COSC 522 – Machine Learning

Baysian Decision Theory

Hairong Qi, Gonzalez Family Professor
Electrical Engineering and Computer Science
University of Tennessee, Knoxville
https://www.eecs.utk.edu/people/hairong-qi/
Email: hqi@utk.edu

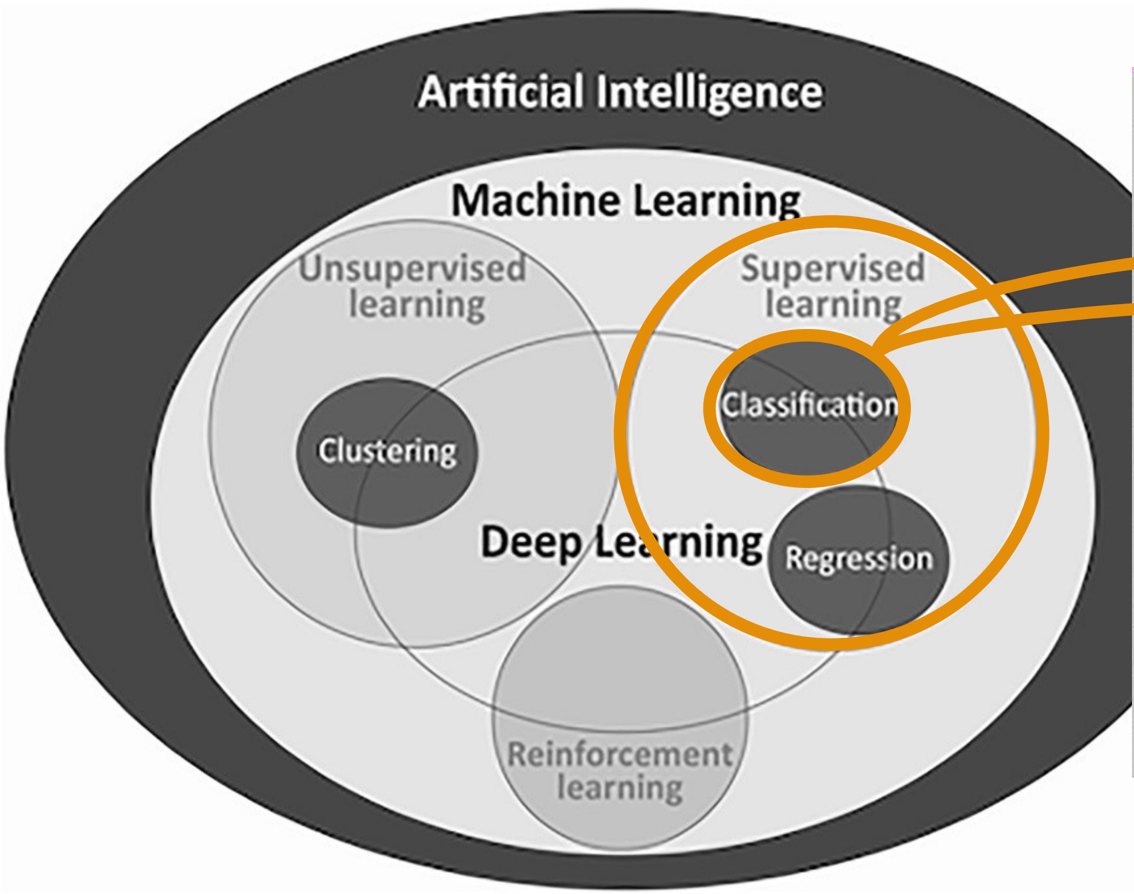
Outline/Questions

- Supervised vs. Unsupervised Learning
- Training set vs. Test set
- Features vs. Samples vs. Dimension
- Classification vs. Regression
- Parametric Learning vs. Non-parametric Learning

- What is pdf? pdf vs. histogram?
- What is Bayes' Formula?
- What is the difference between probability and pdf?
- What is the role of "evidence"?
- In Bayes' Formula, what is conditional pdf? Prior probability? Posterior probability?
- What does the normalization factor (or evidence) do?

- What is Baysian decision rule? or MPP?
- What are decision regions?
- How to calculate conditional probability of error and overall probability of error?
- What are cost function (or objective function) and optimization method?

Where Are We?



