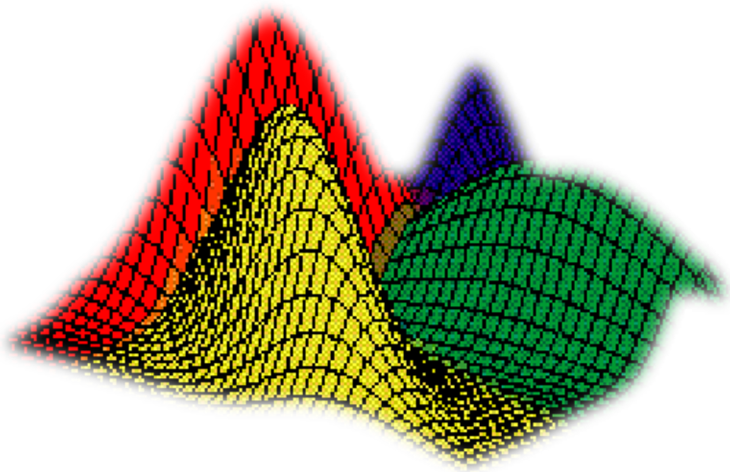
CSE 473
Pattern Recognition

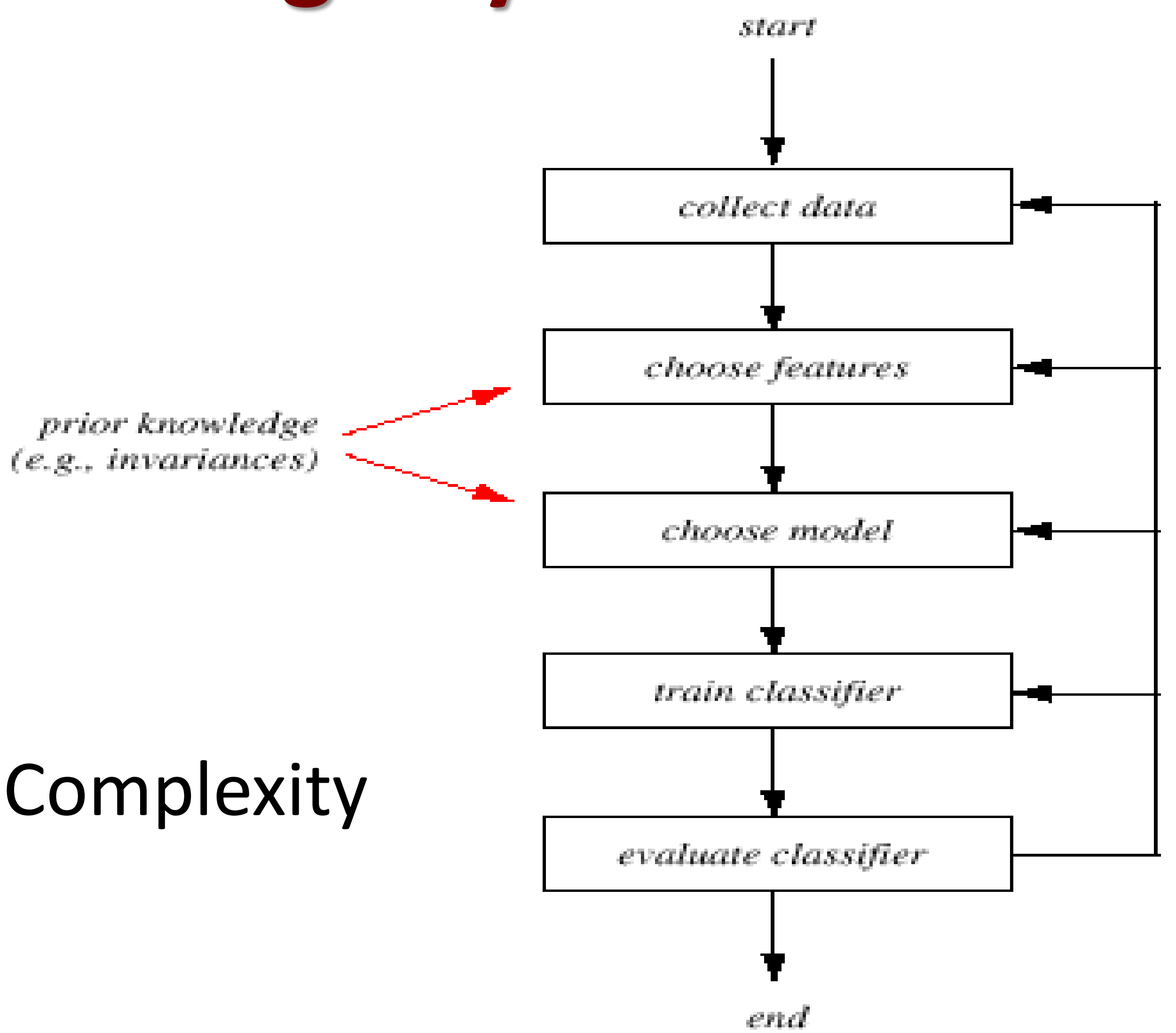


Lecturer:

Dr. Md. Monirul Islam

The Design Cycle

- Data collection
- Feature Choice
- Model Choice
- Training
- Evaluation
- Computational Complexity



- Data Collection

  - How do we know when we have collected an adequately large and representative set of examples for training and testing the system?

- Feature Choice
  - Depends on the characteristics of the problem domain.
  - Requirement
    - simple to extract
    - invariant to irrelevant transformation
    - insensitive to noise.

- Model Choice
  - too many classification models?
  - which one is best?

- Training
  - Use data to determine the classifier. Many different procedures for training classifiers and choosing models

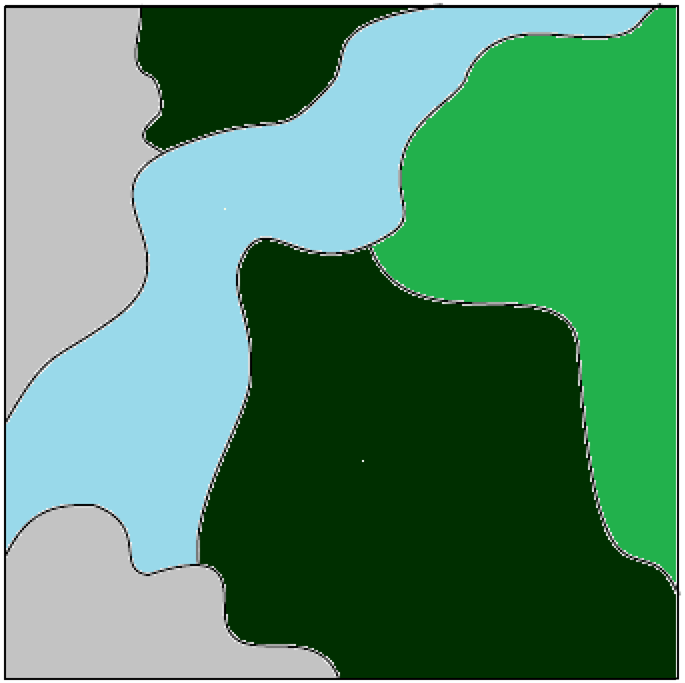
- Evaluation
  - Measure the error rate (or performance) and switch from one set of features to another

- Computational Complexity
  - What is the trade-off between computational ease and performance?

