THE UNIVERSITY OF TENNESSEE KNOXVILLE

AICIP RESEARCH

COSC 522 – Machine Learning

Introduction

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/

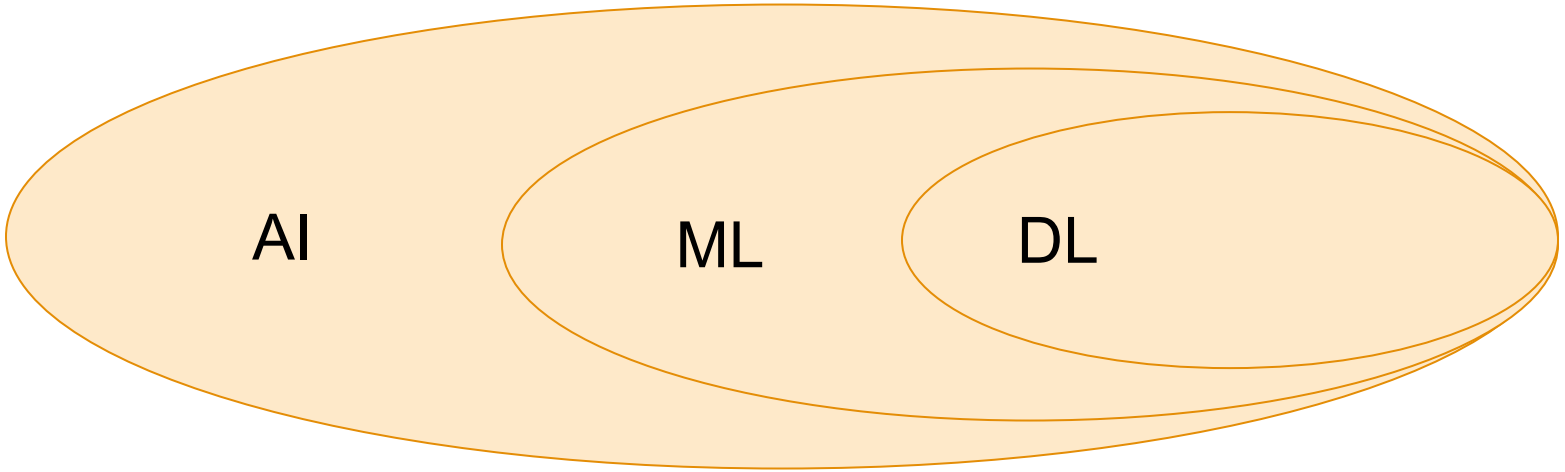
Questions

AICIP RESEARCH

- Q1: Exactly what are the differences between
  - Machine Learning (ML)
  - Deep Learning (DL)
  - Artificial Intelligence (AI)
- Q2: What are the differences of the related courses offered at UT?
- Q3: Do I have enough background?
- Q4: What are we going to learn in this class?
- Q5: Course policy
- Q6: What's the difference between this class and some very popular online ML lectures?

Q1: AI vs. ML vs. DL

AICIP RESEARCH



A Comparison between AI and ML Contents

Artificial Intelligence: A Modern Approach

(Fourth edition, 2020)

by Stuart Russell and Peter Norvig

The leading textbook in Artificial Intelligence, used in over **1400** schools in over **120** countries.

Table of Contents

http://aima.cs.berkeley.edu/

AICIP RESEARCH

A Bit of History of AI Development

AICIP RESEARCH

- 1956-1976
  - 1956, The Dartmouth Summer Research Project on Artificial Intelligence, organized by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon

> We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College ... The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

  - The rise of symbolic methods, systems focused on limited domains, deductive vs. inductive systems
  - 1973, the Lighthill report by James Lighthill, “Artificial Intelligence: A General Survey” - automata, robotics, neural network
  - 1976, the AI Winter
- 1976-2006
  - 1986, BP algorithm
  - ~1995, The Fifth Generation Computer
- 2006-???
  - 2006, Hinton (U. of Toronto), Bingio (U. of Montreal), LeCun (NYU)
  - 2012, ImageNet by Fei-Fei Li (2010-2017) and AlexNet

https://en.wikipedia.org/wiki/Dartmouth_workshop
https://en.wikipedia.org/wiki/Lighthill_report

Questions

AICIP RESEARCH

- Q1: Exactly what are the differences between
  - Machine Learning (ML)
  - Deep Learning (DL)
  - Artificial Intelligence (AI)
- Q2: What are the differences of the related courses offered at UT?
- Q3: Do I have enough background?
- Q4: What are we going to learn in this class?
- Q5: Course policy
- Q6: What's the difference between this class and some very popular online ML lectures?