	Part 1: Statistical Methods
	Baysian Learning
08/20 (T)	Introduction
08/22 (R)	Baysian Decision Theory and Parametric Learning
08/27 (T)	Non-Parametric Learning
08/29 (R)	ML with Python (taught by TA)
09/03 (T)	Recap
09/05 (R)	Homework and Project Discussion (taught by TA)
	Neural Networks
09/10 (T)	Biological Neuron and Perceptron
09/12 (R)	Back Propagation and Gradient Descent
09/17 (T)	Kernel Methods
09/19 (R)	Support Vector Machine
09/24 (T)	SVM

M. Mafu, “Advances in artificial intelligence and machine learning for quantum communication applications,” IET Quantum Communication, 2024, DOI: 10.1049/qtc2.12094

Reading Assignment in HW1

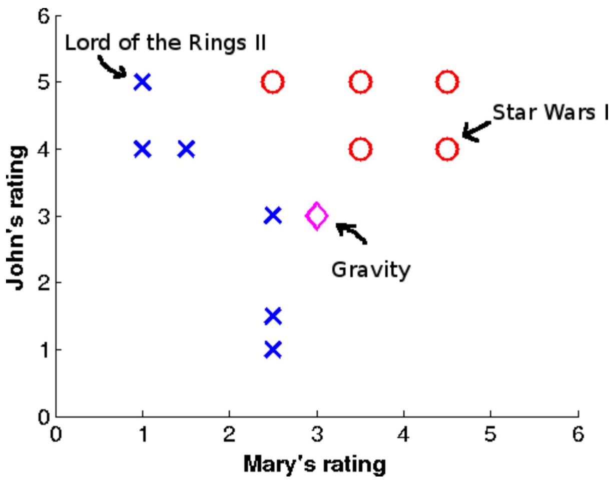
AICIP RESEARCH

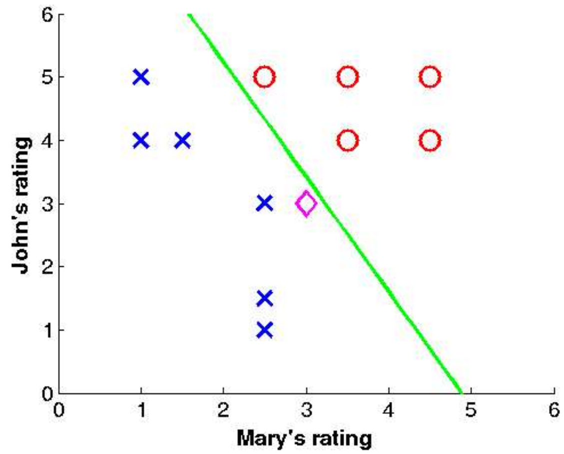
- Leo Breiman, "Statistical modeling: The two cultures," *Statistical Science* 16(3):199-231, 2001.

Revisiting the Toy Example – An Intuitive Solution

AICIP RESEARCH

Movie name	Mary's rating	John's rating	I like?
Lord of the Rings II	1	5	No
...	...	...	...
Star Wars I	4.5	4	Yes
Gravity	3	3	?





- Supervised learning:
  - Training data vs. testing data
  - Training: given input-output pairs
- Features
- Samples
- Dimensions

Bayes' Formula (Bayes' Rule)

AICIP RESEARCH

conditional probability density function (pdf) or "likelihood"

from training data

prior probability (*a-priori* probability)

from domain knowledge

$$P\left(w_j|x\right) = \frac{p\left(x|w_j\right)P\left(w_j\right)}{p(x)}$$

j index for different classes
w_j: different classes
x: training sample

posterior probability (*a-posteriori* probability)

$$p(x) = \sum_{j=1}^{c} p\left(x|w_j\right)P\left(w_j\right)$$

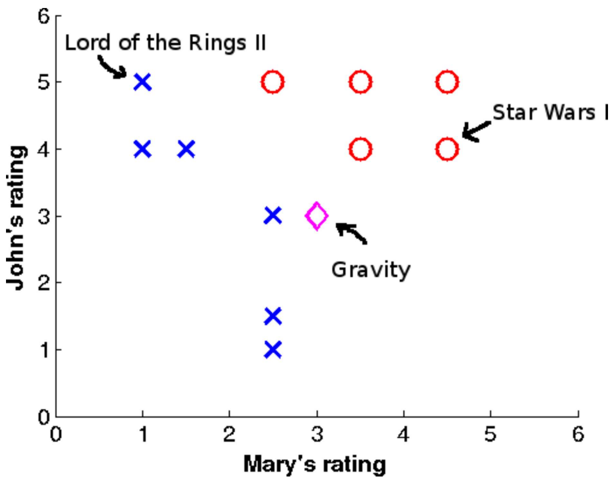
normalization constant (evidence)

