student
at the University of Edinburgh from 1719-1722.

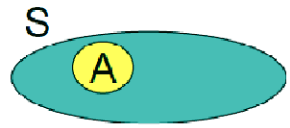
Background Material
(Sample Space and Events)

Consider an experiment

Sample space S:

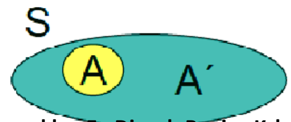
Example:
S={1,2,…,6} rolling a dice
S={head,tail} flipping a coin

Event A:



Example:
A={1,6} when rolling a dice

Complementary event A´:

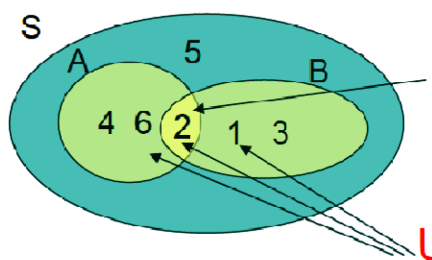


Example:
A´={2,3,4,5} rolling a dice

Background Material
(Probability Theory)

Example:
Rolling a dice

S={1,2,3,4,5,6}
A={2,4,6}
B={1,2,3}

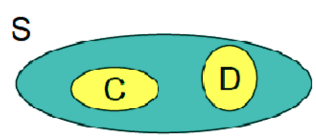


Intersection:
A∩B={2}

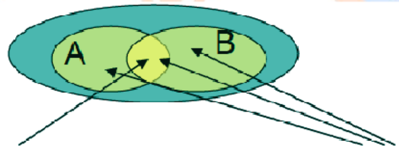
Union:
A∪B={1,2,3,4,6}

Disjoint events: C∩D = Ø

C={1,3,5} and D={2,4,6} are disjoint



Background Material
(Rules for Probabilities)



Intersection: $A \cap B$

Union: $A \cup B$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(B) = P(B \cap A) + P(B \cap A')$$

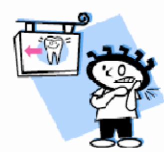
If A and B are disjoint: $P(A \cup B) = P(A) + P(B)$

In particular: $P(A) + P(A') = 1$

Background Material
(Joint Probability Distribution)

- **Probability assignment to all combinations of values of random variables (i.e. all elementary events)**

	toothache	¬ toothache
cavity	0.04	0.06
¬ cavity	0.01	0.89

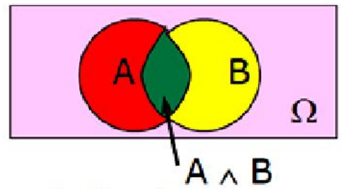


- **The sum of the entries in this table has to be 1**
- ***Every question about a domain can be answered by the joint distribution***

- **Probability of a proposition is the sum of the probabilities of elementary events in which it holds**
  - P(cavity) = 0.1 [marginal of row 1]
  - P(toothache) = 0.05 [marginal of toothache column]

Background Material
(Conditional Probability) – [1]

	toothache	¬ toothache
cavity	0.04	0.06
¬ cavity	0.01	0.89



- *P(cavity)=0.1* and *P(cavity ∧ toothache)=0.04* are both *prior* (unconditional) probabilities
- Once the agent has new evidence concerning a *previously unknown* random variable, e.g. Toothache, we can specify *a posterior* (conditional) probability e.g. *P(cavity | Toothache=true)*

$$P(a \mid b) = P(a \wedge b)/P(b)$$

[Probability of a with the Universe Ω restricted to b]

- *So P(cavity | toothache) = 0.04/0.05 = 0.8*

Background Material
(Conditional Probability) – [2]

- Definition of Conditional Probability:
  $P(a \mid b) = P(a \wedge b)/P(b)$
- Product rule gives an alternative formulation:
  $$P(a \wedge b) = P(a \mid b) * P(b)$$
  $$= P(b \mid a) * P(a)$$
- **Chain rule** is derived by successive application of product rule:

$$\begin{aligned} P(A,B,C,D,E) &= P(A|B,C,D,E)\ P(B,C,D,E) \\ &= P(A|B,C,D,E)\ P(B|C,D,E)\ P(C,D,E) \\ &= \ldots \\ &= P(A|B,C,D,E)\ P(B|C,D,E)\ P(C|D,E)\ P(D|E)\ P(E) \end{aligned}$$

Background Material
(Proof of Bayes' Theorem)

Let A and B be events such that $0 < P(A) < 1$ and $P(B) > 0$.