Q2: Different Courses at UT

AICIP RESEARCH

- Intro to Machine Learning (COSC425), Fall
- Machine Learning (COSC522) ← Pattern Recognition (ECE471/571), Fall
- Artificial Intelligence (COSC523), Fall
- Natural Language Processing (COSC 524), Fall
- Deep Learning (COSC525), Spring
- Data Mining (COSC526), Spring
- Bioinspired Computation (COSC527), Spring
- Reinforcement Learning (ECE517), Fall
- Special Topic Class: Adversarial Learning (ECE599), Spring

Questions

AICIP RESEARCH

- Q1: Exactly what are the differences between
  - Machine Learning (ML)
  - Deep Learning (DL)
  - Artificial Intelligence (AI)
- Q2: What are the differences of the related courses offered at UT?
- Q3: Do I have enough background?
- Q4: What are we going to learn in this class?
- Q5: Course policy
- Q6: What's the difference between this class and some very popular online ML lectures?

Q3: Pre-/Co-requisites

AICIP RESEARCH

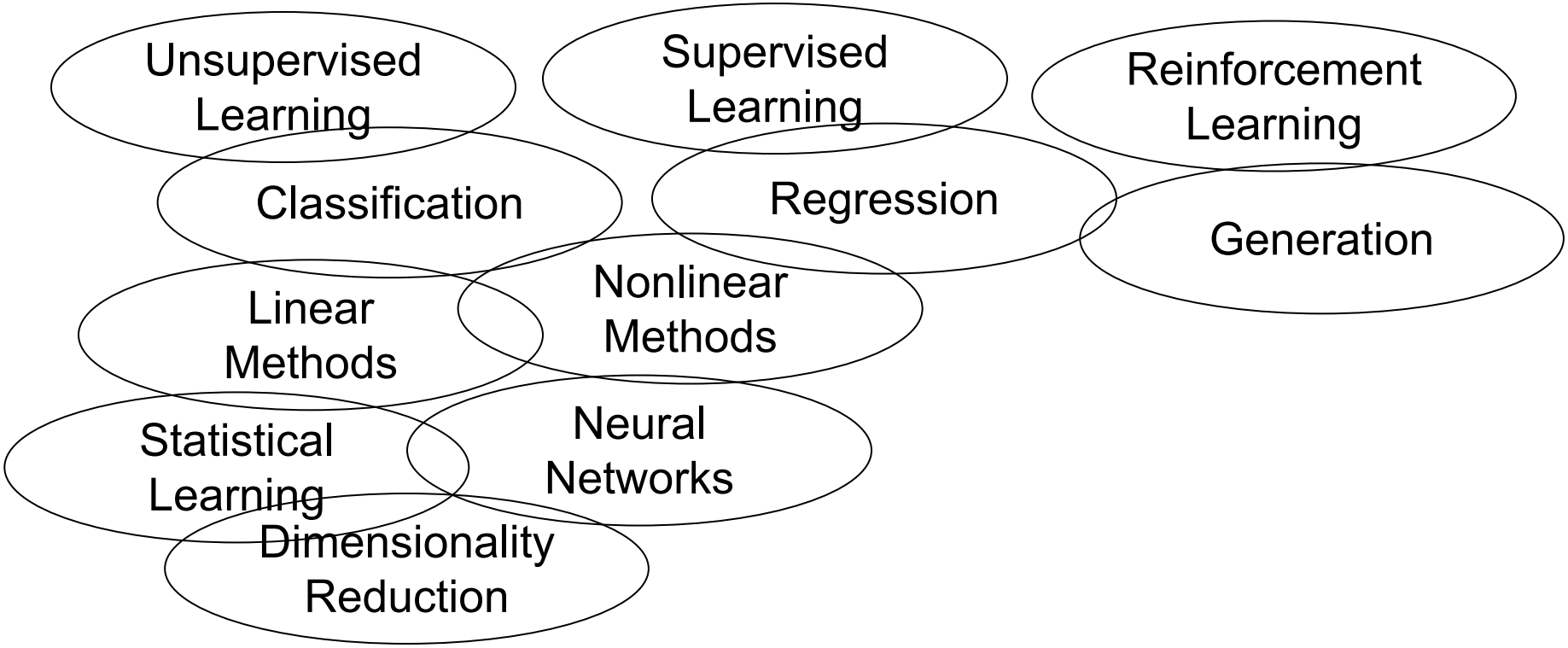
- Probability
- Linear Algebra
- Multivariate Calculus
- Python Programming
  - Jupyter notebook
  - colab