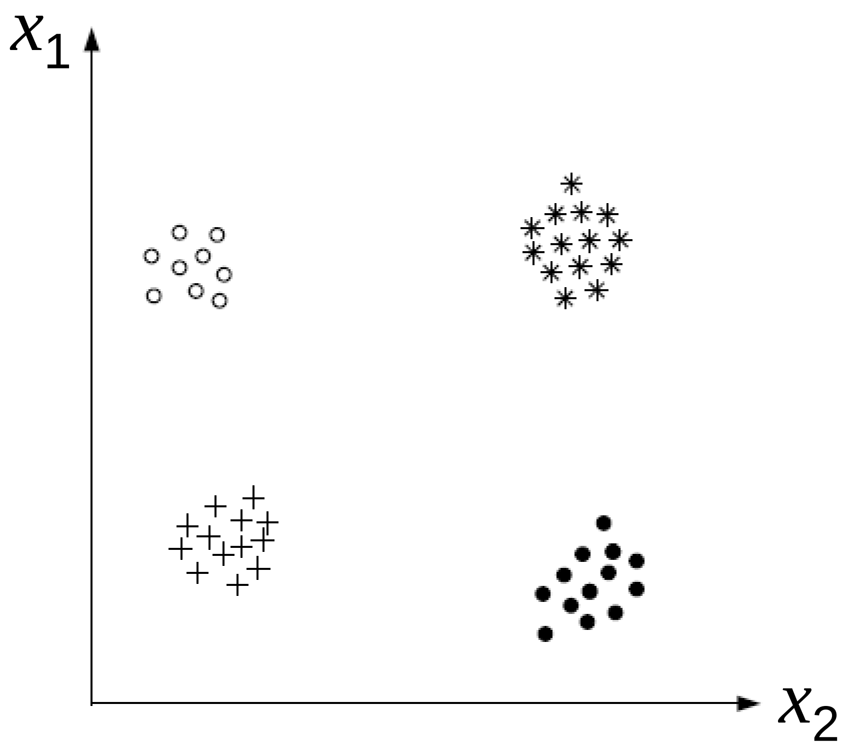
Supervised vs. Unsupervised Learning

- Supervised learning
  - A teacher provides a category label or cost for each pattern in the training set
- Unsupervised learning
  - The system forms clusters or “natural groupings” of the input patterns