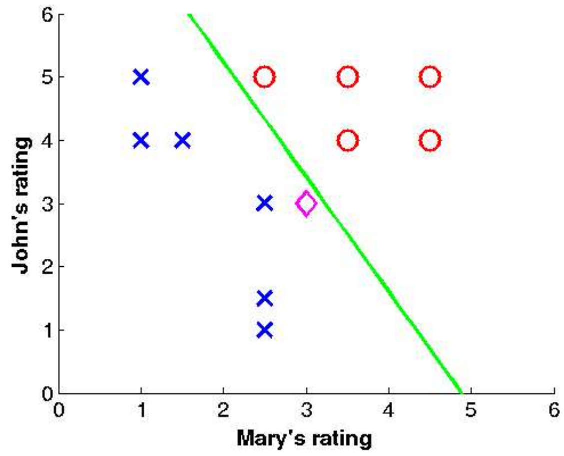
How do They Apply to the Toy Example

AICIP RESEARCH

Movie name	Mary's rating	John's rating	I like?
Lord of the Rings II	1	5	No
...	...	...	...
Star Wars I	4.5	4	Yes
Gravity	3	3	?

$$P\left(w_j | x\right) = \frac{p\left(x | w_j\right) P\left(w_j\right)}{p(x)}$$



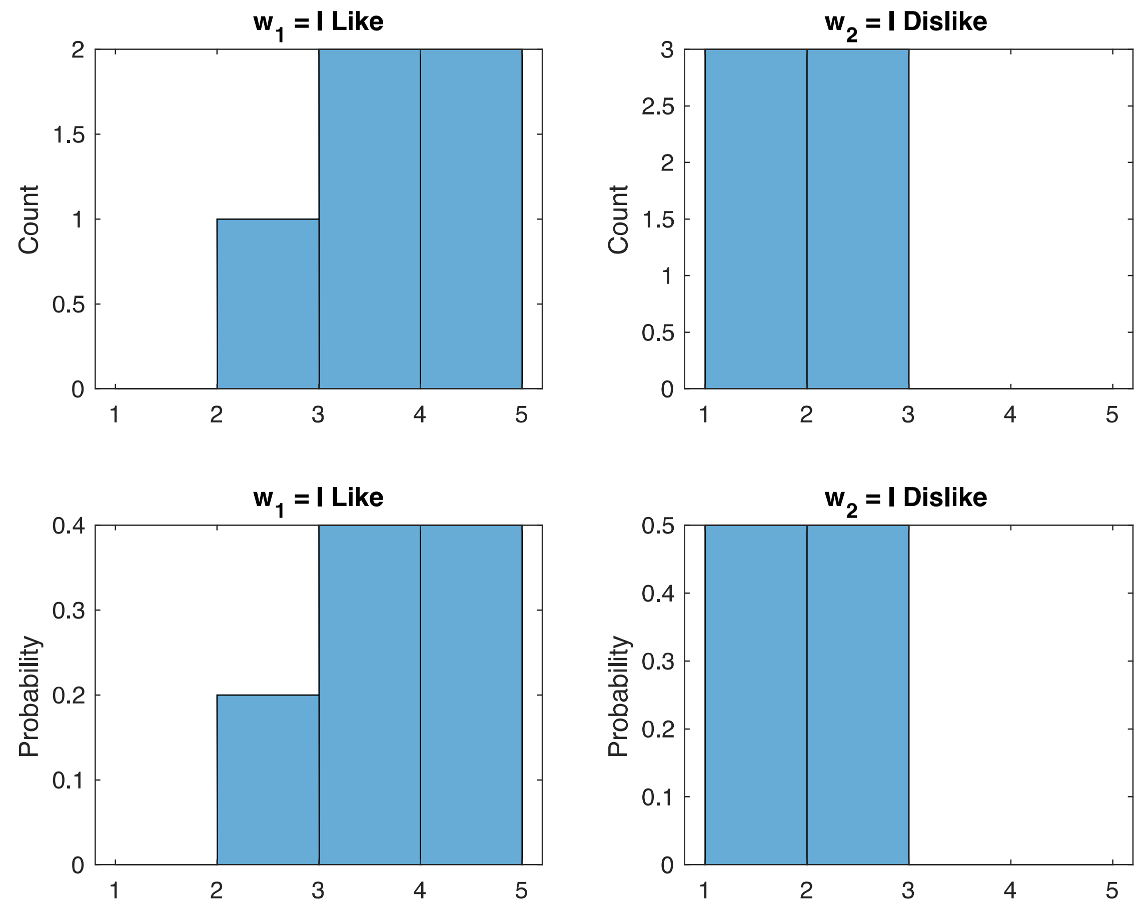


$$P(w_j|x) = \frac{p(x|w_j)P(w_j)}{p(x)}$$

Toy Example Reduced to 1-D Feature (Mary's rating) from histogram to probability

```
Training Set:
    Movies Mary rated that I liked
    x1 = [2.5
          3.5
          3.5
          4.5
          4.5]
    Movies Mary rated that I disliked
    x2 = [1
          1
          1.5
          2.5
          2.5
          2.5]

Test Set:
    x = [3]
    Will I like it?
```



$$P\left(w_j|x\right) = \frac{p\left(x|w_j\right)P\left(w_j\right)}{p\left(x\right)}$$

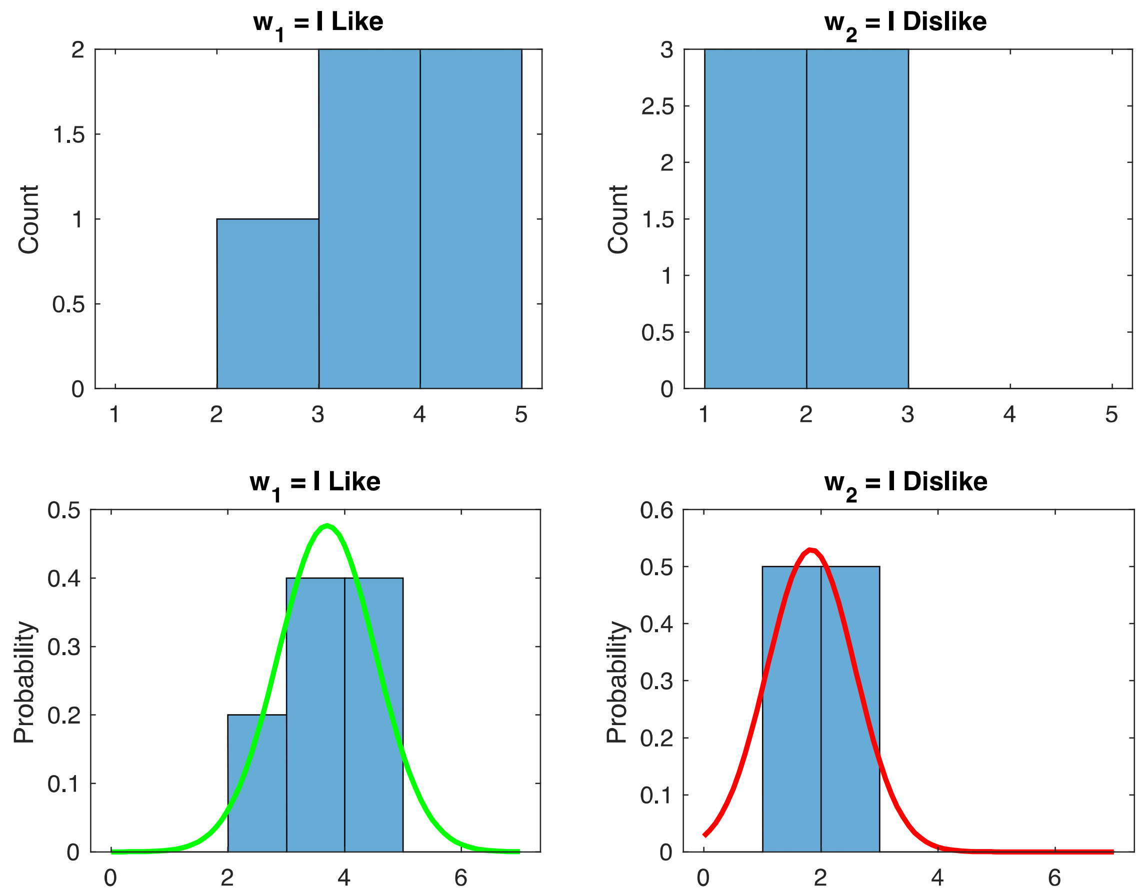
Toy Example Reduced to 1-D Feature (Mary's rating) from probability to model fitting (Gaussian)