Questions

AICIP RESEARCH

- Q1: Exactly what are the differences between
  - Machine Learning (ML)
  - Deep Learning (DL)
  - Artificial Intelligence (AI)
- Q2: What are the differences of the related courses offered at UT?
- Q3: Do I have enough background?
- Q4: What are we going to learn in this class?
- Q5: Course policy
- Q6: What's the difference between this class and some very popular online ML lectures?

Q4: Topics Covered



Solve Real-World Problems!!!

Syllabus

Date	Topic	Reading	Assignment
R 08/24	Introduction		
T 08/29	Supervised Learning - Baysian Decision Theory Last Day to Drop without "W"		Project 1 - Supervised Learning (Due 09/19) HW1 (Due 09/07)
R 08/31	Supervised Learning - Baysian Decision Theory		
T 09/05	Supervised Learning - Discriminant Functions		
R 09/07	Supervised Learning - Discriminant Functions		
T 09/12	Supervised Learning - Non-parametric Learning (kNN)		HW2 (Due 09/21)
R 09/14	Supervised Learning - Non-parametric Learning (kNN)		
T 09/19	Dimensionality Reduction		Project 2 - Unsupervised Learning (Due 10/05)
R 09/21	Dimensionality Reduction		
T 09/26	Unsupervised Learning		HW3 (Due 10/05)
R 09/28	Unsupervised Learning		
T 10/03	Classifier Fusion		
R 10/05	Performance Evaluation		
T 10/10	Fall Break		
R 10/12	Test 1		
T 10/17	Regression		Project 3 - Regression (Due 10/31) HW4 (Due 10/26)
R 10/19	Regression		
T 10/24	Neural Networks		
R 10/26	Neural Networks		
T 10/31	Support Vector Machine		Project 4 - Neural Network (Due 11/14) HW5 (Due 11/09)
R 11/02	Support Vector Machine		
T 11/07	Decision Tree		
R 11/09	Decision Tree		
T 11/14	Semi-Supervised Learning Last Day for Graduate Students to Drop with "W"		Final Project (Due 12/13)
R 11/16	Test 2		
T 11/21	Self-Supervised Learning		
R 11/23	Thanksgiving		
T 11/28			
R 11/30			
T 12/05			
W 12/13	Final Presentation		

Questions

AICIP RESEARCH

- Q1: Exactly what are the differences between
  - Machine Learning (ML)
  - Deep Learning (DL)
  - Artificial Intelligence (AI)
- Q2: What are the differences of the related courses offered at UT?
- Q3: Do I have enough background?
- Q4: What are we going to learn in this class?
- Q5: Course policy
- Q6: What's the difference between this class and some very popular online ML lectures?

Q5: Course Policy

AICIP RESEARCH

- Assignment is due 11:59pm on the due date with electronic submission through Canvas.
  - Please name your project report and homework using the following style
    uid_proj#.pdf or uid_hw#.pdf
  - Name your source code with the following style
    uid_proj#.tar
- Late policy: Each student is given a 48-hour grace period cumulatively for all assignment. The unused grace period will be counted toward bonus (0 ~ 1pt) added to the final average. The bonus is modeled as a Gaussian with an std TBD.
- Grading
  - Homework (5): 25%
  - Project (4): 32%
  - Tests (2): 28%
  - Final Project Report and Presentation: 10+5%

Questions

AICIP RESEARCH

- Q1: Exactly what are the differences between
  - Machine Learning (ML)
  - Deep Learning (DL)
  - Artificial Intelligence (AI)
- Q2: What are the differences of the related courses offered at UT?
- Q3: Do I have enough background?
- Q4: What are we going to learn in this class?
- Q5: Course policy
- Q6: What's the difference between this class and some very popular online ML lectures?

