THE UNIVERSITY OF TENNESSEE KNOXVILLE

AICIP RESEARCH

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

Course Website: http://web.eecs.utk.edu/~hqi/cosc522/

Questions

- What is supervised learning (vs. unsupervised learning)?
- What is the difference between the training set and the test set?
- What is the difference between classification and regression?
- What are features and samples?
- What is dimension?
- What is histogram?
- What is pdf?
- What is Bayes' Formula?
- What is conditional pdf?
- What is the difference between prior probability and posterior probability?
- 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?

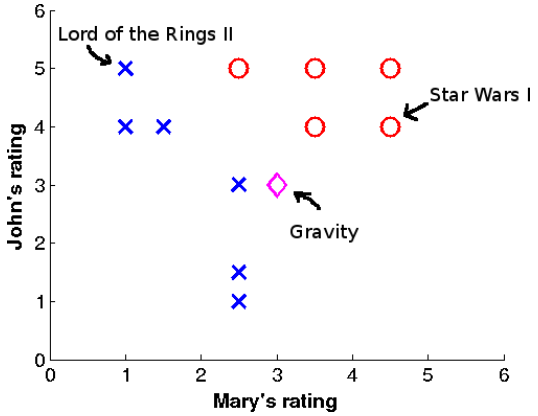
terminologies

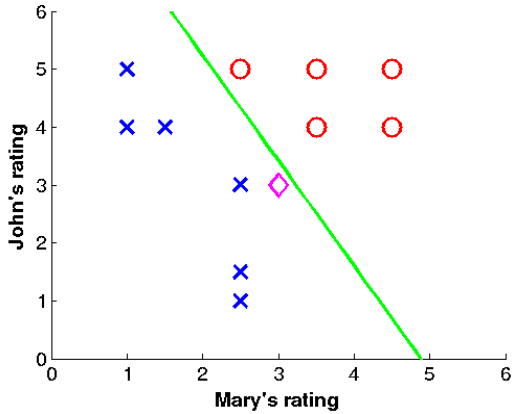
the Formula

decision rule

The Toy Example 1

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	?





The Toy Example 2

AICIP RESEARCH

- Student (taking COSC522 in F23) covid test
- Feature: temperature (1-D)
- Data collection: For the entire class, we take temperature of each student; also ask the student to take a covid test
- Data:
  - Training set: For half of the class, use temperature measurement as “feature”, and their test result as “label”
  - Testing set: For the other half of the class, given temperature information, determine if the student might have covid or not
- Question: Why do we need to ask students in test set to take covid test but didn’t use that test results?

Terminologies

AICIP RESEARCH

- Supervised learning:
  - Training data vs. testing data vs. validation data
  - Training: given input-output pairs
- Features (e.g., temperature)
- Samples
- Dimensions
- Classification vs. Regression

Questions

- What is supervised learning (vs. unsupervised learning)?
- What is the difference between the training set and the test set?
- What is the difference between classification and regression?
- What are features and samples?
- What is dimension?
- What is histogram?
- What is pdf?
- What is Bayes' Formula?
- What is conditional pdf?
- What is the difference between prior probability and posterior probability?
- 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?

