CSC872 Pattern Analysis and Machine Intelligence

Fall 2021
Department of Computer Science
San Francisco State University

Course Information

• **URL:** [https://bidal.sfsu.edu/~kazokada/csc872/](https://bidal.sfsu.edu/~kazokada/csc872/)

• Instructor
  – Kaz Okada, kazokada@sfsu.edu
  – OH: Zoom, Thr: 3:30 – 4:30pm

• TA:
  – Tuba Senbabaoglu, tsenbabaoglu@mail.sfsu.edu
  – OH: Zoom, Thr: 11:30-12:30pm

• See the *iLearn* course page for Zoom IDs and passcodes of the office hours.

• Policies: Please read the course website above.
• Be aware of the deadline and the late policy etc!!!
**Zoom Lecture**

- Zoom lecture links are in your *iLearn* course page
- Keep your video on during the lecture
- Voice muted but can be activated by the instructor
- Use Chat to ask questions
- Course slides shared online after each lecture through the course website: Click links in the “note” column of the “lecture plan” table.
- Zoom lectures recorded and shared in the *ilearn* course page in “Recorded Lectures”. Stay tuned

**Attendance**

- Your attendance is kept using *iLearn* attendance link (see under the zoom link)
- Passcode shared at the beginning of each lecture.
- Today’s passcode: “welcome”
- Go click the link and give that passcode.
- And report your presence there.
- We will do this for every lecture.
Waitlisted

• For those who were waitlisted, we still have to see if we can accommodate you. Please stay tuned and keep attending lectures and do assignments.

• Automatic email should be sent to you with the permit number that you can use to add to this course. Hope to resolve this soon.

Evaluations

• No Midterm/Final Exams

• Homework (50% of total grades)
  – Five HWs (See course website for schedule)
  – Shared from/Submitted to iLearn links
  – Due in one/two week(s)
  – Involves some difficult analytical problem solving

• Final Project (25% report, 10% presen)
  – Final Presentation (Presentation) on Dec 7.
  – Assignments given in the course website (follow the link)

• Fast Prototyping (15%: 5% each)
Text Books

• AIMA by Russell-Norvig: our text, general AI
  – Duda-Hurt: for PR foundation
  – Hastie: advanced ML
  – Gonzales-Woods: comprehensive IP&CV

• Read the AIMA chapters before the classes
• Additional reading assignments given as appropriate

Course Schedule

• Consult the course website for details of the lecture plan

• First part: lecture 4:00 – 5:30 ca (90min)
• 10 min break, for exercise and restroom
• Next part: in-class exercise 5:40 – 6:45 (65min)

• Drop deadline: Sep 13 (three weeks)
## Course Overview

- Intro & Agent (AI: Ch1-2)
- Search Methods (AI: Ch3-4)
- Logic and Inference (AI: Ch7-9)
- Bayesian Framework (PR: Ch13-14 etc)
- Statistical Modeling (PR: Ch20 etc)
- Statistical Classification (PR: Ch3 etc)
- Machine Learning (ML: Ch18 etc)
- Supervised Classification (ML: Ch3 etc)
- Supervised Regression (ML: Ch3 etc)
- Function Learning (NN: Ch20 etc)
- Deep Learning (NN)

## MATLAB

- You will learn how to use a powerful prototyping software environment !!!
  - Exercise Tutorials
  - Fast Prototyping
- You need to bring a laptop with MATLAB by the next lecture
  - Free copy available for all SFSU students:
    https://athelp.sfsu.edu/hc/en-us/articles/360011475074-Getting-MATLAB-for-students
  - Follow instruction to install and make sure it can start on your laptop
  - Play a little before next lecture
In-Class Exercises

- MATLAB Exercises
  - Basics of MATLAB
  - Three exercise sessions
  - Hands-on tutorials
  - TA and my help during office hours
  - End up learning a useful tool

- Fast Prototyping Exercises
  - Hands-on MATLAB software prototyping guided exercise
  - Three classic algorithms: PCA, Mean Shift, LDA
  - Three computer vision problems
  - 15% of the total grades!
  - End up learning how to quickly implement your ideas defying all the nice thing you learned in SE classes.