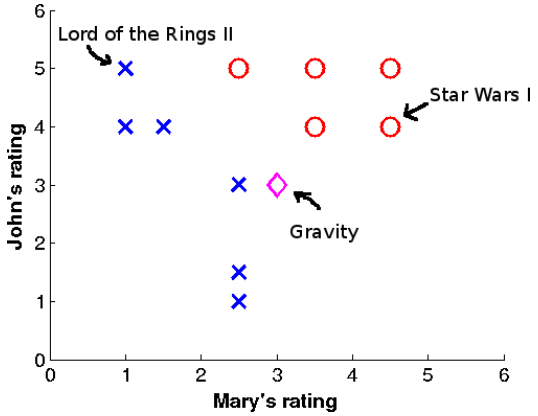
Baysian Decision Theory

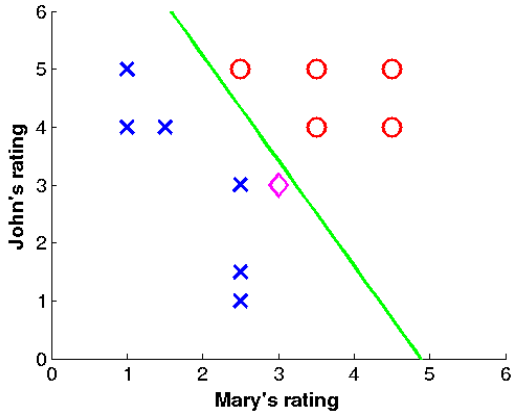
Questions

- 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?

The Toy Example

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	?





Terminologies

AICIP RESEARCH

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

Questions

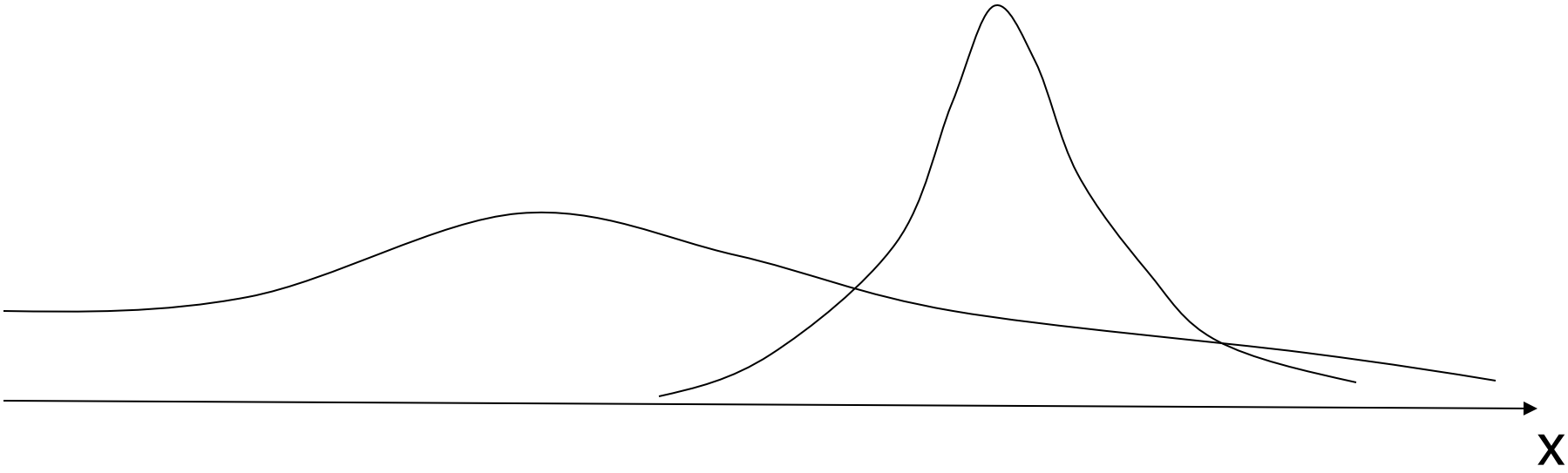
- What is supervised learning (vs. unsupervised learning)?
- What is the difference between the training set and the test set?
- What is the difference between classification and regression?
- What are features and samples?
- What is dimension?
- What is histogram?
- What is pdf?
- What is Bayes' Formula?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

