Wellcome back

CSC872 Pattern Analysis and Machine Intelligence

Spring 2025

Department of Computer Science

San Francisco State University

Course Information

- **URL:** https://bidal.sfsu.edu/~kazokada/csc872/
- Instructor
  - **Kaz Okada**, kazokada@sfsu.edu
  - OH: TH911, Thr: 11:30am – 12:30pm

- TA: (CA Building
  - **Shivani Bokka**, sbokka@sfsu.edu
  - OH: ~~Zoom~~, Mon: 2:00pm – 3:00pm

- See the *Canvas* page for other information.
- Policies: Please read the course website above.
- Be aware of the deadline and the late policy etc!!!

Basics

Hello! :)

- Course slides shared online after each lecture through the course website: Click links in the "note" column of the "lecture plan" table.
- Attendance are taken in-person.
- Honor Code:
  - I request all students to read and pledge on the course honor code (found in Canvas: general information).
  - No assistive sites (chegg, numerade) No ChatGPT etc
  - Read it thoroughly and send your pledge on Canvas **by Feb 4th**. See instructions given in the Canvas link.

Waitlisted

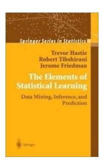
- For those who were waitlisted, we still have to see if we can accommodate you. Please stay tuned and keep attending lectures and do assignments.
- If I choose to add you with permit number after the 2nd week, automatic email should be sent to you with the permit number that you can use to add to this course. Hope to resolve this soon.

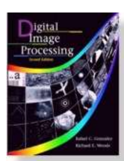
Evaluations

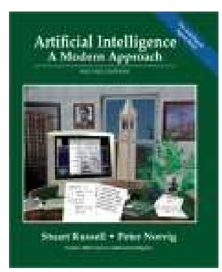
- No Midterm/Final Exams
- Homework (50% of total grades)
  - **Five** HWs (See course website for schedule)
  - **Shared from/Submitted to *Canvas* links**
  - Due in one/two week(s)
  - Includes some difficult analytical problem solving
- Final Project (25% report, 10% presen)
  - Final Report (**Literature Survey Report**) due on May 13.
  - Final Presentation (Presentation) on May 13.
  - Assignments given in the course website (follow the link)
- Fast Prototyping (15%: 5% each for 3 exers)

Text Books

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







- **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: Feb 17 (3 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 (Go to Canvas:Matlab)
  - Follow instruction to install and make sure it can start on your laptop
  - Play a little before next lecture

In-Class Exercises

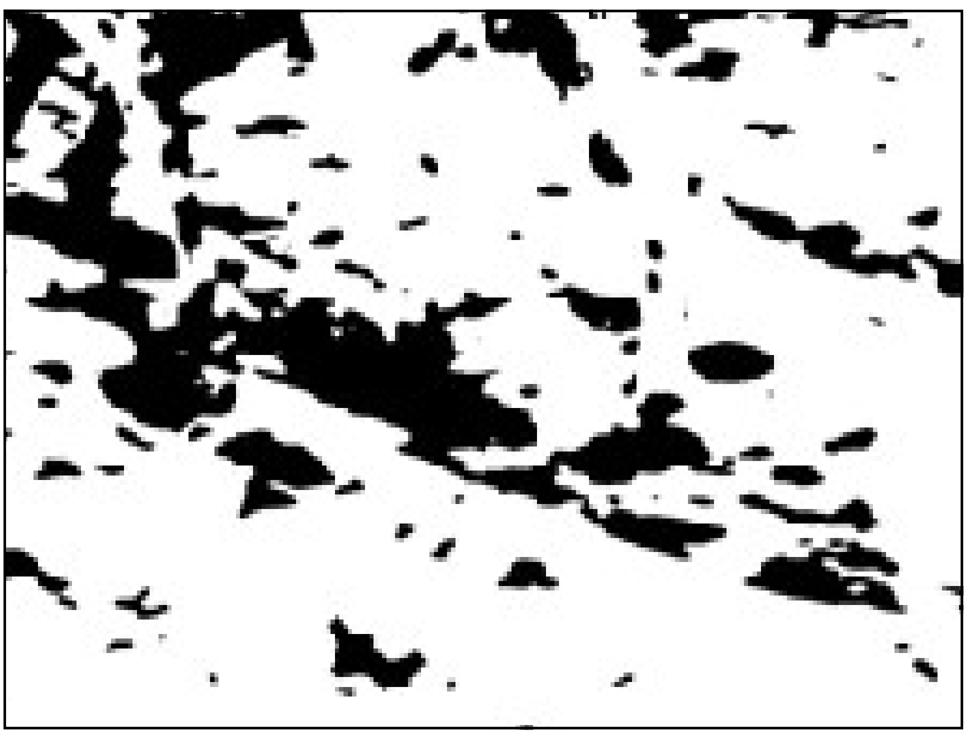
- MATLAB Exercise
  - Basics of MATLAB
  - Three exercise-sessions
  - Hands-on tutorials
  - TA and my help during office hours
  - End up learning a useful tool
- Fast Prototyping Exercises
  - Hands-on MATLAB software prototyping guided exercise
  - Three Exercises
    - 3 classic algorithms: PCA, Mean Shift, LDA
    - 3 computer vision problems
  - 3 sessions for each exercise
  - 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!!!**