- Bring your own laptop with MATLAB!!!
From the intelligent eye by R.L. Gregory, 1973

Photo by R.C. James
What is PAMI?

- Pattern Analysis and Machine Intelligence is a study for:
  - A modern artificial intelligence
  - Understanding the foundation of different approaches to make machines behave intelligently
  - Applying AI techniques to various engineering tasks
  - Type of researches that get published in IEEE trans on PAMI …
What is PAMI?

Collectively, we call them PAMI studies

QUESTIONS:
Commonalities? Differences? Relationships?

Enormous!!!

What is PAMI?

Focus on
AI, PR, ML, NN
& Statistics

Emphasizing
- the three questions
- quantitative statistical approach

Upon further interests, you should study each subject further
The 3 questions: Common Framework

• What is PATTERN?
  – Codifying Properties of World
  – Data & Knowledge Representation

• What is MACHINE INTELLIGENCE?
  – Formalizing Intelligence for Machines
  – Problem Formulation

• What is ANALYSIS?
  – Analyzing Data & Knowledge to solve formulated prob.
  – Problem Solving

Common Questions For Different Approaches
Different Answers you get

Data & Knowledge Representation

• How to formally describe data/knowledge?

  • Algebraic Variables
    – Boolean, Scalar, Vector, Matrix, Tensor

  • Probabilistic Variables and Distributions
    – Random Variables, Probabilistic Mass/Density Function

  • Formal Rules
    – Rational Statement, Causality

  • Discrete & Continuous Relations
    – Tree, Graph, Function, Ontology
Problem Formulation: Problems?

- **Problems**: what is the computational task?
  - Inference
  - Modeling
  - Learning
  - Classification
  - Regression

Problem Formulation: Formulations?

- **Formalisms**: How to describe the task?
  - Agents
  - First Order Logic
  - Bayesian Inference/Classification
  - Maximum Likelihood Estimation (MLE)
  - Maximum A Posteriori Estimation (MAP)
  - Statistical Regression
  - Energy/Error Minimization
  - Maximum Information
  - Ensemble Learning
Problem Solving: Basics

• How to solve the problem w/ given data?
  – Search: Depth-First, Width-First, A*
  – Logical Inference: Resolution
  – Kernel Density Estimation (KDE)
  – Expectation-Maximization (EM) Algorithm
  – Principal Component Analysis (PCA)
  – Linear Discriminant Analysis (LDA)
  – Hill-Climbing/Gradient Descent
  – Simulated Annealing
  – Back Propagation
  – Support Vector Machine (SVM)
  – Markov Chain Monte Carlo (MCMC)
  – AdaBoost, Random Forest, XGBoost, CNN, RNN…

PAMI Framework

• KR = Data & Knowledge Representation
• PF = Problem Formulation
• PS = Problem Solving
• Make your habit to think everything in the form of (KR-PF-PS)
• Example: you as a PAMI problem…
  – KR: your brain with all the details therein
  – PF: maximize amount and quality of learning
  – PS: taking and working in this course
Review: Basic Concepts

• Some relevant mathematical ideas:
  – Calculus (high-school to lower-division)
  – Algebra (high-school to lower-division)
  – Probability (basic + some advanced)
  – Statistics (basic + some advanced)

• You want to make sure you are comfortable with these concepts and notations
• OK… some refresher now;

KR: Variable: Scalar & Vector

• Variable is:
  – Symbolic representation of quantity
  – Unknown quantity that can change in algebraic sense
  – Measurable attribute of a system in statistics

• Scalar \( \mathbf{X} \): Variable indicating a single-valued entity
• Vector \( \mathbf{X} \): Variable indicating a multiple-valued entity

\[
\begin{align*}
\mathbf{x} &= (x \ y)^T \\
x &= a : \text{area} \\
x &= \mu : \text{angle}
\end{align*}
\]