By definition, $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$. So: $P(A \cap B) = P(A \mid B)P(B)$.

Likewise, $P(B \cap A) = P(B \mid A)P(A)$.

Likewise, $P(B \cap \overline{A}) = P(B \mid \overline{A})P(\overline{A})$. (Note that $P(\overline{A}) > 0$.)

Note that $P(A \mid B)P(B) = P(A \cap B) = P(B \mid A)P(A)$. So,

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Furthermore,

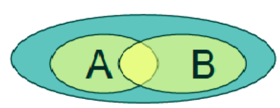
$$\begin{aligned} P(B) &= P((B \cap A) \cup (B \cap \overline{A})) = P(B \cap A) + P(B \cap \overline{A}) \\ &= P(B \mid A)P(A) + P(B \mid \overline{A})P(\overline{A}) \end{aligned}$$

So:
$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid \overline{A})P(\overline{A})}.$$

Background Material
(Summary of Bayes Rule)

Conditional probability for A given B:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{where } P(B) > 0$$



Bayes' Rule: $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

Rewriting Bayes' rule:

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

Background Material
(Bayes' Rule – Example)

Example: Lung disease & Smoking

According to "The American Lung Association" 7% of the population suffers from a lung disease, and 90% of these are smokers. Amongst people without any lung disease 25.3% are smokers.

Events:	**Probabilities:**
A: person has lung disease	P(A) = 0.07
B: person is a smoker	P(B|A) = 0.90
	P(B|A´) = 0.253

What is the probability that a smoker suffers from a lung disease?

$$P(A\mid B)=\frac{P(B\mid A)P(A)}{P(B\mid A)P(A)+P(B\mid A')P(A')}=\frac{0.9\cdot 0.07}{0.9\cdot 0.07+0.253\cdot 0.93}=0.211$$

Bayesian Spam Filtering

Problem: Suppose it has been observed empirically that the word "Congratulations" occurs in 1 out of 10 spam emails, but that "Congratulations" only occurs in 1 out of 1000 non-spam emails. Suppose it has also been observed empirically that about 4 out of 10 emails are spam.

Suppose we get a new email that contains "Congratulations".

Let C be the event that a new email contains "Congratulations".
Let S be the event that a new email is spam.

We have observed C. We want to know $P(S \mid C)$.

Bayesian Spam Filtering

Bayesian solution: By Bayes' Theorem:

$$P(S \mid C) = \frac{P(C \mid S)P(S)}{P(C \mid S)P(S) + P(C \mid \overline{S})P(\overline{S})}$$

From the "empirical probabilities", we get the estimates:

$P(C \mid S) \approx 1/10$; $P(C \mid \overline{S}) \approx 1/1000$;

$P(S) \approx 4/10$; $P(\overline{S}) \approx 6/10$.

So, we estimate that:

$$P(S \mid C) \approx \frac{(1/10)(4/10)}{(1/10)(4/10) + (1/1000) * (6/10)}$$

$$\approx \frac{.04}{.0406} \approx 0.985$$

Bayes' Rule & Disease Diagnosis – [1]

Likelihood Prior

$$P(a|b) = \frac{P(b|a) * P(a)}{P(b)}$$

Posterior Normalization

Useful for assessing diagnostic probability from causal probability:

$$P(Cause|Effect) = \frac{P(Effect|Cause) * P(Cause)}{P(Effect)}$$

Bayes' Rule & Disease Diagnosis – [2]

$$P(Disease \mid Symptom) = \frac{P(Symptom \mid Disease) * P(Disease)}{P(Symptom)}$$

Imagine:

- disease = TB, symptom = coughing
- *P(disease | symptom)* is different in TB-indicated country vs. USA
- *P(symptom | disease)* should be the same
- What about P(symptom)?
  - Use *conditioning* (next slide)

Importance of Conditioning

- *Idea:* Use *conditional probabilities* instead of joint probabilities
- $P(a) = P(a \wedge b) + P(a \wedge \neg b)$
  $= P(a \mid b) * P(b) + P(a \mid \neg b) * P(\neg b)$

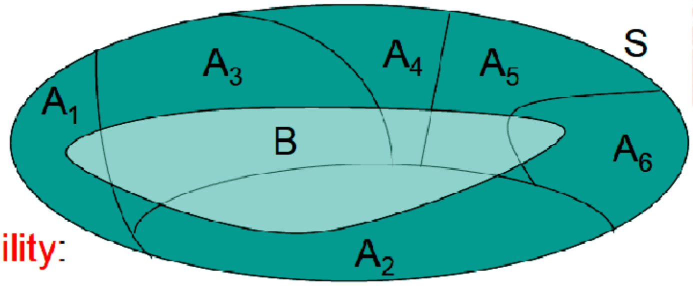
  Here:

  $P(symptom) = P(symptom \mid disease) * P(disease) + P(symptom \mid \neg disease) * P(\neg disease)$

- More generally: $P(Y) = \sum_z P(Y|z) * P(z)$

Bayes' Rule – Extended Version

$A_1, \ldots, A_k$ is a partitioning of S



Law of total probability:

$$P(B) = \sum_{i=1}^{k} P(B \mid A_i) P(A_i)$$

Bayes' formula extended:

$$P(A_r \mid B) = \frac{P(B \mid A_r) P(A_r)}{\sum_{i=1}^{k} P(B \mid A_i) P(A_i)}$$

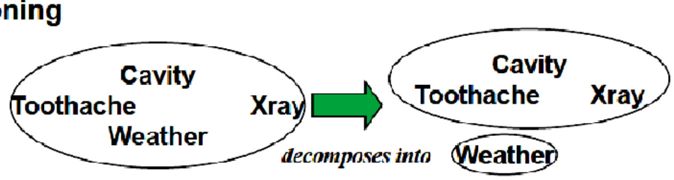
Estimating Joint Probabilities
(Maybe Infeasible)

- For $|D|$ diseases, $|S|$ symptoms where a person can have n of the diseases and m of the symptoms
  - —$P(s|d_1, d_2, \ldots, d_n)$ requires $|S|\ |D|^n$ values
  - —$P(s_1, s_2, \ldots, s_m)$ requires $|S|^m$ values

- **These numbers get big fast**
  - If $|S|$ =1,000, $|D|$=100, n=4, m=7
    - —$P(s|d_1, \ldots d_n)$ requires 1000*100^4 =10^{11} values (-1)
    - —$P(s_1 \ldots s_m)$ requires 1000^7 = 10^{21} values (-1)

Estimating Joint Probabilities
(Solution:- Independence)

- Random variables A and B are *independent* iff
  - $P(A \wedge B) = P(A) * P(B)$
  - equivalently: $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$
- *A and B are independent if knowing whether A occurred gives no information about B (and vice versa)*
- Independence assumptions are *essential* for efficient probabilistic reasoning



$$P(T, X, C, W) = P(T, X, C) * P(W)$$

- 15 entries (2^4-1) reduced to 8 (2^3-1 + 2-1)

Dependence: Example

Example:

	Employed	Unemployed	Total
Man	460	40	500
Woman	140	260	400
Total	600	300	900

$$P(\text{man}|\text{employed}) = \frac{460/900}{600/900} = 76.7\%$$

$$P(\text{man}) = 500/900 = 55.6\%$$

Conclusion: the two events “man” and “employed” are dependent.

Alternative to Complete Independence
(Conditional Independence) – [1]

- BUT *absolute* independence is rare
- Dentistry is a large field with hundreds of variables, none of which are independent. What to do?

- A and B are *conditionally independent given C* iff
  - $P(A \mid B, C) = P(A \mid C)$
  - $P(B \mid A, C) = P(B \mid C)$
  - $P(A \wedge B \mid C) = P(A \mid C) * P(B \mid C)$

- Toothache (T), Spot in Xray (X), Cavity (C)
  - None of these are independent of the other two
  - But ***T and X are conditionally independent given C***

Alternative to Complete Independence
(Conditional Independence) – [2]

- If I have a cavity, the probability that the XRay shows a spot doesn't depend on whether I have a toothache (and vice versa):
  $P(X|T,C) = P(X|C)$
- From which follows:
  $P(T|X,C) = P(T|C)$ and $P(T,X|C) = P(T|C) * P(X|C)$
- By the chain rule , given conditional independence:
  $$P(T,X,C) = P(T|X,C) * P(X,C) = P(T|X,C) * P(X|C) * P(C)$$
  $$= P(T|C) * P(X|C) * P(C)$$
- P(*Toothache, Cavity, Xray*) has $2^3 - 1 = 7$ independent entries
- Given conditional independence, chain rule yields
  $2 + 2 + 1 = 5$ independent numbers

Alternative to Complete Independence
(Conditional Independence) – [3]

- In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from ***exponential*** in *n* to ***linear*** in *n*.