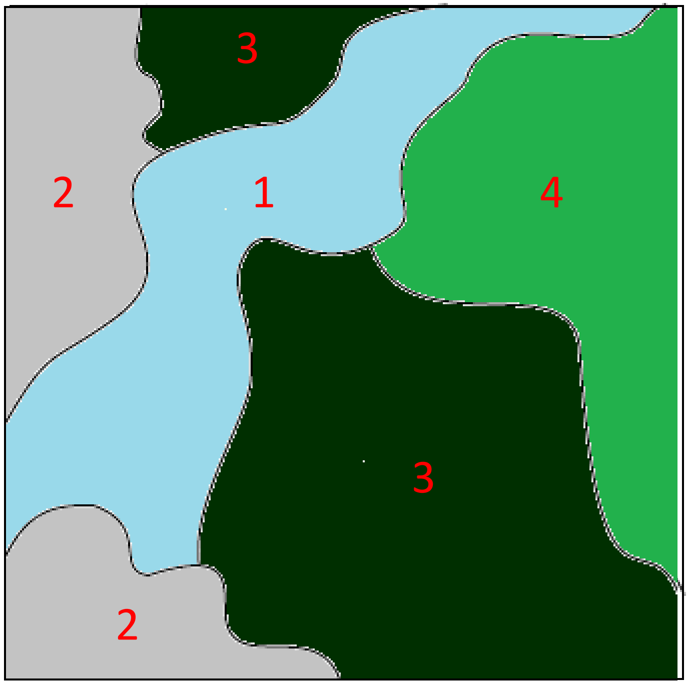
Unsupervised Learning

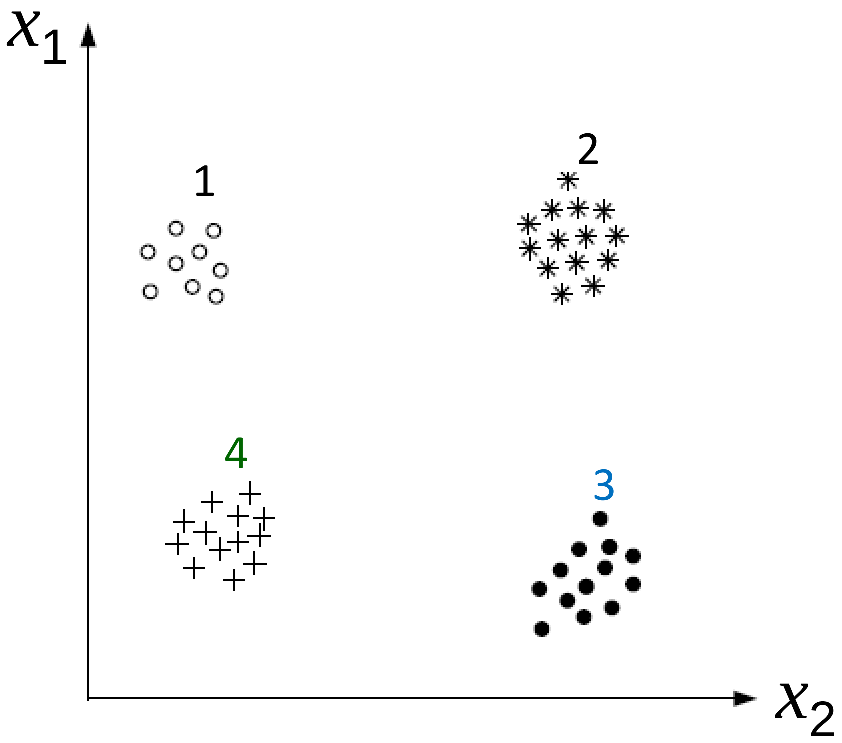


Unsupervised Learning

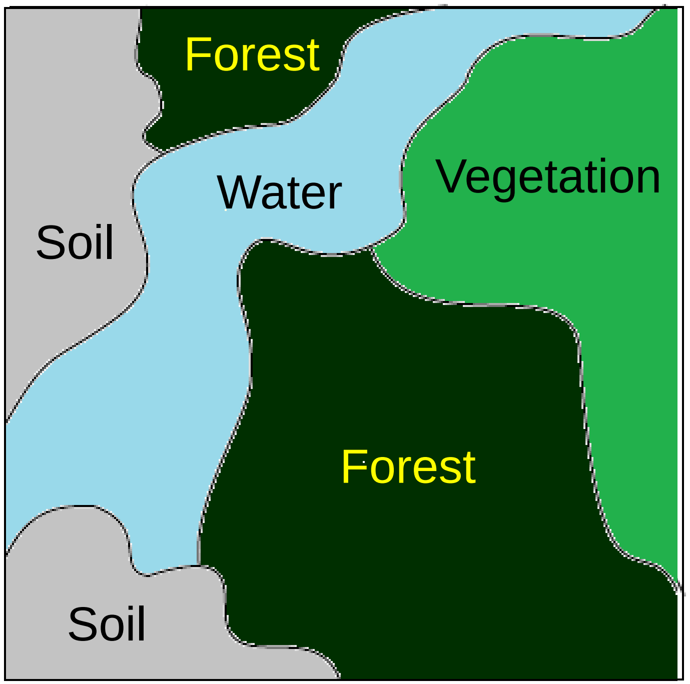


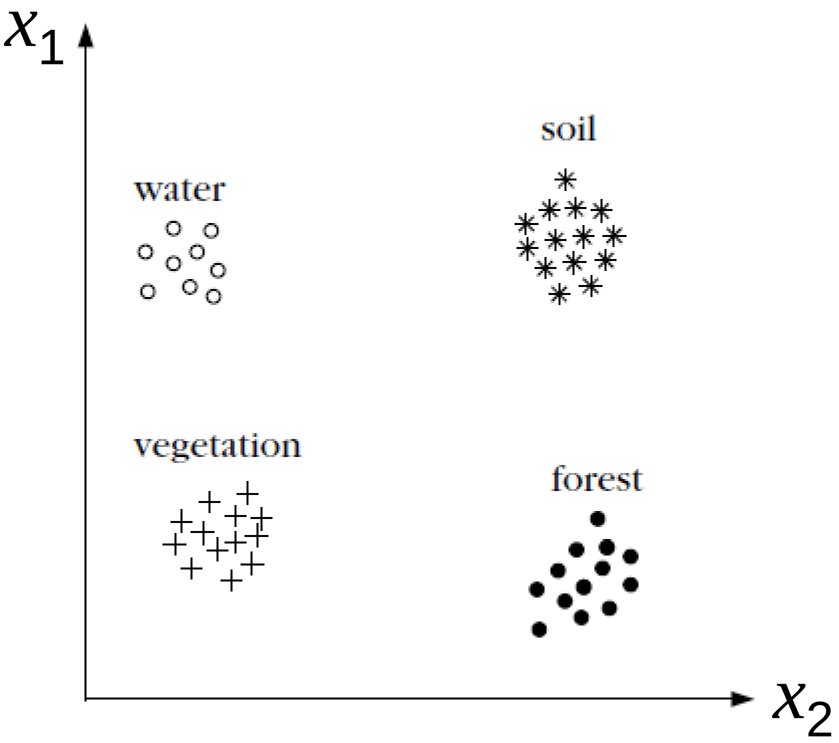
Unsupervised Learning





Unsupervised Learning





Bayesian Classifier and its Variants

Classification Example 1

- Given:
  - A doctor knows that meningitis causes stiff neck 50% of the time
  - one of every 50,000 persons has meningitis
  - one of every 20 persons has stiff neck

Classification Example 1

- Given:
  - A doctor knows that meningitis causes stiff neck 50% of the time
  - one of every 50,000 persons has meningitis
  - one of every 20 persons has stiff neck

- If a patient has stiff neck, does he/she has meningitis?