Dimension := number of coeffs
KR: Continuous vs. Discrete

- **Continuous Variable X**
  - indicates real-value entities
  - $x \in R$
  - $x = (x_1, ..., x_n)^T \in R^n$  
    \hspace{1cm} $N$-Dim vector

- **Discrete Variable $X_i$**
  - only take a set of predetermined discrete values
  - $x_i \in N$  \hspace{.5cm} Natural number: $i = 1, 2, ...$
  - $x_i \in \{MO, TU, WE, TH, FR, SU, SA\}$

KR: Function

- Deterministic dependence of two quantities/sets, associating input $X$ to output $Y$ by a binary relation

- $f: X \rightarrow Y$
- Map, Mapping, Transformation = Function
- Inverse function: $g = f^{-1}: Y \rightarrow X$
KR: Function Properties

- Rules of $f$ described in a graphical plot or sometimes in analytic formula when known

\[
\begin{array}{c}
\text{y} \\
\bullet \\
\text{x} \\
\end{array}
\]

\[
y = f(x) \\
x \in X \\
y \in Y
\]

- Continuous Function
- Differentiability
- Smooth Function
  - All-order differentiable over entire domain

KR: Matrix

- Product does not commute: $AB \neq BA$
- Transpose: $A^T: a_{ij} \rightarrow a_{ji}$ (swapping rows & columns): $(AB)^T = B^T A^T$
- Symmetric matrix $A$: $A^T = A: a_{ij} = a_{ji}$
- Inverse matrix of $A^{-1}: A A^{-1} = A^{-1} A = I_n$
- Orthogonal matrix $A$: $A^T = A^{-1}: A A^T = A^T A = I_n$
- Outer and inner product

\[
\begin{align*}
xx^T & = \begin{pmatrix} x^2 & xy \\ xy & y^2 \end{pmatrix} \\
x^T x & = x^2 + y^2 = tr(xx^T)
\end{align*}
\]
PF: Matrix Equations

- **Linear equations**
  \[ ax_1 + bx_2 = e \]
  \[ cx_1 + dx_2 = f \]
  \[ Ax = b; A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}; b = \begin{pmatrix} e \\ f \end{pmatrix} \]
  \[ x = A^{-1}b \]

- **Linear transform**
  \[ y = Ax \]
  \[ x = A^{-1}y \]

- **Eigen-values & vectors**
  \[ \lambda v = Av; \lambda \in \mathbb{R} \]

KR: Graph

- **Undirected graph**: \( G = (V, E) \)
  - vertices and edges (no direction)
- **Directed graph**: \( G = (V, A) \)
  - vertices and arrows
- **Directed acyclic graph (DAG)**
  - Directed graph without a loop
- **Connected graph**
  - Can reach from any vertex from any other vertices
- **Connected DAG**
  - Tree
KR: Boolean Variable

- A is a Boolean variable if it indicates two-valued system, a statement or event
  - e.g., indicator variable $A = \{\text{Yes, No}\}$
  - e.g., $A$ = My name is George
  - e.g., $A$ = I teach CSC872

- Some event has intrinsic degree of **uncertainty** as to whether $A$ occurs
  - e.g., $A$ = There will be an earthquake tomorrow
  - e.g., $A$ = My stock price will go up tomorrow

- **Random Variable** is a function that chooses a value from the event space \{True, False\} according to probability $P(A)$

---

KR: Basic Probability

- $P(A)$ means “the fraction of possible worlds in which $A$ is true”

  ![Diagram of Event Space and A being True](image)

  - Event space of all possible worlds
  - Its area is 1
  - $P(A)$ is the area of the pink circle

- The axioms of probability !!!
  - $0 \leq P(A) \leq 1$
  - $P(\text{True}) = 1$
  - $P(\text{False}) = 0$
  - $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

• Joint Probability
  – Probability of two events in conjunction
  – \( P(A \text{ and } B) := P(A \land B) := P(A, B) \)

