Pattern Recognition
CSE 467

Classification: Naïve Bayes' Classifier

Modification of the content created by

Dr. Debasis Samanta

Associate Professor

Department of Computer Science & Engineering, IIT (Khog.)

Today's discussion…

- Introduction to Classification
- Classification Techniques
  - Supervised and unsupervised classification
- Formal statement of supervised classification technique
- Bayesian Classifier
  - Principle of Bayesian classifier
  - Bayes' theorem of probability
- Naïve Bayesian Classifier

A Simple Quiz: Identify the objects

Introduction to Classification

Example 8.1

- Teacher classify students as A, B, C, D and F based on their marks. The following is one simple classification rule:

Mark $\geq$ 90	**:**	**A**
90 > Mark $\geq$ 80	**:**	**B**
80 > Mark $\geq$ 70	**:**	**C**
70 > Mark $\geq$ 60	**:**	**D**
60 > Mark	**:**	**F**

Note:

Here, we apply the above rule to a specific data (in this case a table of marks).

Examples of Classification in Data Analytics

- Life Science: Predicting tumor cells as benign or malignant
- Security: Classifying credit card transactions as legitimate or fraudulent
- Prediction: Weather, voting, political dynamics, etc.
- Entertainment: Categorizing news stories as finance, weather, entertainment, sports, etc.
- Social media: Identifying the current trend and future growth