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

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 100K did not refund the loan previously

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 100K did not refund the loan previously
- *Can we trust him?*

Classification Example 3

- The sea bass/salmon example
  - We know the previous counts of salmon/sea bass
  - Can we predict which fish is coming in the conveyor?

Bayes Classifier

- A probabilistic framework for solving classification problems

- Bayes theorem:

$$P(A,C) = P(A)P(C \mid A) = P(A \mid C)P(C)$$

Bayes Classifier

- A probabilistic framework for solving classification problems

- Conditional Probabilities:

$$P(C \mid A) = \frac{P(A, C)}{P(A)}$$

$$P(A \mid C) = \frac{P(A, C)}{P(C)}$$

Example 1

- 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, does he/she has meningitis?

$$P(M \mid S) = \frac{P(S \mid M)P(M)}{P(S)} = \frac{0.5 \times 1/50000}{1/20} = 0.0002$$

Example 3

- The sea bass/salmon example
  - We know the previous counts of salmon/sea bass
    - amount of catches
    - *class* or *state of nature* (Salmon/Sea bass)
      state of nature is a random variable

Example 3

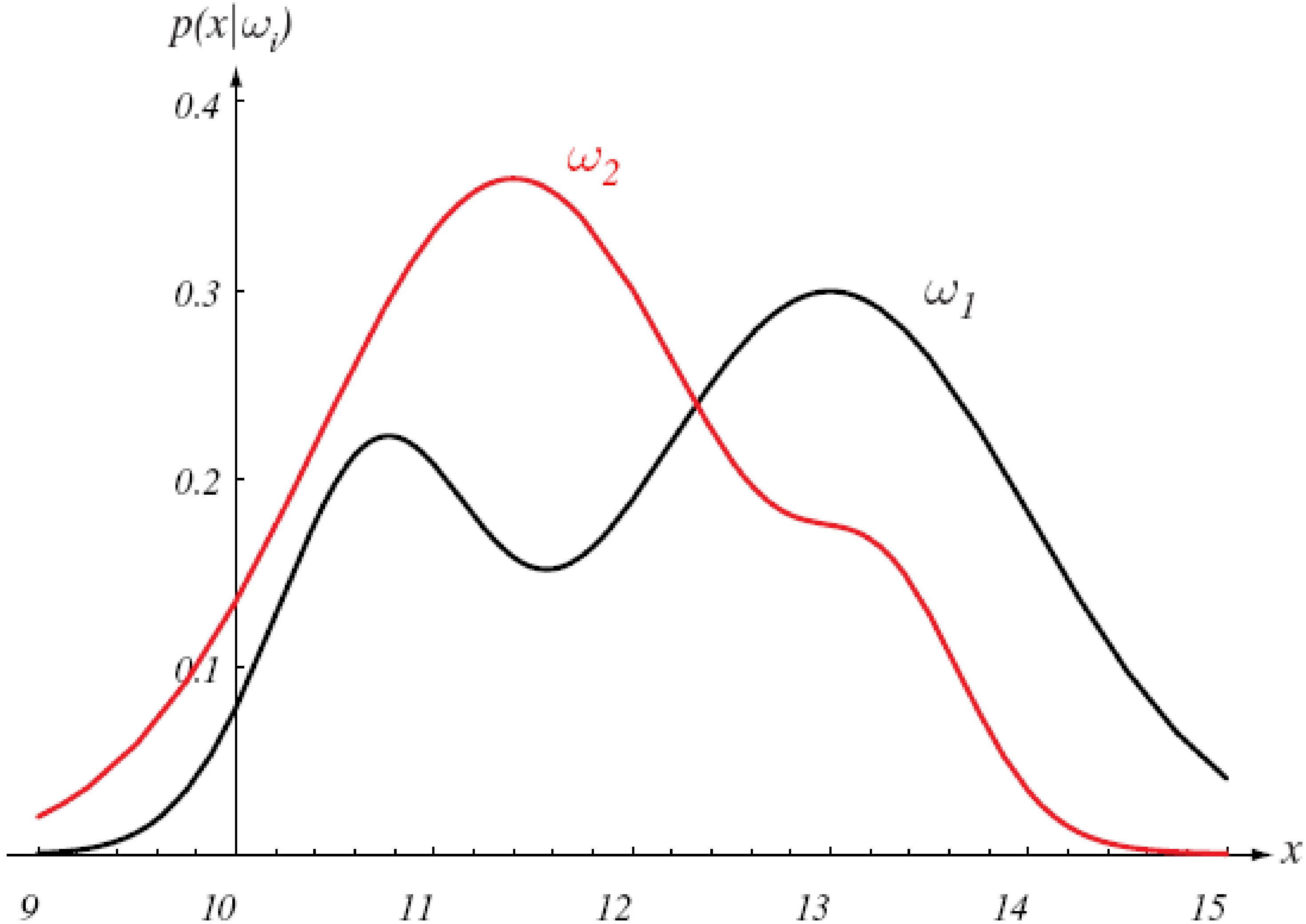
- The sea bass/salmon example
  - if the catch of salmon and sea bass is equi-probable
    - $P(\omega_1) = P(\omega_2)$ (uniform priors)
    - $P(\omega_1) + P(\omega_2) = 1$ (exclusivity and exhaustivity)

