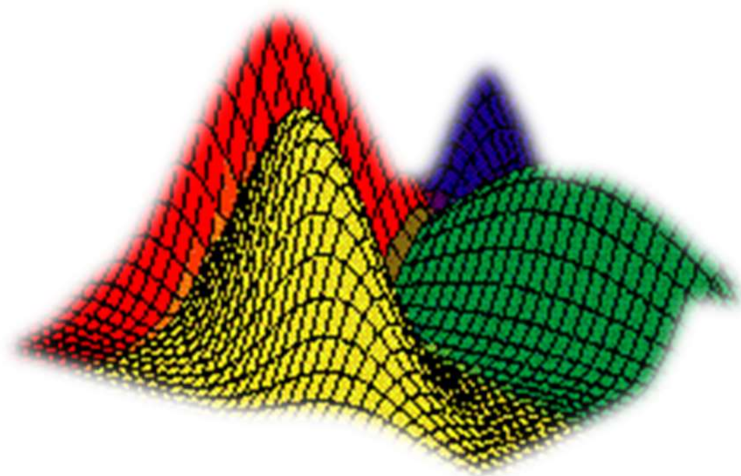
# CSE 473
# Pattern Recognition




**Instructor:**

**Dr. Md. Monirul Islam**

# Classification Example 2

| *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 |

- A married person with income 120K did not refund the loan previously
- *Can we trust him*?

# Bayesian Classifiers

- We have multiple attributes ($A_1, A_2, \ldots, A_n$)
  - Goal is to predict class $C$
  - Specifically, we want to find the value of C that maximizes $P(C| A_1, A_2, \ldots, A_n)$
- Can we estimate $P(C| A_1, A_2, \ldots, A_n)$ directly from data?

# Bayesian Classifiers

- Approach:
  - compute the posterior probability $P(C \mid A_1, A_2, \ldots, A_n)$ for all values of $C$ using the Bayes theorem

$$\boxed{P(C \mid A_1 A_2 \ldots A_n) = \frac{P(A_1 A_2 \ldots A_n \mid C) P(C)}{P(A_1 A_2 \ldots A_n)}}$$

# Bayesian Classifiers

- Approach:
  - compute the posterior probability $P(C \mid A_1, A_2, \ldots, A_n)$ for all values of $C$ using the Bayes theorem

$$\boxed{P(C \mid A_1 A_2 \ldots A_n) = \frac{P(A_1 A_2 \ldots A_n \mid C) P(C)}{P(A_1 A_2 \ldots A_n)}}$$

  - Choose value of $C$ that maximizes
    $P(C \mid A_1, A_2, \ldots, A_n)$

# Bayesian Classifiers

- Approach:
  - compute the posterior probability $P(C \mid A_1, A_2, \ldots, A_n)$ for all values of $C$ using the Bayes theorem

$$\boxed{P(C \mid A_1 A_2 \ldots A_n) = \frac{P(A_1 A_2 \ldots A_n \mid C) P(C)}{P(A_1 A_2 \ldots A_n)}}$$

  - Choose value of $C$ that maximizes
    $P(C \mid A_1, A_2, \ldots, A_n)$
  - Equivalent to choosing value of $C$ that maximizes
    $P(A_1, A_2, \ldots, A_n \mid C)\, P(C)$

# Bayesian Classifiers

- Approach:
  - compute the posterior probability $P(C \mid A_1, A_2, \ldots, A_n)$ for all values of $C$ using the Bayes theorem

$$\boxed{P(C \mid A_1 A_2 \ldots A_n) = \frac{P(A_1 A_2 \ldots A_n \mid C) P(C)}{P(A_1 A_2 \ldots A_n)}}$$

  - Choose value of $C$ that maximizes

    $P(C \mid A_1, A_2, \ldots, A_n)$

  - Equivalent to choosing value of $C$ that maximizes

    $P(A_1, A_2, \ldots, A_n|C)\, P(C)$

- How to estimate $P(A_1, A_2, \ldots, A_n \mid C)$?