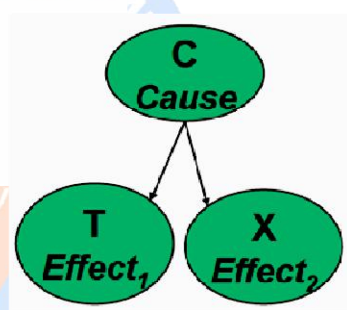
- ***Conditional independence is our most basic and robust form of knowledge about uncertain environments.***

Naïve Bayes Model

By Bayes Rule $P(C|T,X) = \frac{P(T,X|C)P(C)}{P(T,X)}$

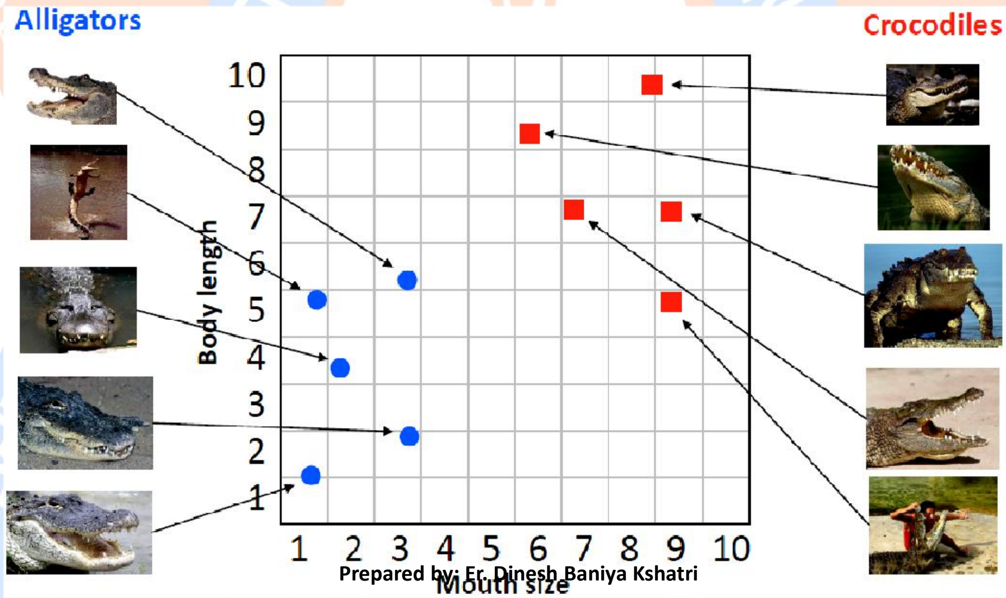
If T and X are *conditionally independent given C*:

$$P(C|T,X) = \frac{P(T|C)P(X|C)P(C)}{P(T,X)}$$



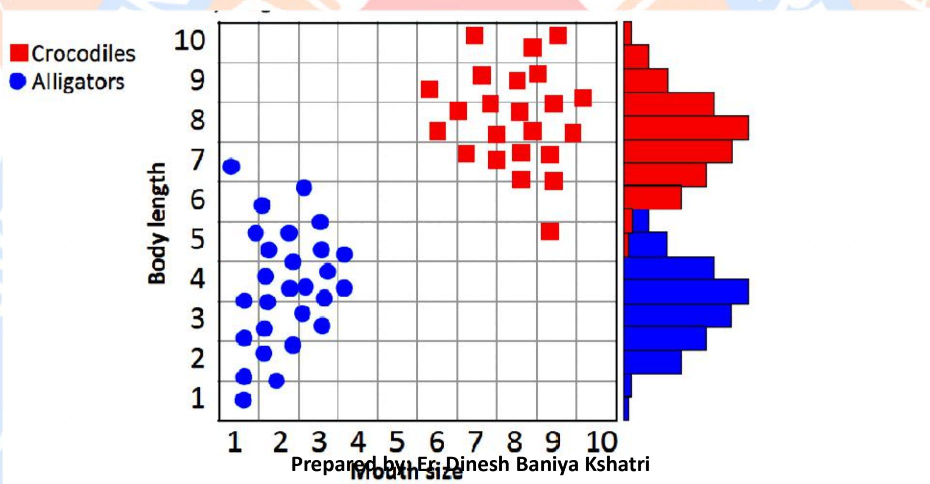
All effects assumed conditionally independent given Cause

Visual Intuition
(Naïve Bayes Model) – [1]