- Decision rule with only the prior information
  - Decide ω_1 if $P(\omega_1) > P(\omega_2)$ otherwise decide ω_2

- Decision rule with only the prior information
  - Decide ω_1 if $P(\omega_1) > P(\omega_2)$ otherwise decide ω_2
  - Misclassify many fishes

- Use of the class –conditional information
  - Use lightness

- $P(x \mid \omega_1)$ and $P(x \mid \omega_2)$ describe lightness in sea bass and salmon



Lightness

- Given the lightness evidence *x*, calculate *Posterior* from *Likelihood* and evidence

  - $$P(\omega_j \mid x) = \frac{p(x \mid \omega_j) P(\omega_j)}{p(x)}$$

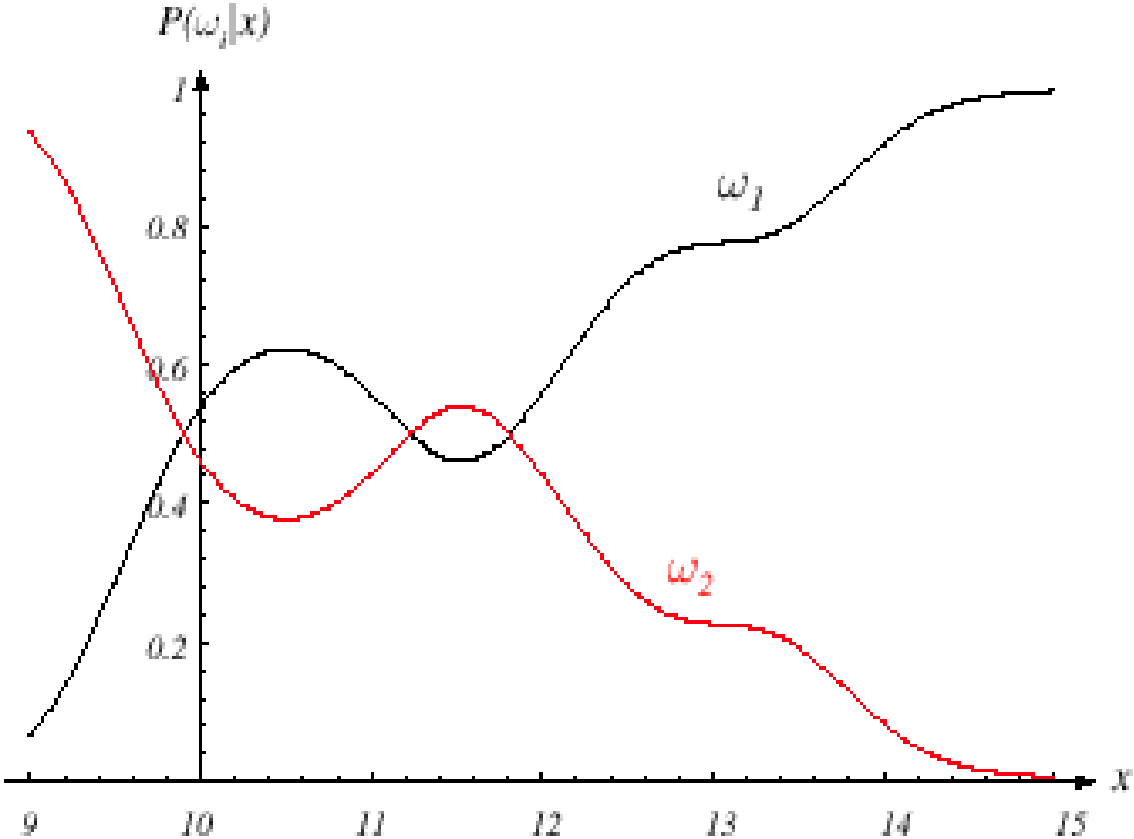
  - Posterior = (Likelihood. Prior) / Evidence