# Naïve Bayes Classifier

- Assume independence among attributes $A_i$ when class is given:
  - $P(A_1, A_2, \ldots, A_n | C_j) = P(A_1 | C_j)\, P(A_2 | C_j) \ldots P(A_n | C_j)$

# Naïve Bayes Classifier

- Assume independence among attributes $A_i$ when class is given:
  - $P(A_1, A_2, \ldots, A_n | C_j) = P(A_1| C_j)\, P(A_2| C_j) \ldots P(A_n| C_j)$
  - can estimate $P(A_i| C_j)$ for all $A_i$ and $C_j$ from training data
  - *the new pattern* is classified to $C_j$ if $P(C_j) \Pi P(A_i| C_j)$ is maximum

# How to 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 |

- Class: $P(C) = N_c/N$
  - e.g., $P(No) = 7/10$,
    $P(Yes) = 3/10$

# How to 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 |

- Class: $P(C) = N_c/N$
  - e.g., $P(No) = 7/10$,
    $P(Yes) = 3/10$
- For discrete attributes:

$$P(A_i \mid C_k) = |A_{ik}| / N_c$$

  - where $|A_{ik}|$ is number of instances having attribute $A_i$ and belongs to class $C_k$
  - Examples:

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

# How to Estimate Probabilities from Data?

- For continuous attributes:
  - Discretize the range into bins
    - one ordinal attribute per bin
  - Two-way split: $(A < v)$ or $(A > v)$
    - choose only one of the two splits as new attribute
  - Probability density estimation:
    - Assume attribute follows a normal distribution
    - Use data to estimate parameters of distribution (e.g., mean and standard deviation)
    - Once probability distribution is known, can use it to estimate the conditional probability $P(A_i|c)$

# How to 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:

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

  - One for each $(A_i, c_j)$ pair

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

# How to 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:

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

  - One for each $(A_i, c_j)$ pair
- For (Income, Class=No):
  - If Class=No
    - sample mean = 110K
    - sample variance = 2975

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

# Example of Naïve Bayes Classifier

Given a Test Record: $X = (\text{Refund} = \text{No}, \text{Married}, \text{Income} = 120\text{K})$

**naive 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) = 2/7
P(Marital Status=Divorced|Yes)=1/7
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|Class=No) = P(Refund=No|Class=No)
  × P(Married| Class=No)
  × P(Income=120K| Class=No)
  = 4/7 × 4/7 × 0.0072 = 0.0024

# Example of Naïve Bayes Classifier

Given a Test Record: $X = (\text{Refund} = \text{No}, \text{Married}, \text{Income} = 120\text{K})$

**naive 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) = 2/7
P(Marital Status=Divorced|Yes)=1/7
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|Class=No) = P(Refund=No|Class=No) $\times$ P(Married| Class=No) $\times$ P(Income=120K| Class=No) = 4/7 $\times$ 4/7 $\times$ 0.0072 = 0.0024

- P(X|Class=Yes) = P(Refund=No| Class=Yes) $\times$ P(Married| Class=Yes) $\times$ P(Income=120K| Class=Yes) = 1 $\times$ 0 $\times$ 1.2 $\times$ $10^{-9}$ = 0

# Example of Naïve Bayes Classifier

Given a Test Record: $X = (\text{Refund} = \text{No}, \text{Married}, \text{Income} = 120\text{K})$

**naive 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) = 2/7
P(Marital Status=Divorced|Yes)=1/7
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|Class=No) = P(Refund=No|Class=No) $\times$ P(Married| Class=No) $\times$ P(Income=120K| Class=No) = 4/7 $\times$ 4/7 $\times$ 0.0072 = 0.0024

- P(X|Class=Yes) = P(Refund=No| Class=Yes) $\times$ P(Married| Class=Yes) $\times$ P(Income=120K| Class=Yes) = 1 $\times$ 0 $\times$ 1.2 $\times$ $10^{-9}$ = 0