• Marginal Probability
  – Probability of one event \( A \) regardless of the other events \( B \)
  – Obtained by summing (integrating) a joint probability over the event space \( \Omega \) for unwanted events \( B \)
  
  \[
  P(A) = \sum_{v \in \Omega} P(A \land B = v) = P(A \land B) + P(A \land \neg B)
  \]
  
  – \( P(\neg A) := P(\neg A) \)
  – \( P(A) + P(\neg A) = 1 \) total probability theorem


• Conditional Probability
  – Probability of an event \( A \) given other event \( B \)
  – \( P(A|B) = \frac{\text{area of } A \text{ and } B}{\text{area of } B} \)
  – \( P(A|B) = \frac{P(A \land B)}{P(B)} \)
  – \( P(\text{Yes/Math}) \) ?

• Product Rule
  – Joint probability can be written as a product of a conditional and a marginal
  
  \[
  P(A \land B) = P(A|B)P(B) = P(B|A)P(A)
  \]

• Statistical Independence satisfies
  – \( P(A \land B) = P(A)P(B) \)
  – \( P(A|B) = P(A) \)
  – \( P(B|A) = P(B) \)

\[
\begin{array}{c|c|c}
X & Y & \\
\hline
\text{Math} & \text{Yes} & \\
\text{History} & \text{No} & \\
\text{CS} & \text{Yes} & \\
\text{Math} & \text{No} & \\
\text{Math} & \text{No} & \\
\text{CS} & \text{Yes} & \\
\text{History} & \text{No} & \\
\text{Math} & \text{Yes} & \\
\end{array}
\]

\( X = \text{College Major} \)
\( Y = \text{Likes “XBOX”} \)
KR: Beyond Boolean Events

• When more than one state (over a discrete variable):
  – e.g., \( X = \text{day}, \Omega := \{\text{Mon}, \ldots, \text{Sun}\} \)
  – Discrete Random Variable
    – \( P(X = v_i \text{ and } X = v_j) = 0 \text{ if } i \neq j \) (mutually exclusive)
    – \( P(X = v_1, \ldots, X = v_k) = \sum_{i=1}^{k} P(X = v_i) = 1 \) (total prob. Th.)
    – \( P(Y) = \sum_{i=1}^{k} P(Y \text{ and } X = v_i) \) (marginal)

• When over continuous variable:
  – Continuous Random Variable
  – e.g., \( X = \text{temperature of SF} \)

KR: Probability Distribution

• For a discrete random variable \( X \)
  – Probability Mass Function
    \[ P(X = x_i) \]
    \[ \sum_{i=1}^{k} P(X = x_i) = 1 \]

• For a real-valued random variable \( X \)
  – Probability Density Function
    \[ p(x) \]
    \[ \int_{a}^{b} p(x) \, dx = 1 \]
    \[ P(a < X \leq b) = \int_{a}^{b} p(x) \, dx \]
KR: Expectation

- For a discrete random variable $X$
  - $E[X] = \sum_\omega x_i P(X=x_i) = \mu$ (population mean)
  - $E[f(X)] = \sum_\omega f(X=x_i)P(X=x_i)$

- For a real-valued random variable $X$
  - $E[X] = \int_\Omega xP(y)dy$

- Linearity
  - $E[aX+Y] = aE[X]+E[Y] \approx a\mu_1 + \mu_2$

PF: Statistics

- Independent and Identically-Distributed (i.i.d.) Random Variable
  - Rolling a fair dice for instance.
  - If $x_1, x_2, x_3, \ldots, x_i, \ldots, x_k$ are i.i.d. of $X$ then
    - $P(x_1, x_2, x_3, \ldots, x_i, \ldots, x_k) = P(X=x_1)P(X=x_2)\ldots P(X=x_k)$

- Central limit theorem
  - The sum of i.i.d. random variables with finite variance will be approximately normally (Gaussian) distributed as we go towards an infinite number of samples.
  - A reason why you see a lot of Gaussians …