Training Set:

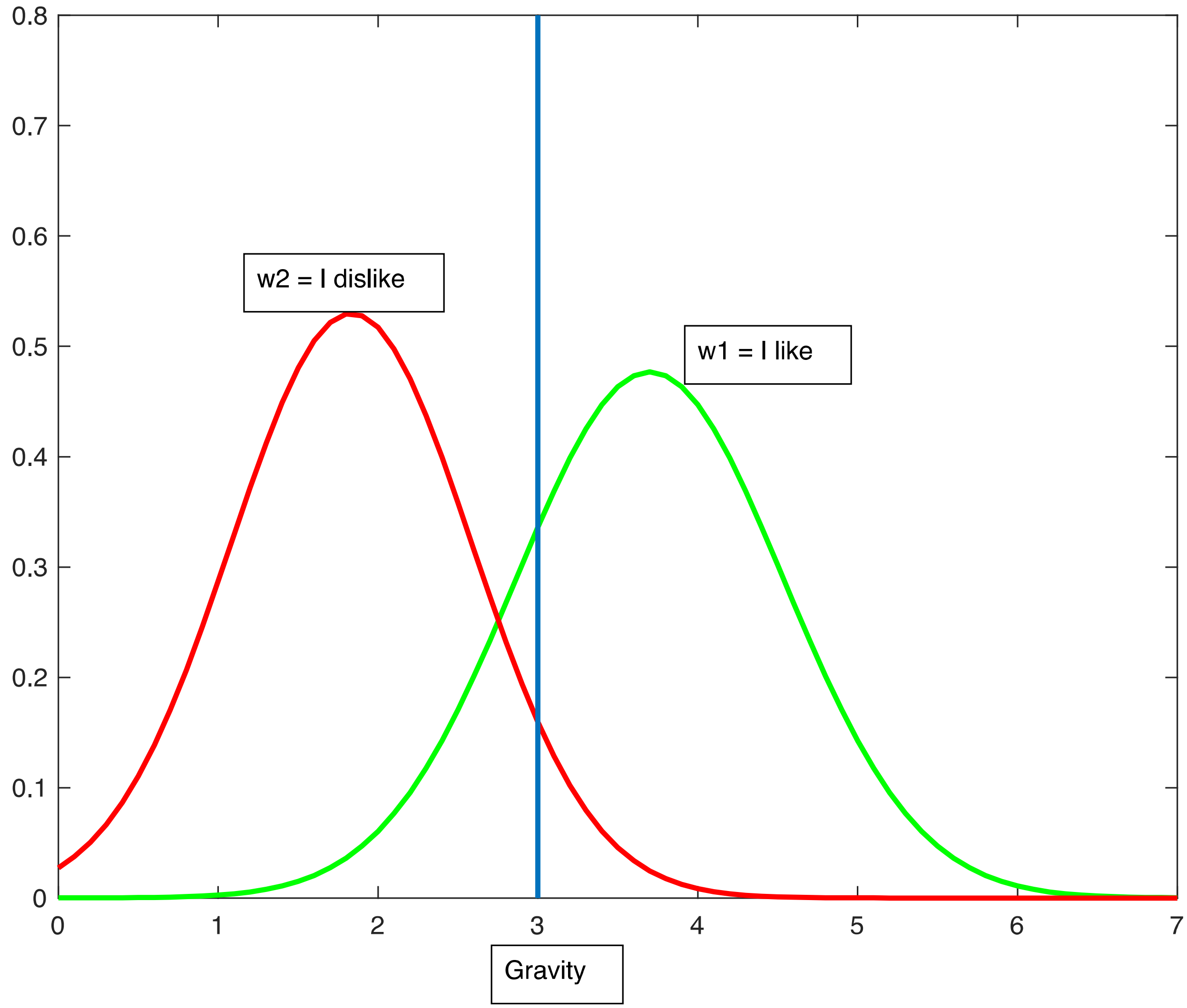
```
    Movies Mary rated that I liked
    x1 = [2.5
          3.5
          3.5
          4.5
          4.5]
    Movies Mary rated that I disliked
    x2 = [1
          1
          1.5
          2.5
          2.5
          2.5]
```