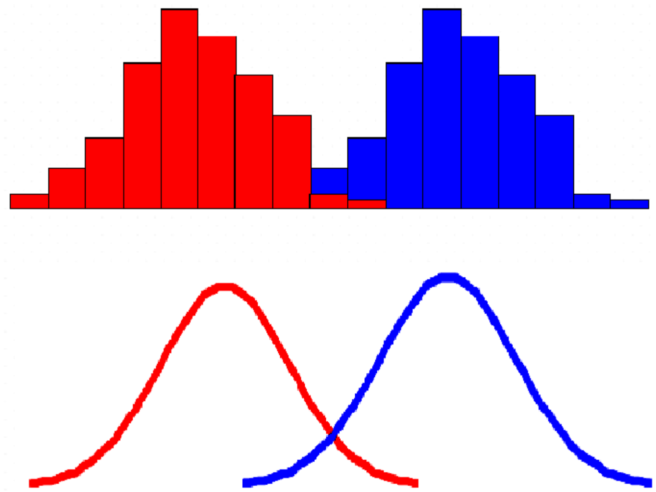
Visual Intuition
(Naïve Bayes Model) – [2]

- **Suppose we had lots of data. We could use that data to build a histogram. Below is one built for the body length feature**



Visual Intuition
(Naïve Bayes Model) – [3]

We can summarize these histograms as two normal distributions.



Visual Intuition
(Naïve Bayes Model) – [4]

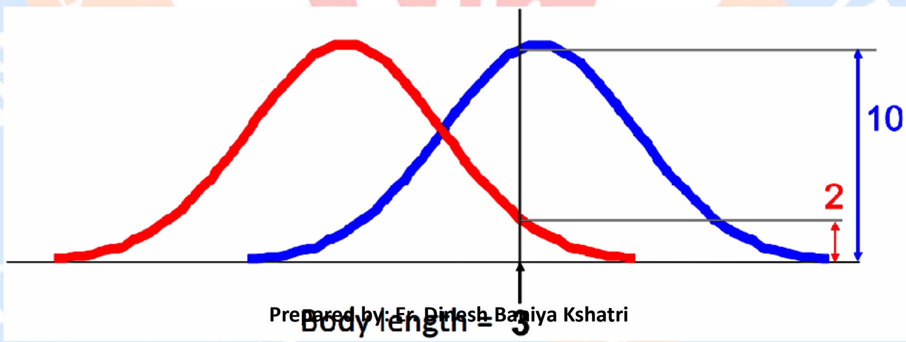
- **Suppose we wish to classify a new animal that we just found. Its body length is (X) units. How can we classify it?**
  - One way to do this is, given the distributions of that feature, we can analyze which class is more *probable: Crocodile or Alligator.*

$$p(c_j|d) = probability\ of\ class\ \mathbf{c_j}, given\ that\ we\ observed\ \mathbf{d}$$

Visual Intuition
(Naïve Bayes Model) – [5]

$$p(c_j|d) = probability\ of\ class\ \boldsymbol{c_j}, given\ that\ we\ observed\ \boldsymbol{d}$$

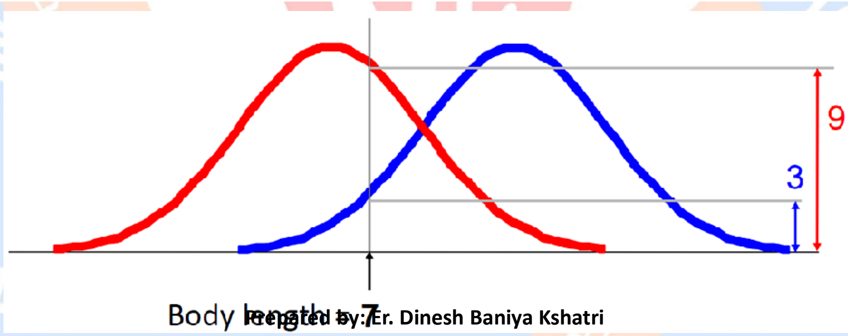
$$p(Alligator|body\ length = 3) = 10/(10+2) = \mathbf{0.833}$$
$$p(Crocodile|body\ length = 3) = 2/(10+2) = \mathbf{0.166}$$



Visual Intuition
(Naïve Bayes Model) – [6]

$$p(c_j|d) = probability\ of\ class\ \boldsymbol{c_j}, given\ that\ we\ observed\ \boldsymbol{d}$$

$$p(Alligator|body\ length = 7) = 3/(3+9) = \mathbf{0.25}$$
$$p(Crocodile|body\ length = 7) = 9/(3+9) = \mathbf{0.75}$$



Essence of Naïve Bayes Classifier

- **Naïve Bayes also called Simple Bayes**
  - This is because it makes the assumption that features of a measurement are independent of each other
- **Basic Idea of Naïve Bayes Classifier:**
  - Find the probability of the previously unseen instance belonging to each class
  - Then simply pick the most probable class

How to Estimate Probabilities from Data?