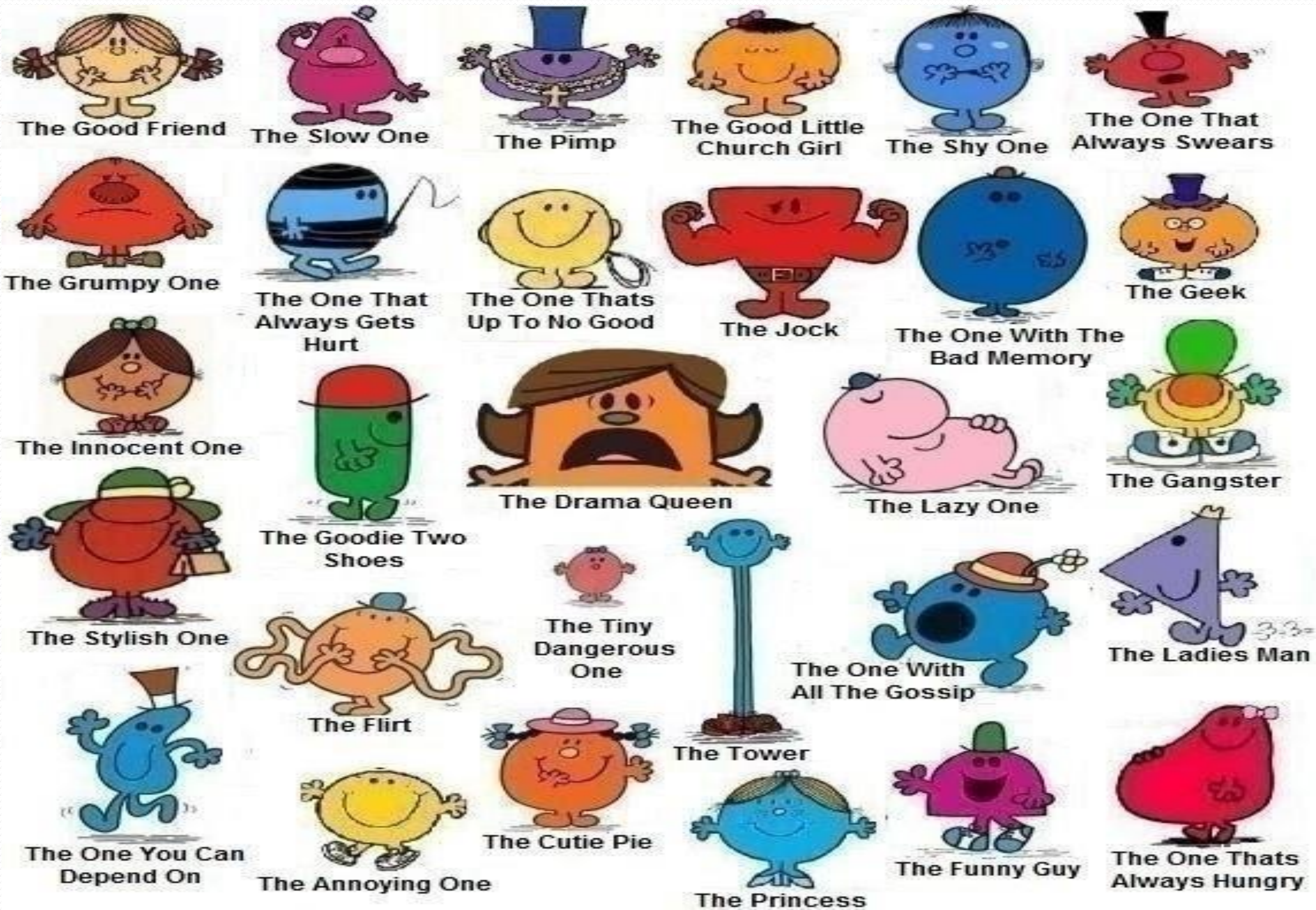
Classification : Definition

- Classification is a form of data analysis to extract models describing important data classes.
- Essentially, it involves dividing up objects so that each is assigned to one of a number of mutually exhaustive and exclusive categories known as classes.
    - The term "mutually exhaustive and exclusive" simply means that each object must be assigned to precisely one class
        - That is, never to more than one and never to no class at all.

Classification Techniques

- Classification consists of assigning a class label to a set of unclassified cases.
- **Supervised Classification**
  - The set of possible classes is known in advance.
- **Unsupervised Classification**
  - Set of possible classes is not known. After classification we can try to assign a name to that class.
    - Unsupervised classification is called **clustering**.

Supervised Classification



Unsupervised Classification

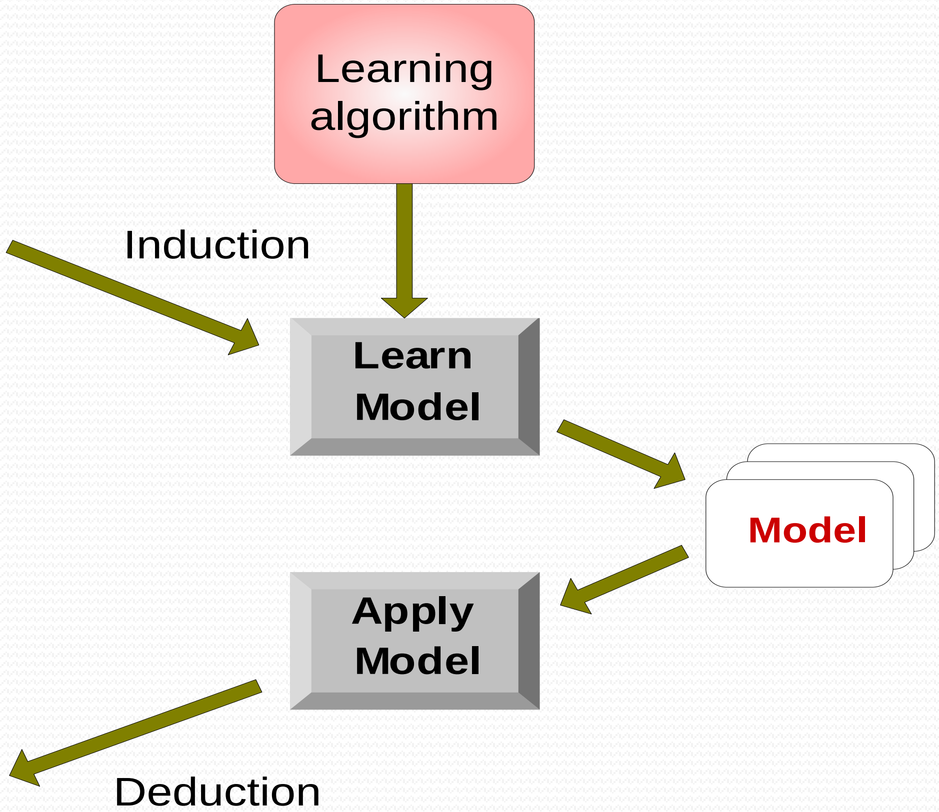
Supervised Classification Technique

- Given a collection of records (*training set*)
  - Each record contains a set of *attributes*, one of the attributes is the *class*.
- Find a *model* for class attribute as a function of the values of other attributes.
- Goal: Previously unseen records should be assigned a class as accurately as possible.
  - Satisfy the property of "mutually exclusive and exhaustive"

Illustrating Classification Tasks

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set



Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set

Classification Problem

- More precisely, a classification problem can be stated as below:

Definition 8.1: Classification Problem

Given a database D = $\{t_1, t_2, \ldots\ldots, t_m\}$ of tuples and a set of classes C = $\{c_1, c_2, \ldots\ldots, c_k\}$, the classification problem is to define a mapping f : D → C,

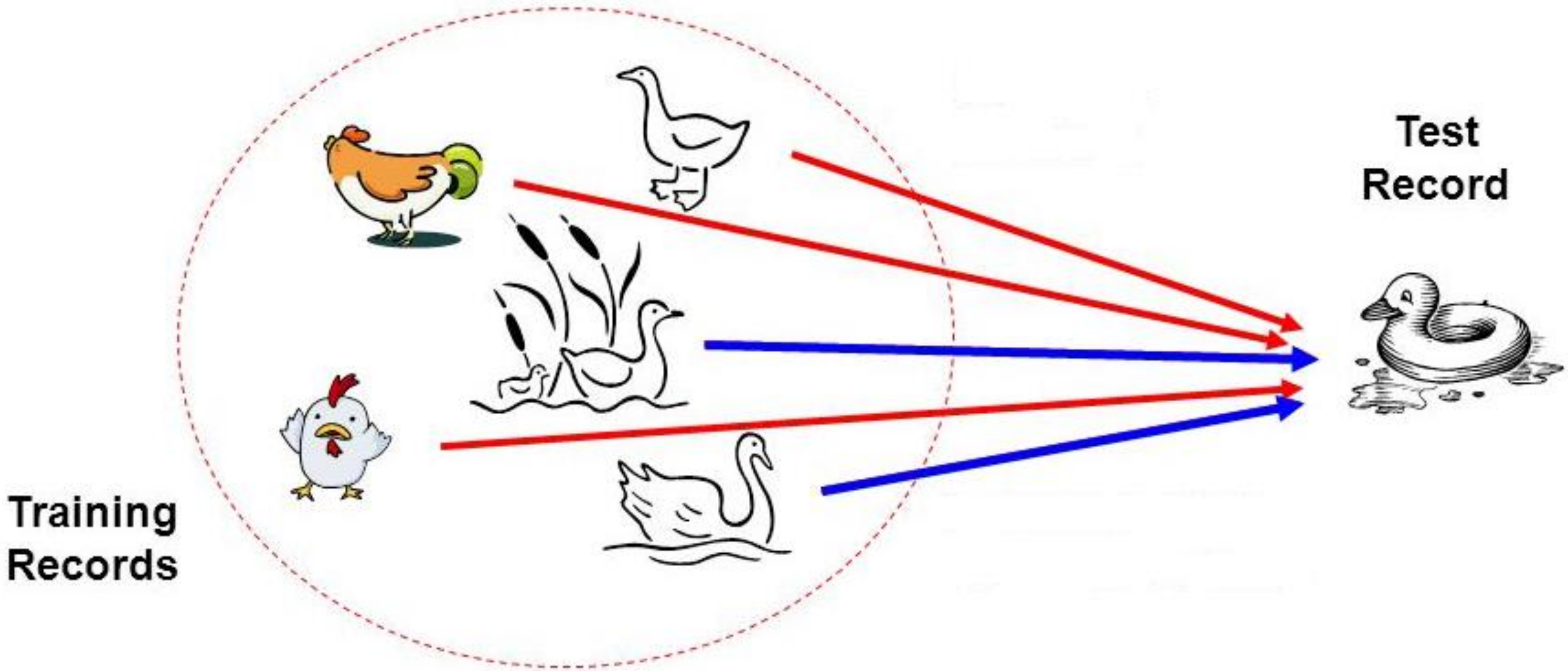
Where each t_i is assigned to one class.

Note that tuple $t_i \in D$ is defined by a set of attributes $A = \{A_1, A_2, \ldots\ldots, A_n\}$.

Bayesian Classifier

Bayesian Classifier

- Principle
  - If it walks like a duck, quacks like a duck, then it is probably a duck



Bayesian Classifier

- A statistical classifier
  - Performs *probabilistic prediction*, *i.e.*, predicts class membership probabilities
- Foundation
  - Based on Bayes' Theorem.
- Assumptions
  1. The classes are mutually exclusive and exhaustive.
  2. The attributes are independent given the class.
- Called "Naïve" classifier because of these assumptions.
  - Empirically proven to be useful.
  - Scales very well.