Test Set:

```
    x = [3]
    Will I like it?
```



Toy Example Reduced to 1-D Feature (Mary's rating)
from model fitting to decision boundary



AICIP RESEARCH

A Snippet to Discriminant Functions – Minimum Distance Classifier (Assumptions)

- 1. pdf is Gaussian

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right]$$

- 2. equal prior probability

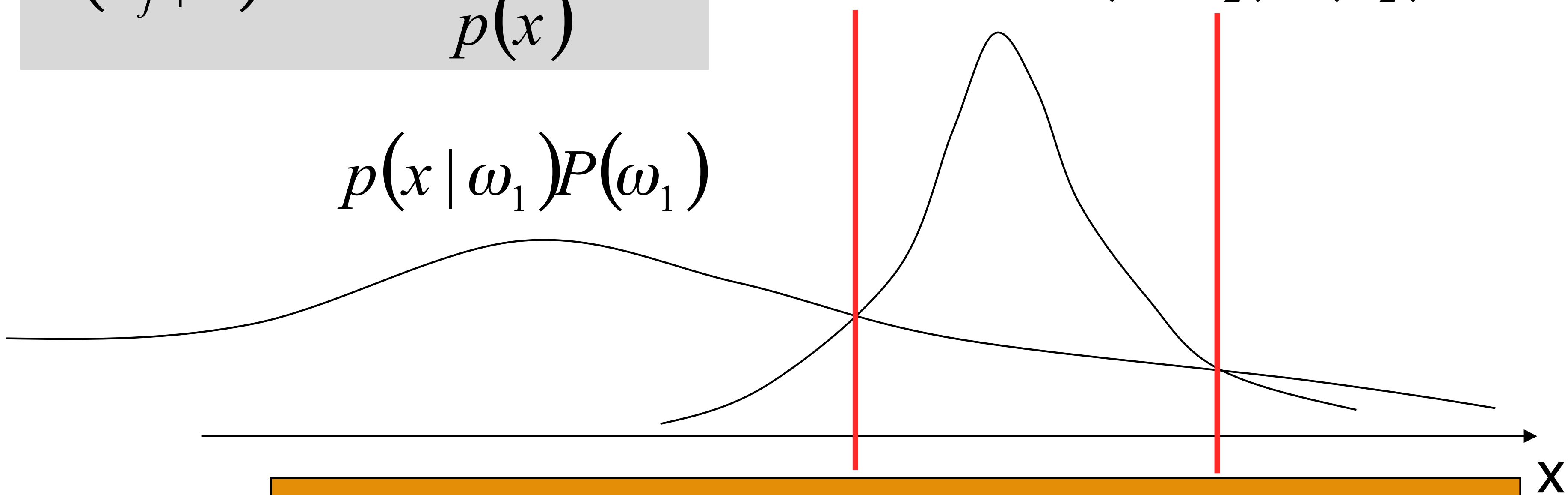
$$P(w_1) = P(w_2)$$

- 3. equal standard deviation