- Consider each attribute and class label as random variables

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

Evade C
Event space: {Yes, No}
P(C) = (0.3, 0.7)

Refund A_1
Event space: {Yes, No}
$P(A_1) = (0.3,0.7)$

Martial Status A_2
Event space: {Single, Married, Divorced}
$P(A_2) = (0.4,0.4,0.2)$

Taxable Income A_3
Event space: R
$P(A_3) \sim Normal(\mu,\sigma)$

- Assume attribute follows a normal distribution
- Use data to estimate parameters of distribution (i.e., mean μ and standard deviation σ)

Bayes Theorem
(Revisited)

$$p(c_j \mid d) = \frac{p(d \mid c_j)\, p(c_j)}{p(d)}$$

$p(c_j \mid d)$ = probability of instance d being in class c_j,
This is what we are trying to compute

$p(d \mid c_j)$ = probability of generating instance d given class c_j,
We can imagine that being in class c_j, causes you to have feature d with some probability

$p(c_j)$ = probability of occurrence of class c_j,
This is just how frequent the class c_j, is in our database

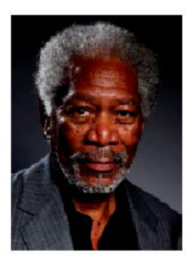
$p(d)$ = probability of instance d occurring
This can actually be ignored, since it is the same for all classes

Example of Naïve Bayes Classifier
(Guessing Gender) – [1]

Suppose we have another binary classification problem with the following two classes: $c_1 = male$, and $c_2 = female$

We now have a person called *Morgan*. How do we classify them as male or female?

Morgan Fairchild



Morgan Freeman

What is the probability of being called Morgan given that you are a male?

What is the probability of being a male?

$$P(male \mid morgan) = \frac{P(morgan \mid male)\, P(male)}{P(morgan)}$$

What is the probability of being called Morgan

Example of Naïve Bayes Classifier
(Guessing Gender) – [2]

Suppose this individual on your left (Morgan) was arrested for money laundering. Is Morgan male or female?

Assume we are given the following database of names. We can then apply Bayes rule.

Name	Sex
Morgan	Male
Reid	Female
Morgan	Male
Morgan	Female
Everaldo	Male
Francis	Male
Jennifer	Female

Example of Naïve Bayes Classifier
(Guessing Gender) – [3]

Name	Sex
Morgan	Male
Reid	Female
Morgan	Male
Morgan	Female
Everaldo	Male
Francis	Male
Jennifer	Female

$$P(c_j|d) = \frac{P(d|c_j)P(c_j)}{P(d)}$$

$$P(female|morgan) = \frac{1/3 * 3/7}{3/7} = \frac{0.143}{3/7}$$

$$P(male|morgan) = \frac{2/4 * 4/7}{3/7} = \frac{0.286}{3/7}$$

Money launderer Morgan is more likely to be male.

How to deal with Multiple Attributes?

- Both examples that we looked at considered only a single feature (i.e., *body length* and *name*)
- What if we have several features?

Name	Over 6ft	Eye color	Hair style	Sex
Morgan	Yes	Blue	Long	Female
Bob	No	Brown	None	Male
Vincent	Yes	Brown	Short	Male
Amanda	No	Brown	Short	Female
Reid	No	Blue	Short	Male
Lauren	No	Blue	Long	Female
Elisa	Yes	Brown	Long	Female

How to deal with Multiple Attributes? (Assume Conditionally Independent Features)

- Naïve Bayes assumes that all features are independent (i.e., they have independent distributions).
- The probability of class c_j generating instance d can then be estimated as:

$$P(d|c_j) = P(d_1|c_j) \times P(d_2|c_j) \times \ \dots \ \times \ P(d_n|c_j)$$

Probability of class c_j generating the observed value for feature 1

Probability of class c_j generating the observed value for feature 2

...