- Given the lightness evidence *x*, calculate *Posterior* from *Likelihood* and evidence

  - $$P(\omega_j \mid x) = \frac{p(x \mid \omega_j) P(\omega_j)}{p(x)}$$

    where in case of two categories

$$P(x) = \sum_{j=1}^{j=2} P(x \mid \omega_j) P(\omega_j)$$



Posterior function

- Decision given the posterior probabilities

X is an observation for which:

if $P(\omega_1 \mid x) > P(\omega_2 \mid x) \Rightarrow$ True state of nature = ω_1

if $P(\omega_1 \mid x) < P(\omega_2 \mid x) \Rightarrow$ True state of nature = ω_2

- Decision given the posterior probabilities

X is an observation for which:

if $P(\omega_1 \mid x) > P(\omega_2 \mid x) \Rightarrow$ True state of nature = ω_1
if $P(\omega_1 \mid x) < P(\omega_2 \mid x) \Rightarrow$ True state of nature = ω_2

Therefore:
whenever we observe a particular x, the probability of error is :

$$P(error \mid x) = P(\omega_1 \mid x) \text{ if we decide } \omega_2$$
$$P(error \mid x) = P(\omega_2 \mid x) \text{ if we decide } \omega_1$$

- Minimizing the probability of error

- Decide ω_1 *if* $P(\omega_1 \mid x) > P(\omega_2 \mid x)$*;*
  otherwise decide ω_2

  Therefore:

  $$P(error \mid x) = \min\,[P(\omega_1 \mid x), P(\omega_2 \mid x)]$$

  (Bayes decision)

Bayesian Decision Theory – Continuous Features

- Generalization of the preceding ideas
  - Use of more than one feature
  - Use more than two states of nature
  - Allowing actions and not only decide on the state of nature
  - Introduce a loss of function which is more general than the probability of error

Classification to Minimize Loss

Let $\{\omega_1, \omega_2, \ldots, \omega_m\}$ be the set of m states of nature (or "categories or classes")

Let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be the set of possible actions

Let $\lambda(\alpha_i \mid \omega_j)$ be the loss incurred for taking

action α_i when the state of nature is ω_j

Classification to Minimize Loss

The risk to take decision α_i is

$$R(\alpha_i \mid x) = \sum_{j=1}^{j=m} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid x)$$

Overall risk

$R = Sum\ of\ all\ R(\alpha_i \mid x)\ for\ i = 1,\ldots,n$

Minimizing R ⟺ Minimizing $R(\alpha_i \mid x)\ for\ i = 1, \ldots, n$

Classification to Minimize Loss

- Two-category classification

$$\alpha_1 : deciding\ \omega_1$$

$$\alpha_2 : deciding\ \omega_2$$

Classification to Minimize Loss

- Two-category classification

$$\alpha_1 : deciding\ \omega_1$$

$$\alpha_2 : deciding\ \omega_2$$

$$\lambda_{ij} = \lambda(\alpha_i \mid \omega_j)$$

loss incurred for deciding ω_i when the true state of nature is ω_j

Classification to Minimize Loss

- Two-category classification

$$\alpha_1 : deciding\ \omega_1$$

$$\alpha_2 : deciding\ \omega_2$$

$$\lambda_{ij} = \lambda(\alpha_i \mid \omega_j)$$

loss incurred for deciding ω_i when the true state of nature is ω_j

Conditional risk:

$$R(\alpha_1 \mid x) = \lambda_{11}P(\omega_1 \mid x) + \lambda_{12}P(\omega_2 \mid x)$$
$$R(\alpha_2 \mid x) = \lambda_{21}P(\omega_1 \mid x) + \lambda_{22}P(\omega_2 \mid x)$$