$$\sigma_1 = \sigma_2$$

Bayes Decision Rule

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



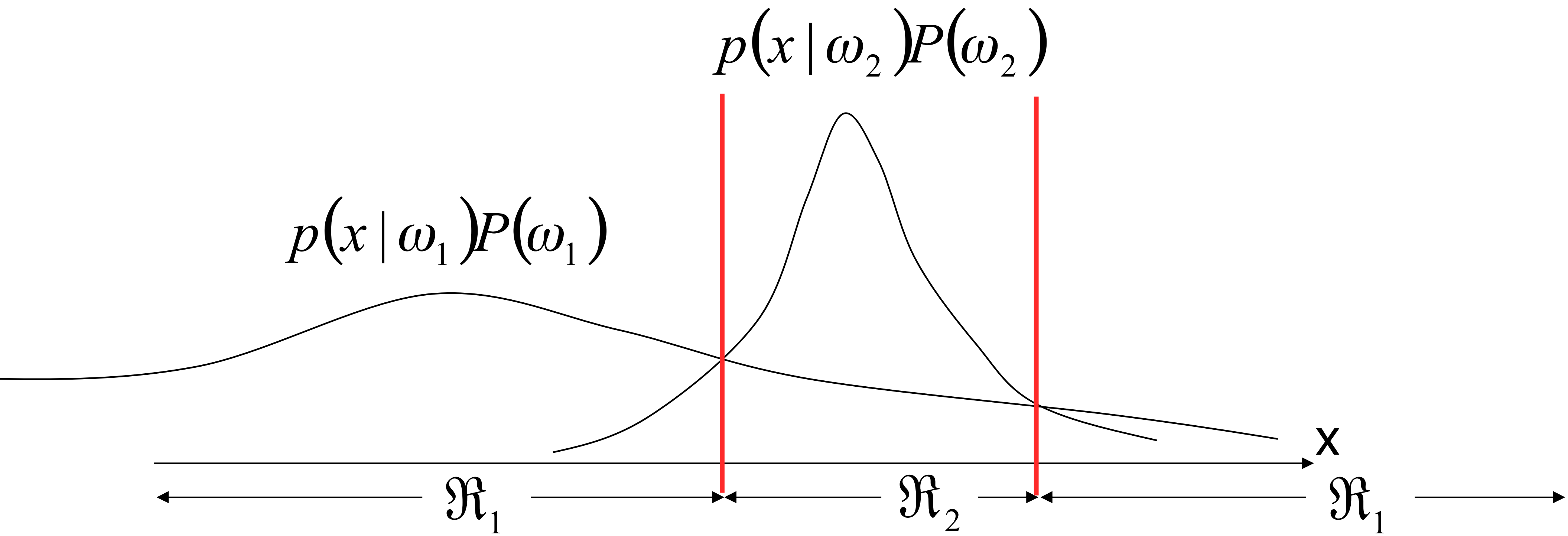
Maximum Posterior Probability (MPP):

For a given x, if $P(\omega_1 \mid x) > P(\omega_2 \mid x)$,

then x belongs to class 1, otherwise, 2.

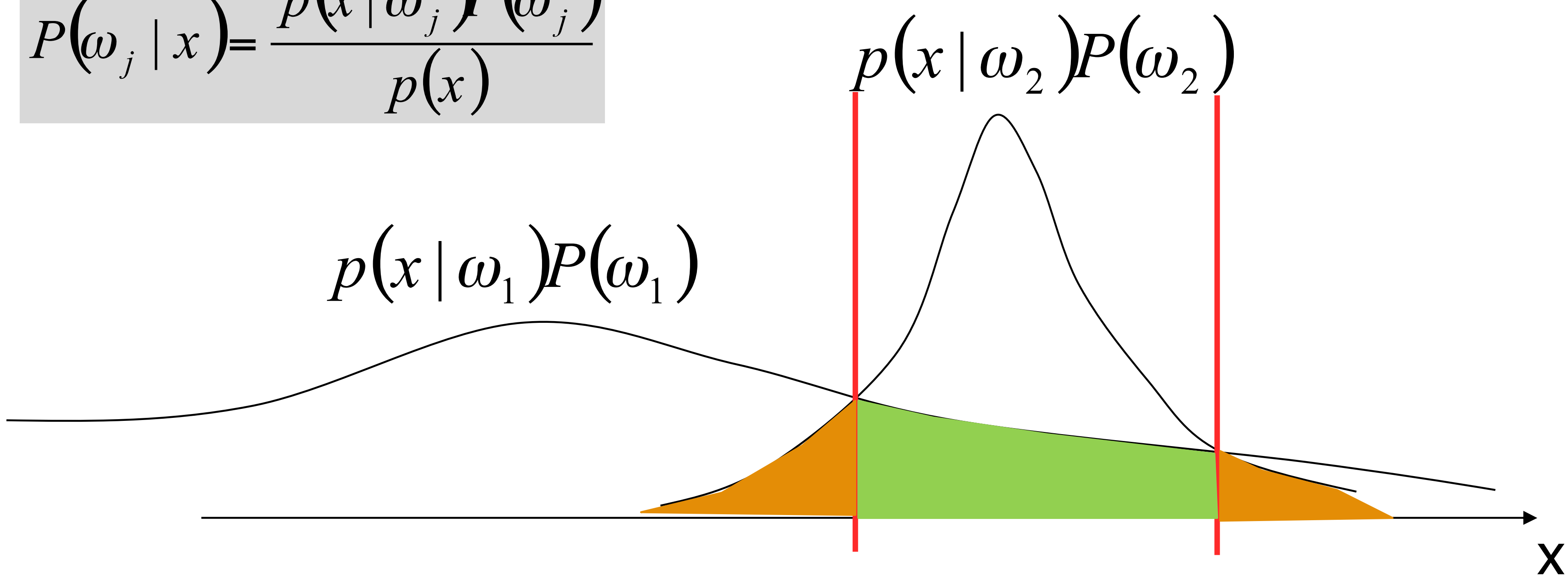
Decision Regions

The effect of any decision rule is to partition the feature space into *c* decision regions $\Re_1, \Re_2, \cdots, \Re_c$