terminologies

the Formula

decision rule

Example 1 – 1-D feature

AICIP RESEARCH

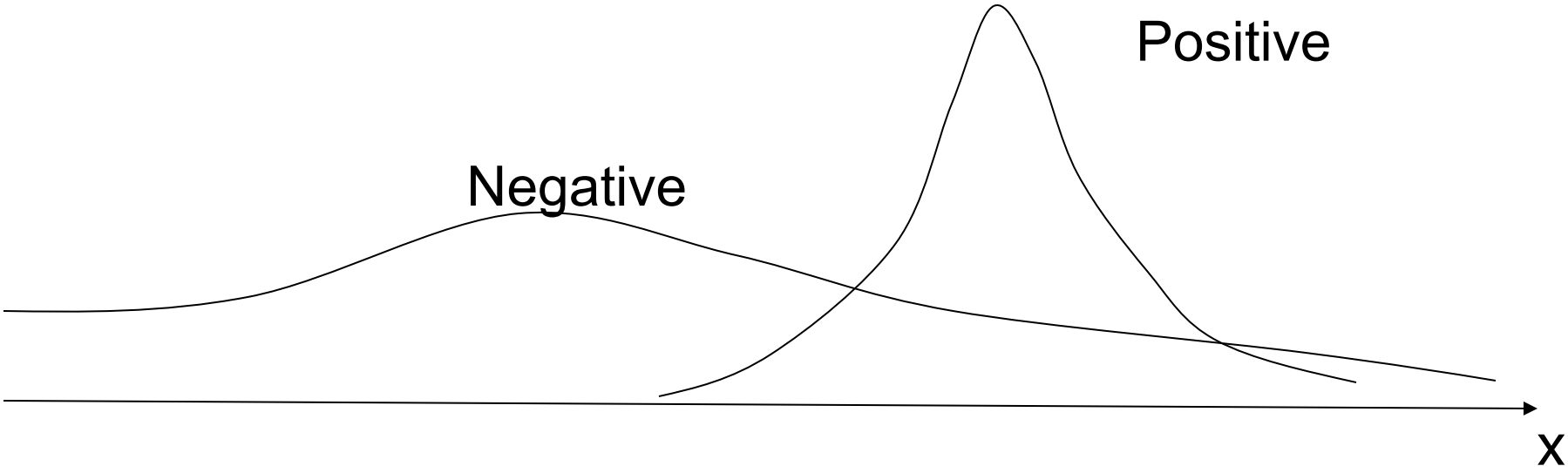
Rating	label
3.5	Y
4.8	N
3.4	Y
3.7	N
4.5	Y
4.8	N
3.6	Y
2.7	N
1	N

Example 2 – covid testing

AICIP RESEARCH

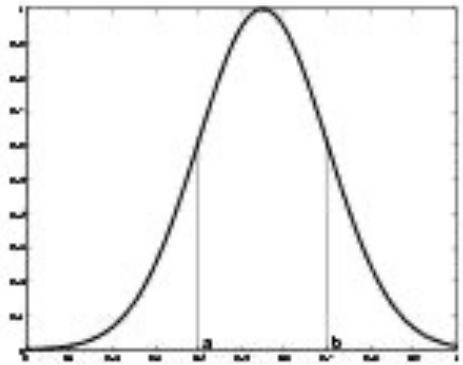
temperature	label
92	N
90	N
100	Y
102	Y
90	Y
101	N
93	N
95	N
103	Y

From Histogram to Probability Density Distribution (pdf)

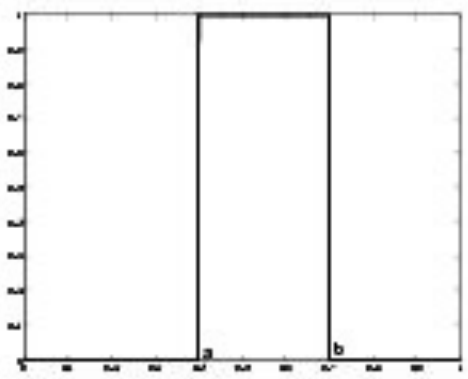


Examples of pdf

- Gaussian distribution
  - Bell curve
  - Normal distribution



- Uniform distribution



Q&A Session - Looking into Gaussian

AICIP RESEARCH

- Two classes with one intersection?
- Two classes with no intersection?
- Two classes with two intersections?

Bayes' Formula (Bayes' Rule)

AICIP RESEARCH

From domain knowledge

prior probability
(*a-priori* probability)

Conditional probability density function
(likelihood)

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

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

posterior probability
(*a-posteriori* probability)

normalization constant
(evidence)

Q&A Session

AICIP RESEARCH

- How do you interpret prior probability in the toy example?

Questions

AICIP RESEARCH

- What is supervised learning (vs. unsupervised learning)?
- What is the difference between the training set and the test set?
- What is the difference between classification and regression?
- What are features and samples?
- What is dimension?
- What is histogram?
- What is pdf?
- What is Bayes' Formula?
- What is conditional pdf?
- What is the difference between prior probability and posterior probability?
- 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?