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

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

# Example-2 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 |

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

# Example-2 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

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

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

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

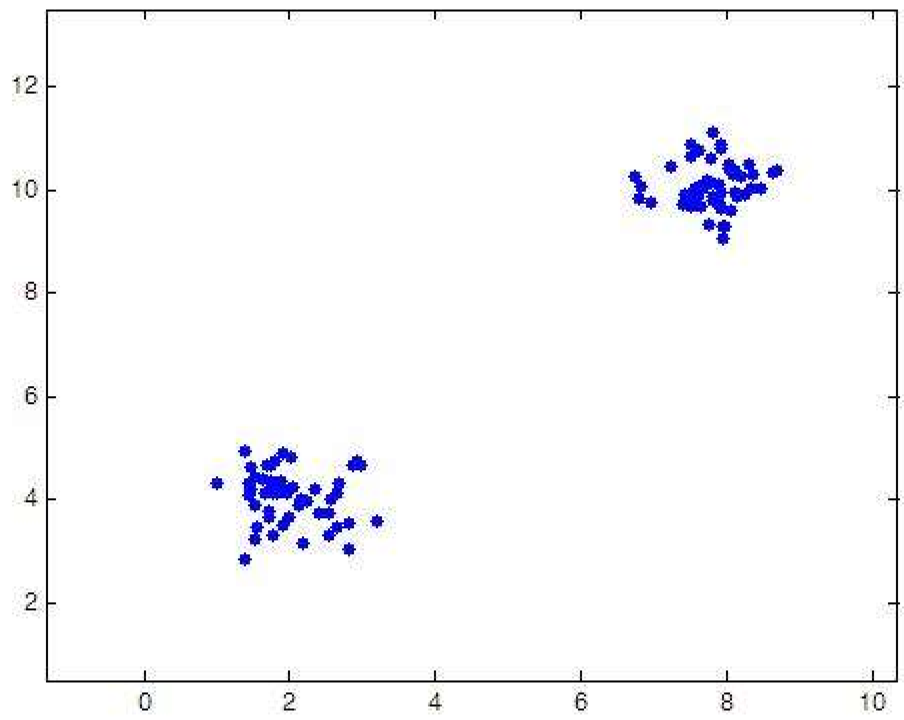
$$P(A|N)P(N) = 0.004 \times \frac{13}{20} = 0.0027$$

| 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

# Sample Data for Sessional on Bayesian Classification



## Sample Data for Bayesian Classification

| Feature 1 | Feature 2 | Class |
|---|---|---|
| 1.7044 | 3.6651 | 1 |
| 1.6726 | 4.6705 | 1 |
| 1.4597 | 4.194 | 1 |
| 1.9761 | 4.1965 | 1 |
| 1.9126 | 3.4987 | 1 |
| 1.5214 | 3.9072 | 1 |
| 2.6463 | 3.473 | 1 |
| 2.2205 | 3.9642 | 1 |
| 6.8104 | 10.0517 | 2 |
| 7.5809 | 9.8897 | 2 |
| 8.1287 | 9.8605 | 2 |
| 7.9081 | 9.6332 | 2 |
| 7.9162 | 9.9677 | 2 |
| 7.9415 | 9.278 | 2 |
| 8.0842 | 10.3062 | 2 |
| 7.7494 | 9.3382 | 2 |
| 8.1146 | 9.9617 | 2 |

# Naïve Bayes (Summary)

- Robust to isolated noise points
- Handle missing values by ignoring the instance during probability estimate calculations
- Robust to irrelevant attributes
- Independence assumption may not hold for some attributes
  - Use other techniques such as Bayesian Belief Networks (BBN)

# Bayesian Belief Networks

- Let we have *l* random variables
- The joint probability is given by,

$$p(x_1, x_2, \ldots, x_\ell)$$

which is calculated as

$$p(x_\ell \mid x_{\ell-1}, \ldots, x_1) \cdot p(x_{\ell-1} \mid x_{\ell-2}, \ldots, x_1) \cdots p(x_2 \mid x_1) \cdot p(x_1)$$

# Bayesian Belief Networks

The formula

$$p(x_1, x_2, \ldots, x_\ell) = p(x_\ell \mid x_{\ell-1}, \ldots, x_1) \cdot p(x_{\ell-1} \mid x_{\ell-2}, \ldots, x_1) \cdot \ldots$$
$$\ldots \cdot p(x_2 \mid x_1) \cdot p(x_1)$$

can be verified as

$$\begin{aligned}
p(x_1, x_2, \ldots, x_\ell) &= p(x_1, x_2, \ldots, x_{\ell-1}, x_\ell) \\
&= p(x_1, x_2, \ldots, x_{\ell-1}) \cdot p(x_\ell \mid x_1, x_2, \ldots, x_{\ell-1}) \\
&= p(x_1, x_2, \ldots, x_{\ell-1}) \cdot p(x_\ell \mid x_{\ell-1}, \ldots, x_2, x_1) \\
&\vdots
\end{aligned}$$

# Bayesian Belief Networks

The formula

$$p(x_1, x_2, \ldots, x_\ell) = p(x_\ell \mid x_{\ell-1}, \ldots, x_1) \cdot p(x_{\ell-1} \mid x_{\ell-2}, \ldots, x_1) \cdot \ldots$$
$$\ldots \cdot p(x_2 \mid x_1) \cdot p(x_1)$$

can be written as

$$p(x_1, x_2, \ldots, x_\ell) = p(x_1) \cdot \prod_{i=2}^{\ell} p(x_i \mid A_i)$$

where

$$A_i \subseteq \{x_{i-1}, x_{i-2}, \ldots, x_1\}$$