Example 1 – 1-D feature

AICIP RESEARCH

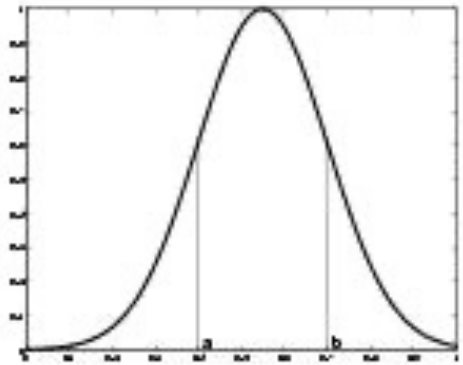
Rating	label
3.5	Y
4.8	N
3.4	Y
3.7	N
4.5	Y
4.8	N
3.6	Y
2.7	N
1	N

From Histogram to Probability Density Distribution (pdf)

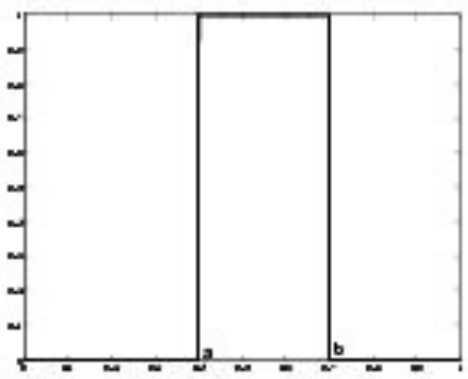


Examples of pdf

- 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

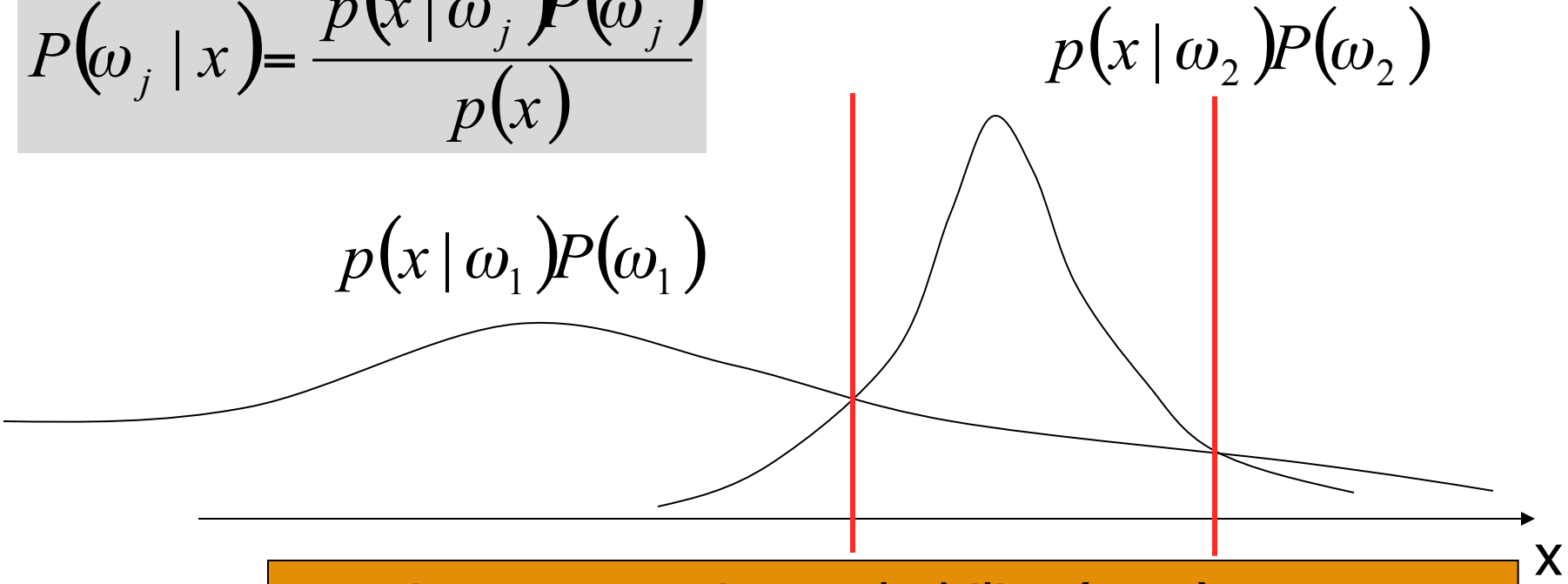
- How do you interpret prior probability in the toy example?