Dealing with Multiple Attributes

- Suppose we have Amanda's data:

Name	Over 6ft	Eye color	Hair style	Sex
Amanda	No	Brown	Short	?

$$\boldsymbol{P(Amanda|c_j)} = P(over6ft = No|c_j) \times P(eyecolor = Brown|c_j) \times P(hair = Short|c_j)$$

Naïve Bayes Classifier
(Class Exercise) – [1]

- Predict if Bob will default his loan

<u>Bob</u>
Home owner: *No*
Marital status: *Married*
Job experience: *3*

Home owner	Marital Status	Job experience (1-5)	Defaulted
Yes	Single	3	No
No	Married	4	No
No	Single	5	No
Yes	Married	4	No
No	Divorced	2	Yes
No	Married	4	No
Yes	Divorced	2	No
No	Married	3	Yes
No	Married	3	No
Yes	Single	2	Yes

Naïve Bayes Classifier
(Class Exercise) – [2]

Bob
Home owner: *No*
Marital status: *Married*
Job experience: *3*

- $P(Y = No) = 7/10$
- $P(Home\ owner = No|Y = No) = 4/7$
- $P(Marital\ status = Married|Y = No) = 4/7$
- $P(Job\ experience = 3|Y = No) = 2/7$

$$P(Bob\ will\ NOT\ default) = \frac{7}{10} \times \frac{4}{7} \times \frac{4}{7} \times \frac{2}{7} = \mathbf{0.065}$$

Home owner	Marital Status	Job experience (1-5)	Defaulted
Yes	Single	3	No
No	Married	4	No
No	Single	5	No
Yes	Married	4	No
No	Divorced	2	Yes
No	Married	4	No
Yes	Divorced	2	No
No	Married	3	Yes
No	Married	3	No
Yes	Single	2	Yes

Naïve Bayes Classifier
(Class Exercise) – [3]

Bob
Home owner: *No*
Marital status: *Married*
Job experience: *3*

- $P(Y = Yes) = 3/10$
- $P(Home\ owner = No|Y = Yes) = 2/3$
- $P(Marital\ status = Married|Y = Yes) = 1/3$
- $P(Job\ experience = 3|Y = Yes) = 1/3$

$$P(Bob\ will\ default) = \frac{3}{10} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} = \mathbf{0.022}$$

Home owner	Marital Status	Job experience (1-5)	Defaulted
Yes	Single	3	No
No	Married	4	No
No	Single	5	No
Yes	Married	4	No
No	Divorced	2	Yes
No	Married	4	No
Yes	Divorced	2	No
No	Married	3	Yes
No	Married	3	No
Yes	Single	2	Yes

Naïve Bayes Classifier
(Class Exercise) – [4]

Bob
Home owner: *No*
Marital status: *Married*
Job experience: *3*

- $P(Bob\ will\ NOT\ default) = \mathbf{0.065}$
- $P(Bob\ will\ default) = \mathbf{0.022}$

Predict: BOB WILL NOT DEFAULT

Naïve Bayes Classifier – Shortcomings
(Zero Conditional Probability)

- **If one of the conditional probabilities is zero then the entire expression becomes zero**

$$P(d|c_j) = P(d_1|c_j) \times P(d_2|c_j) \times \ldots \times P(d_n|c_j)$$
$$= 0.15 \quad \times \quad 0 \quad \times \cdots \times \quad 0.55$$

- **Could be due to:**
  - Incomplete training dataset
  - No combined occurrence of a given class and feature in the training set

Solution to Zero Conditional Probability

- Probability estimation:

$$\text{Original}: P(A_i = a \mid C = c) = \frac{N_{ac}}{N_c}$$

$$\text{Laplace}: P(A_i = a \mid C = c) = \frac{N_{ac} + 1}{N_c + N_i}$$

$$\text{m-estimate}: P(A_i = a \mid C = c) = \frac{N_{ac} + mp}{N_c + m}$$

N_i: number of attribute values for attribute A_i

p: prior probability

m: parameter

Prepared by: Er. Dinesh Baniya Kshatri 45

Solution to Zero Conditional Probability (m-Estimate)

- To avoid trouble when a probability $P(d_1|c_j) = 0$, we fix its prior probability and the number of samples to some non-zero value beforehand
  - Think of it as adding a bunch of fake instances before we start the whole process
- If we create $m > 0$ fake samples of feature X with value of x, and we assign a prior probability p to them, then posterior probabilities are obtained as:

$$P(X = x|c_j) = \frac{\#(X=x, c_j) + mp}{\#(c_j) + m}$$

Prepared by: Er. Dinesh Baniya Kshatri 46

Solution to Zero Conditional Probability (Laplace Smoothing)

- To eliminate zero joint probability, use add-one or Laplace smoothing
- Adds arbitrary low probabilities
- Prevents computation from becoming zero

Solution to Zero Conditional Probability (Laplace Smoothing)

X_i = The i-th attribute in dataset D.

x_i = A particular value of the X_i attribute in dataset D.