Conditional Probability of Error

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



$$P(error \mid x) = \begin{cases} P(\omega_1 \mid x) & \text{if we decide } \omega_2 \\ P(\omega_2 \mid x) & \text{if we decide } \omega_1 \end{cases} = \min[P(\omega_1 \mid x), P(\omega_2 \mid x)]$$

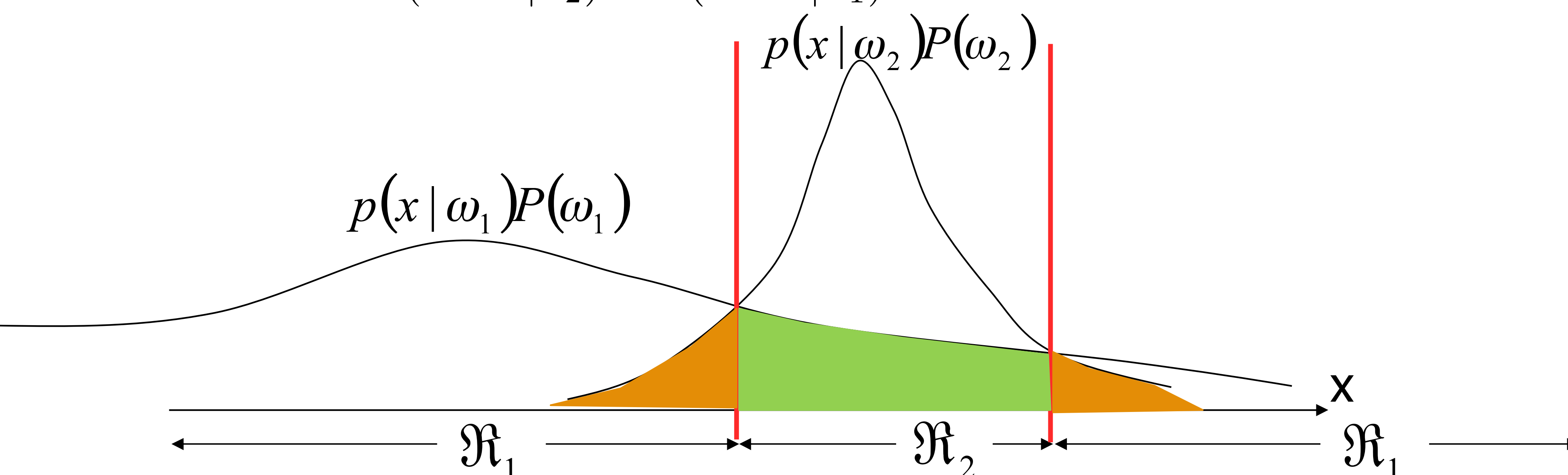
Overall Probability of Error

Or unconditional risk, unconditional probability of error

$$P(error) = \int_{-\infty}^{\infty} P(error, x)\,dx = \int_{-\infty}^{\infty} P(error \mid x)\,p(x)\,dx$$

$$P(error) = \int_{\Re_1} P(\omega_2 \mid x)\,p(x)\,dx + \int_{\Re_2} P(\omega_1 \mid x)\,p(x)\,dx$$

$$= P(error|\omega_2) + P(error|\omega_1)$$



Q&A Session

AICIP RESEARCH

- What is the cost function?
- What is the optimization approach we use to find the optimal solution to the cost function?

Theme 1: Cost functions and Optimization approaches

Recap

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

Maximum Posterior Probability

For a given x, if $P(\omega_1 \mid x) > P(\omega_2 \mid x)$, then x belongs to class 1, otherwise, 2.

Overall probability of error

$$P(error) = \int_{\Re_1} P(\omega_2 \mid x)p(x)dx + \int_{\Re_2} P(\omega_1 \mid x)p(x)dx$$

- Bayes decision rule → maximum posterior probability (MPP)
- Decision regions → How to calculate the overall probability of error