Classification to Minimize Loss

Our rule is the following:

$$\text{if } R(\alpha_1 \mid x) < R(\alpha_2 \mid x)$$

action α_1: "decide ω_1" is taken

Classification to Minimize Loss

Our rule is the following:

if $R(\alpha_1 \mid x) < R(\alpha_2 \mid x)$

action α_1: "decide ω_1" is taken

Now use these formula:

$$R(\alpha_1 \mid x) = \lambda_{11}P(\omega_1 \mid x) + \lambda_{12}P(\omega_2 \mid x)$$

$$R(\alpha_2 \mid x) = \lambda_{21}P(\omega_1 \mid x) + \lambda_{22}P(\omega_2 \mid x)$$

Classification to Minimize Loss

Our rule is the following:

if $R(\alpha_1 \mid x) < R(\alpha_2 \mid x)$

action α_1: "decide ω_1" is taken

This results in the equivalent rule :

decide ω_1 if:

$$(\lambda_{21} - \lambda_{11})\, P(x \mid \omega_1)\, P(\omega_1) > (\lambda_{12} - \lambda_{22})\, P(x \mid \omega_2)\, P(\omega_2)$$

and decide ω_2 otherwise

Classification to Minimize Loss

The preceding rule

$$(\lambda_{21} - \lambda_{11})\, P(x \mid \omega_1)\, P(\omega_1) >$$
$$(\lambda_{12} - \lambda_{22})\, P(x \mid \omega_2)\, P(\omega_2)$$

is equivalent to the following rule:

$$\text{if } \frac{P(x \mid \omega_1)}{P(x \mid \omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)}$$

Then take action α_1 (decide ω_1)
Otherwise take action α_2 (decide ω_2)

Classification to Minimize Loss

Likelihood ratio:

The preceding rule

$$(\lambda_{21}- \lambda_{11})\, P(x \mid \omega_1)\, P(\omega_1) > (\lambda_{12}- \lambda_{22})\, P(x \mid \omega_2)\, P(\omega_2)$$

is equivalent to the following rule:

$$\textit{if } \frac{P(x \mid \omega_1)}{P(x \mid \omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)}$$

Likelihood ratio

Then take action α_1 (decide ω_1)
Otherwise take action α_2 (decide ω_2)

Classification to Minimize Loss

Optimal decision property

“If the likelihood ratio exceeds a threshold value independent of the input pattern x, we can take optimal actions”

$$\frac{P(x|\omega_1)}{P(x|\omega_2)} > \boxed{\frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)}}$$

Minimum Error Rate Classification

Assume the loss function for two class case:

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \qquad i, j = 1, ..., c.$$

The Risk is now:

$$\begin{aligned} R(\alpha_i|\mathbf{x}) &= \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|\mathbf{x}) \\ &= \sum_{j \neq i} P(\omega_j|\mathbf{x}) \\ &= 1 - P(\omega_i|\mathbf{x}) \end{aligned}$$

Minimum Error Rate Classification

Assume the loss function for two class case:

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \qquad i, j = 1, \ldots, c.$$

The Risk is now:

$$\begin{aligned} R(\alpha_i|\mathbf{x}) &= \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|\mathbf{x}) \\ &= \sum_{j \neq i} P(\omega_j|\mathbf{x}) \\ &= 1 - P(\omega_i|\mathbf{x}) \end{aligned}$$

Minimizing *R* $\Leftrightarrow$ Minimizing $R(\alpha_i \mid x)$ $\Leftrightarrow$ Maximizing $P(\omega_i \mid x)$

Minimum Error Rate Classification

Our rule was:

if $R(\alpha_1 \mid x) < R(\alpha_2 \mid x)$

action α_1: "decide ω_1" is taken

which is equivalent to:

$$\text{Decide } \omega_i \text{ if } P(\omega_i|\mathbf{x}) > P(\omega_j|\mathbf{x}) \qquad \text{for all } j \neq i.$$