N = Total number of tuples in dataset D.

k = Laplace Smoothing Factor.

Count ($X_i = x_i$) = Number of tuples where the attribute X_i takes the value x_i

| X_i | = Number of different values attribute X_i can take.

$$P_{Lap,k}(X_i = x_i) = \frac{count(X_i = x_i) + k}{N + k|X_i|}$$

Laplace Smoothing: Example – [1]

- Class buys_computer = yes and an attribute income = {low, medium, high} in some training database, D, containing 1000 tuples such that

  0 tuples with income = low

  990 tuples with income = medium

  10 tuples with income = high
- **The probabilities of these events, without the Laplacian correction, are**

  P(income=low | buys_computer = yes) = 0

  P(income=medium | buys_computer = yes) = 0.990 (i.e. 990/1000)

  P(income=high | buys_computer = yes) = 0.010 (i.e. 10/1000)
- Lets use Laplacian correction, using k = **1** for each of the three attribute values.

Laplace Smoothing: Example – [2]

- Class buys_computer = yes and an attribute income = {low, medium, high} in some training database, D, containing 1000 **+ 3 = 1003** tuples such that

  ~~**0** tuples with income = low~~ **1 tuples with income = low**

  ~~**990** tuples with income = medium~~ **991 tuples with income = medium**

  ~~**10** tuples with income = high~~ **11 tuples with income = high**

- Using Laplacian correction, using k = 1 for each of the three attribute values.
- The "corrected" probability estimates are close to their "uncorrected" counterparts.

Laplace Smoothing: Example – [3]

- The new probabilities of these events, with the Laplacian correction, are

  $P_{LAP,K=1}$ (income=low | buys_computer = yes) = 0.001 (i.e. 1/1003)

  $P_{LAP,K=1}$ (income=medium | buys_computer = yes) = 0.988 (i.e. 991/1003)

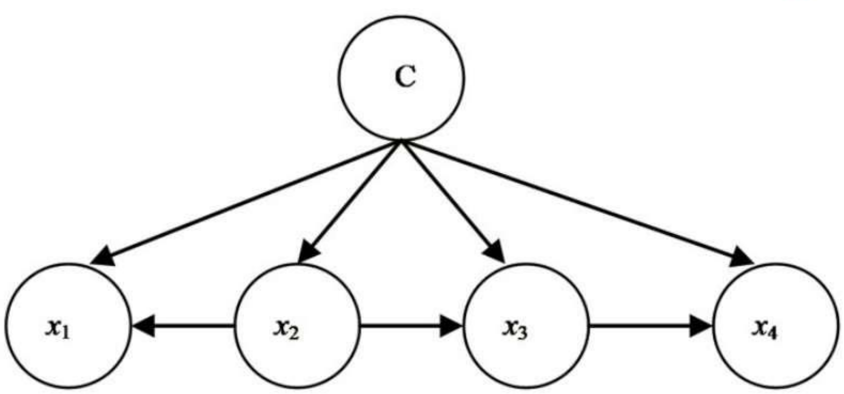
  $P_{LAP,K=1}$ (income=high | buys_computer = yes) = 0.0109 (i.e. 11/1003)

- The "corrected" probability estimates are close to their "uncorrected" counterparts

  The zero probability value is avoided!
- Note: N i.e. total number of tuples is increased to 1003 from 1000.

Naïve Bayes Classifier – Shortcomings
(Correlated Attributes)

What if the attributes have some correlation among themselves?



P (C | x1, x2, x3, x4) = P(C) P(x1|C) P(x2|C) P(x3|C) P(x4|C)

Solution to Correlated Attributes – [1]

1. When it is known beforehand that a few of the attributes are correlated.
   - **Ignore one of the correlated attributes if it's not giving any significant information. For Example: attributes age_group={child,youth,old_aged}, age $\in$ [10,60] in a dataset.**

2. When it is not known which attributes are dependent on the other.
   - **Find the correlation among attributes. For example, Pearson Correlation Test to know the correlation between two attributes.**

Solution to Correlated Attributes – [2] (Pearson Correlation Test)

- To investigate the relationship between two continuous variables/attributes X and Y in the dataset.
- X-bar = Mean of Attribute X, Y-bar = Mean of Attribute Y.
- **'r' measures the strength of the association.**
- **r $\in$ [-1, 1]**

$$r = \frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum(X-\overline{X})^2}\sqrt{\sum(Y-\overline{Y})^2}}$$