Questions

- 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?

Bayes Decision Rule

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



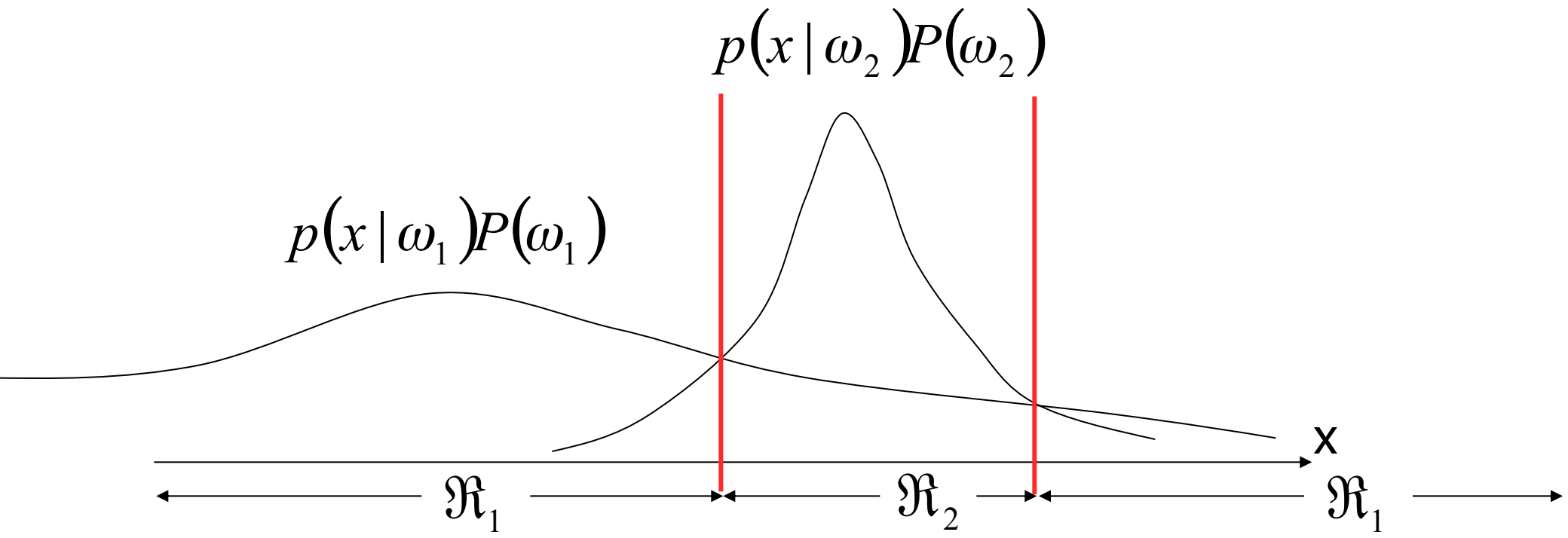
Maximum Posterior Probability (MPP):

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

then x belongs to class 1, otherwise, 2.