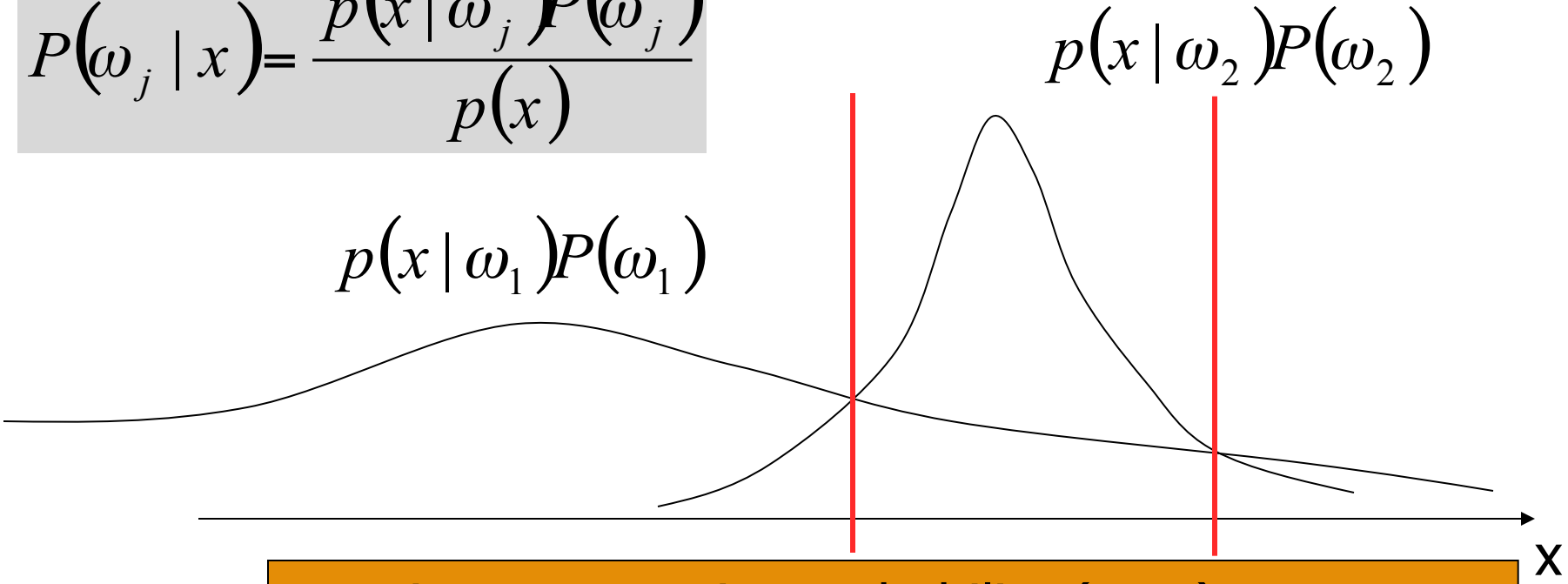
terminologies

the Formula

decision rule

Bayes Decision Rule

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



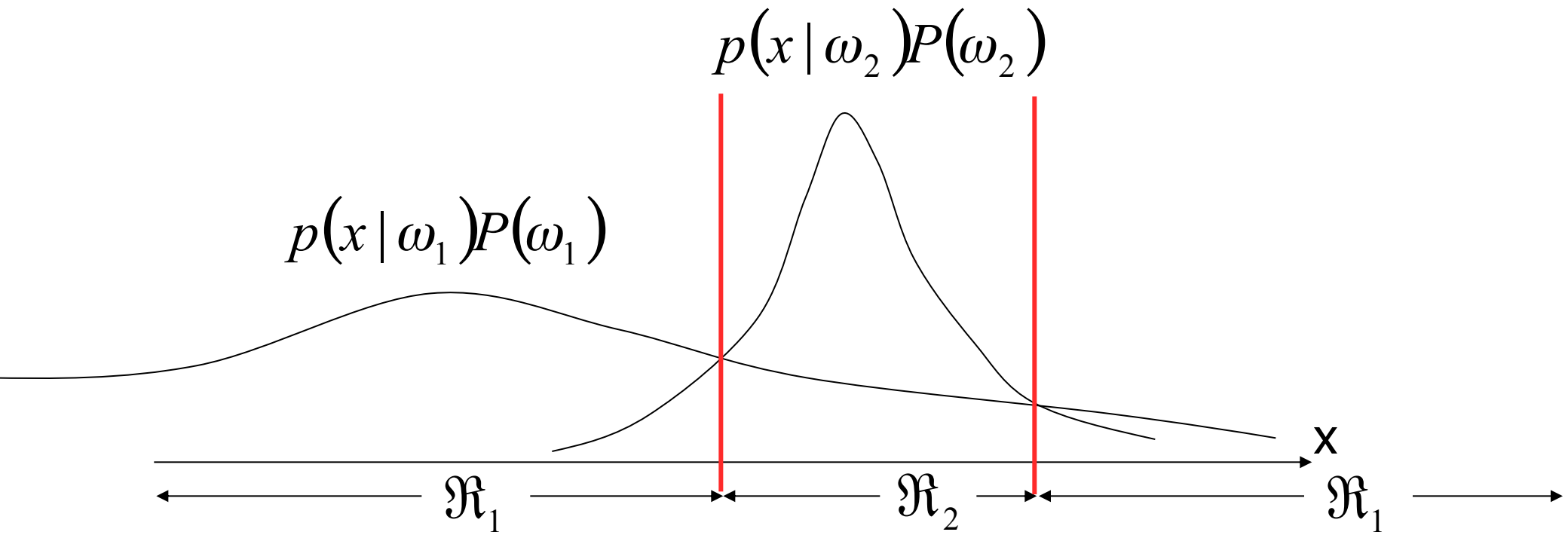
Maximum Posterior Probability (MPP):

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

then x belongs to class 1, otherwise, 2.