Air-Traffic Data

Days	Season	Fog	Rain	Class
Weekday	Spring	None	None	On Time
Weekday	Winter	None	Slight	On Time
Weekday	Winter	None	None	On Time
Holiday	Winter	High	Slight	Late
Saturday	Summer	Normal	None	On Time
Weekday	Autumn	Normal	None	Very Late
Holiday	Summer	High	Slight	On Time
Sunday	Summer	Normal	None	On Time
Weekday	Winter	High	Heavy	Very Late
Weekday	Summer	None	Slight	On Time

Cond. to next slide…

Air-Traffic Data

Cond. from previous slide...

Days	Season	Fog	Rain	Class
Saturday	Spring	High	Heavy	Cancelled
Weekday	Summer	High	Slight	On Time
Weekday	Winter	Normal	None	Late
Weekday	Summer	High	None	On Time
Weekday	Winter	Normal	Heavy	Very Late
Saturday	Autumn	High	Slight	On Time
Weekday	Autumn	None	Heavy	On Time
Holiday	Spring	Normal	Slight	On Time
Weekday	Spring	Normal	None	On Time
Weekday	Spring	Normal	Heavy	On Time

Air-Traffic Data

- In this database, there are four attributes

  A = [Day, Season, Fog, Rain]

  with 20 tuples.
- The categories of classes are:

  C= [On Time, Late, Very Late, Cancelled]

- Given this is the knowledge of data and classes, we are to find most likely classification for any other unseen instance, for example:

Week Day	**Winter**	**High**	**None**	**???**

- Classification technique eventually to map this tuple into an accurate class.

Bayesian Classifier

- In many applications, the relationship between the attributes set and the class variable is non-deterministic.
  - In other words, a test cannot be classified to a class label with certainty.
  - In such a situation, the classification can be achieved probabilistically.
- The Bayesian classifier is an approach for modelling probabilistic relationships between the attribute set and the class variable.
- More precisely, Bayesian classifier use Bayes' Theorem of Probability for classification.
- Before going to discuss the Bayesian classifier, we should have a quick look at the Theory of Probability and then Bayes' Theorem.

Bayes' Theorem of Probability

Simple Probability

Simple Probability

If there are n elementary events associated with a random experiment and m of n of them are favorable to an event A, then the probability of happening or occurrence of A is

$$P(A) = \frac{m}{n}$$

Simple Probability

- Suppose, A and B are any two events and *P(A)*, *P(B)* denote the probabilities that the events *A* and *B* will occur, respectively.

- **Mutually Exclusive Events:**
  - Two events are mutually exclusive, if the occurrence of one precludes the occurrence of the other.

    Example: Tossing a coin (two events)

    Tossing a ludo cube (Six events)

Can you give an example, so that two events are not mutually exclusive?

Hint: Weather (sunny, foggy, warm)

Simple Probability

- **Independent events:** Two events are independent if occurrences of one does not alter the occurrence of other.

Example: Tossing both coin and ludo cube together.
(How many events are here?)

- Can you give an example, where an event is dependent on one or more other events(s)?

Hint: Consider a bag contains 7 black balls, 3 red balls.

Joint Probability

Joint Probability

If *P(A)* and *P(B)* are the probability of two events, then

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

If A and B are mutually exclusive, then $P(A \cap B) = 0$
If A and B are independent events, then $P(A \cap B) = P(A).P(B)$

Thus, for mutually exclusive events

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

Conditional Probability

Conditional Probability

If events are dependent, then their probability is expressed by conditional probability. The probability that A occurs given that B is denoted by $P(A|B)$.

Suppose, A and B are two events associated with a random experiment. The probability of A under the condition that B has already occurred and $P(B) \neq 0$ is given by

$$P(A|B) = \frac{\text{Number of events in } B \text{ which are favourable to } A}{\text{Number of events in } B}$$

$$= \frac{\text{Number of events favourable to } B \cap A}{\text{Number of events favourable to } B}$$

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

Conditional Probability

Conditional Probability

$$P(B \cap A) = P(B).P(A|B), \qquad if\ P(B) \neq 0$$
$$P(A \cap B) = P(A).P(B|A), \qquad if\ P(A) \neq 0$$

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

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

Conditional Probability

- Generalization of Conditional Probability:

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

By the law of total probability : $P(B) = P[(B \cap A) \cup (B \cap \bar{A})]$, where $\bar{A}$ denotes the compliment of event A. Thus,