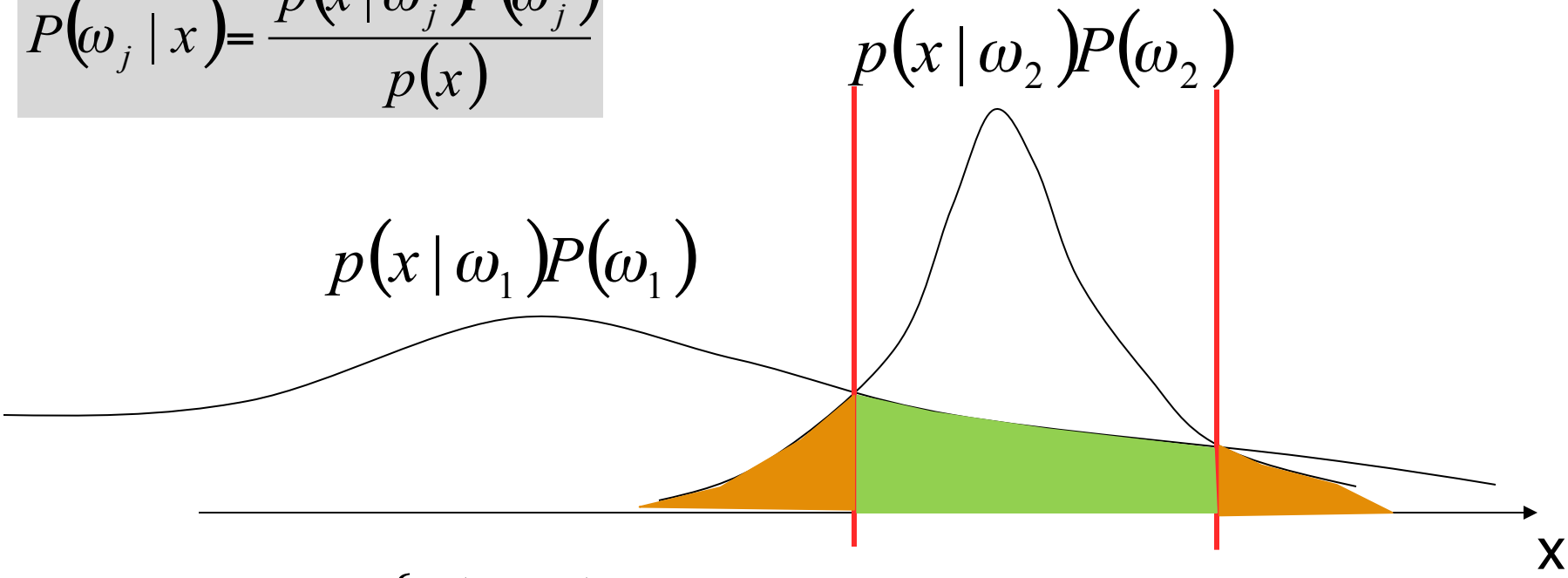
Decision Regions

- 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

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



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

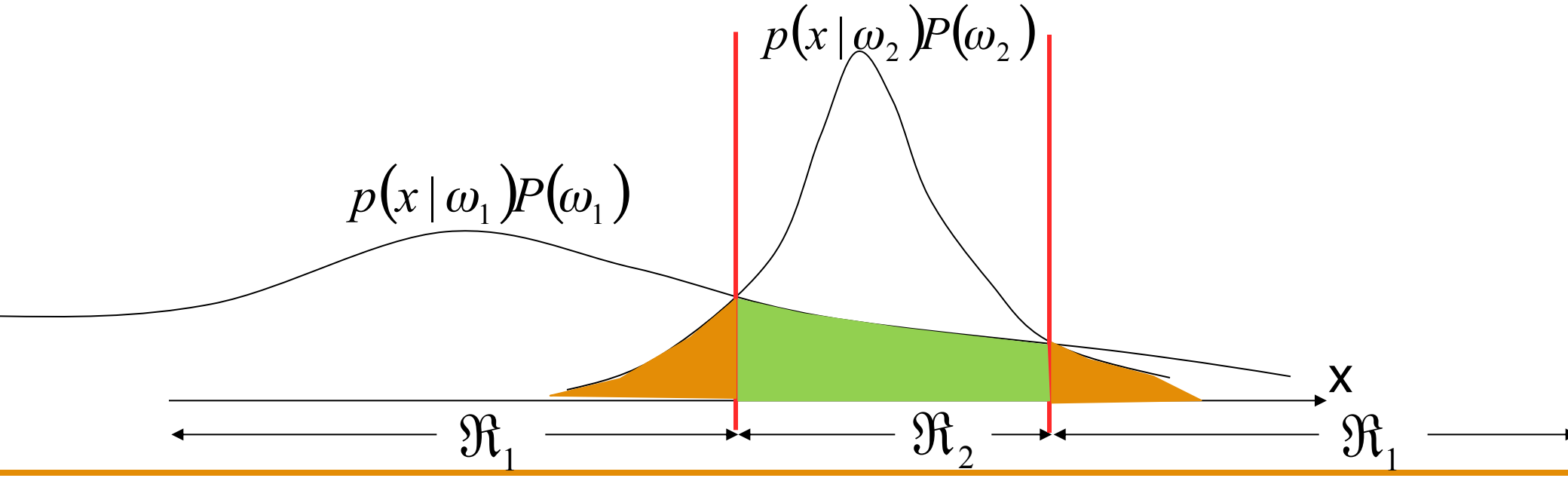
Overall Probability of Error

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 Work Altogether?

AICIP RESEARCH

Training Set:

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

Testing Sample:

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?

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

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