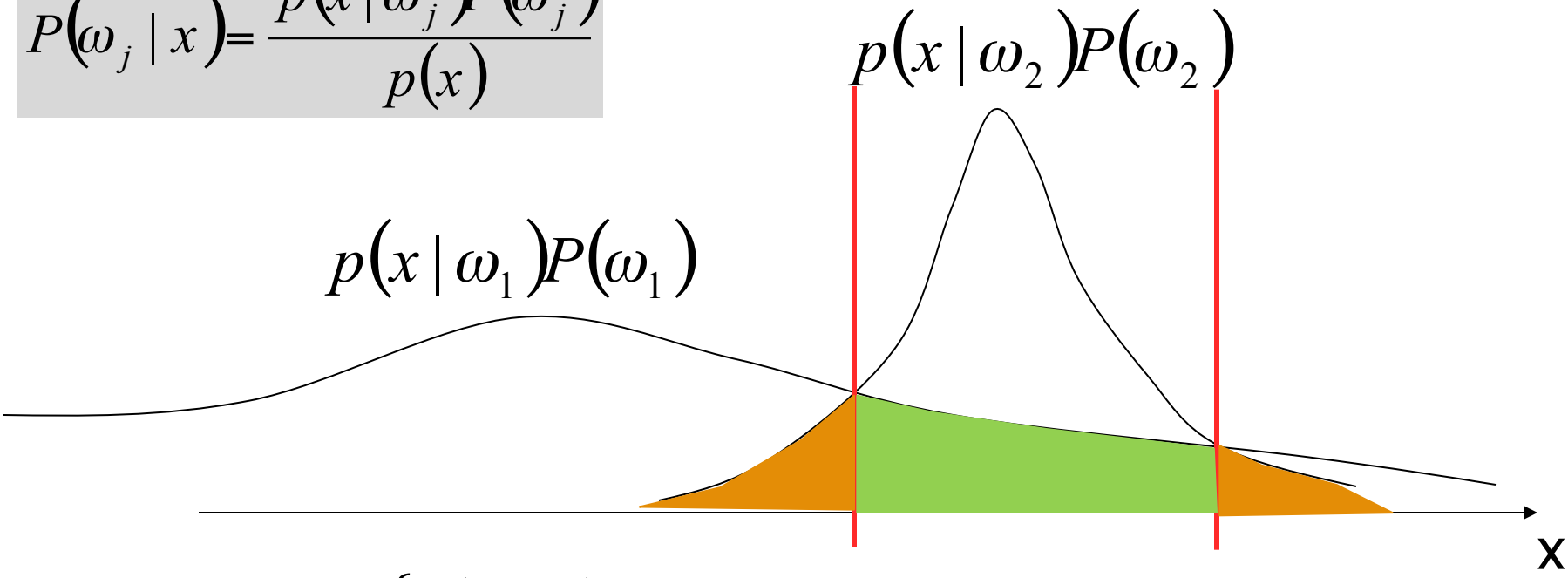
Decision Regions

- 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

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



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

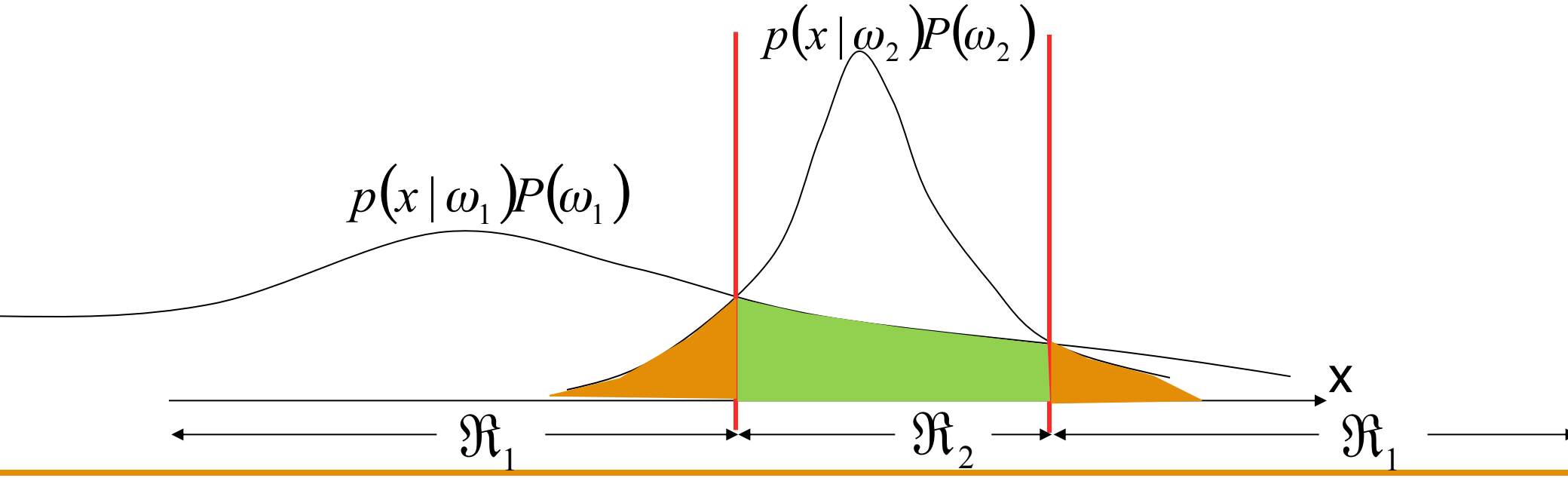
Overall Probability of Error

Or unconditional risk, unconditional probability of error

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

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

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



How Does It Work Altogether?

AICIP RESEARCH

temperature	label
92	N
90	N
100	Y
102	Y
90	Y
101	N
93	N
95	N
103	Y

Questions

AICIP RESEARCH

- What is supervised learning (vs. unsupervised learning)?
- What is the difference between the training set and the test set?
- What is the difference between classification and regression?
- What are features and samples?
- What is dimension?
- What is histogram?
- What is pdf?
- What is Bayes' Formula?
- What is conditional pdf?
- What is the difference between prior probability and posterior probability?
- 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?

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