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P[(B \cap A) \cup (B \cap \bar{A})]}$$

$$= \frac{P(B|A)\cdot P(A)}{P(B|A)\cdot P(A)+P(B\,|\,\bar{A})\cdot P(\bar{A})}$$

Conditional Probability

In general,

$$P(A1|x) = \frac{P(A1)\ \cdot P(x|A1)}{P(A1) \cdot P(x|A1) + P(A2) \cdot P(x|A2) + P(A3) \cdot P(x|A3)}$$

$$P(x) = P(A1) \cdot P(x|A1) + P(A2) \cdot P(x|A2) + P(A3) \cdot P(x|A3)$$

$$P(x) = \sum_{i=1}^{3} P(Ai).P(x|Ai)$$

Let consider we have two classes (A1 and A2) for a feature vector x we can write,

$$P(A1|x) = \frac{P(A1) \cdot P(x|A1)}{P(x)}$$

$$P(x) = P(A1) \cdot P(x|A1) + P(A2) \cdot P(x|A2)$$

Classifier Design

Let consider we have two classes (A1 and A2) for a feature vector x we can write,

$$P(A1|x) = \frac{P(A1) \cdot P(x|A1)}{P(x)}$$

$$P(x) = P(A1) \cdot P(x|A1) + P(A2) \cdot P(x|A2)$$

$$P(A2|x) = \frac{P(A2) \cdot P(x|A2)}{P(x)}$$

If $P(A1|x) > P(A2|x)$, then x belongs to A1 class else A2 class

For more than two classes how can we take decision?

Classifier Design (Continued …..)

Let consider we have two classes (A_1 and A_2) for a feature vector x we can write,

$P(A1|x) = P(A1) \cdot P(x|A1)$
$P(A2|x) = P(A2) \cdot P(x|A2)$
If $P(A_1|x) > P(A_2|x)$, then x belongs to A_1 class else A_2 class

How to calculate $P(x|C)$? Consider, $x = \{x_1, x_2, x_3, \ldots\ldots\ldots, x_n\}$

$P(x|C)=P(\{x_1,x_2,x_3, \ldots\ldots\ldots\ldots, x_n\}|C)= P(x_1|C).\ P(x_2|C).\ P(x_3|C) \ldots\ldots\ldots\ldots P(x_n|C)$

Now for $x = \{x_1, x_2, x_3, \ldots\ldots\ldots, x_n\}$,

$$P(A1|x) = P(A1) \cdot \{P(x1|A1).P(x2|A1).P(x3|A1) \ \ldots\ldots\ldots..P(xn|A1)\}$$

$$P(A2|x) = P(A2) \cdot \{P(x1|A2).P(x2|A2).P(x3|A2) \ \ldots\ldots\ldots..P(xn|A2)\}$$

Example

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

Example (Continued ....)

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

Example (Continued …….)

$$P(x \mid c) = P(Sunny \mid Yes) = 3/9 = 0.33$$

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

Likelihood Table		Play Golf		
		Yes	No	
Outlook	Sunny	3/9	2/5	5/14
	Overcast	4/9	0/5	4/14
	Rainy	2/9	3/5	5/14
		9/14	5/14	

$$P(x) = P(Sunny) = 5/14 = 0.36$$

$$P(c) = P(Yes) = 9/14 = 0.64$$

Posterior Probability: $P(c \mid x) = P(Yes \mid Sunny) = 0.33 \times 0.64 \div 0.36 = 0.60$

$$P(x \mid c) = P(Sunny \mid No) = 2/5 = 0.4$$

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

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
		9	5	14

$$P(x) = P(Sunny) = 5/14 = 0.36$$

$$P(c) = P(No) = 5/14 = 0.36$$

Posterior Probability: $P(c \mid x) = P(No \mid Sunny) = 0.40 \times 0.36 \div 0.36 = 0.40$

Example (Continued ....)

Outlook	Temp	Humidity	Windy	Play
Rainy	Cool	High	True	?

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

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

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

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

Home Work

- Consider the dataset in 16 & 17 number slide and you have to find most likely classification for the unseen instance given below using Naïve Bayes Classifier:

Week Day	Winter	High	None	???

Thanks To

Dr. Debasis Samanta

Associate Professor

Department of Computer Science & Engineering, IIT (Khog.)

&&

https://www.saedsayad.com/naive_bayesian.htm

[contents of slide no. 31 to 34 are taken]

Thank You