Introduction: the PAMI framework

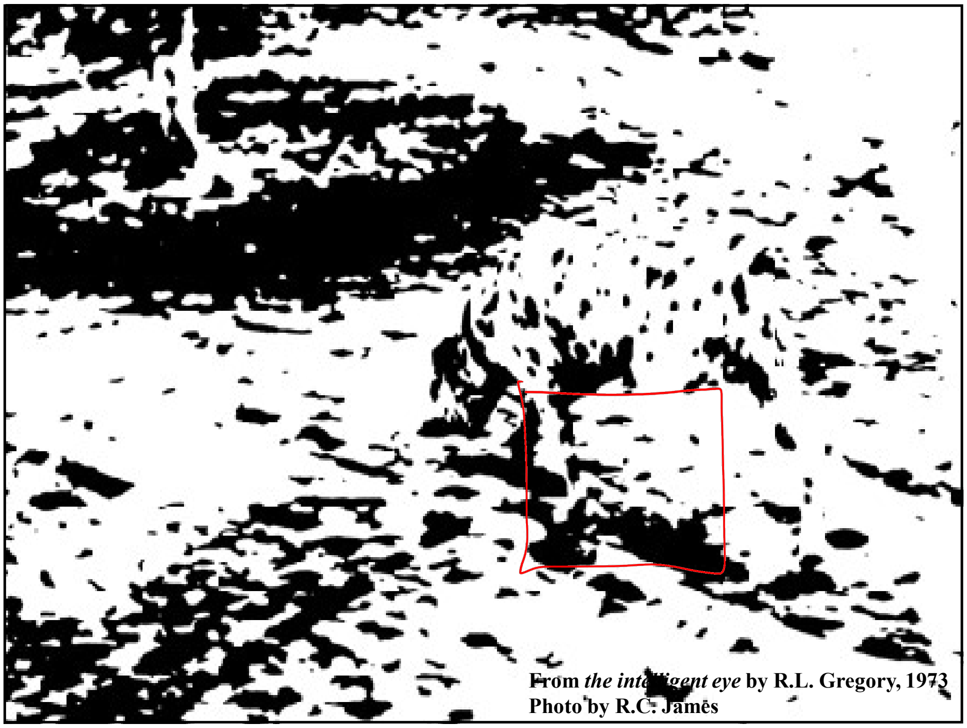
CSC872

Pattern Analysis and Machine Intelligence

11

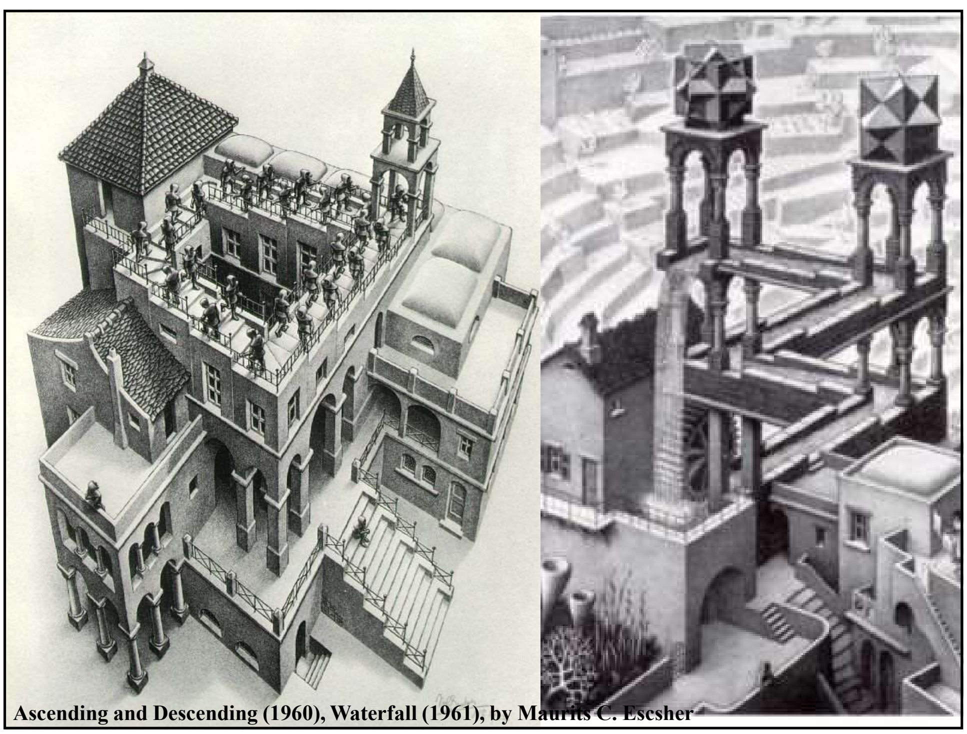


12



From *the intelligent eye* by R.L. Gregory, 1973
Photo by R.C. James

13

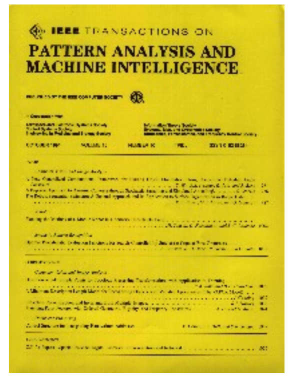


Ascending and Descending (1960), Waterfall (1961), by Maurits C. Escsher

14

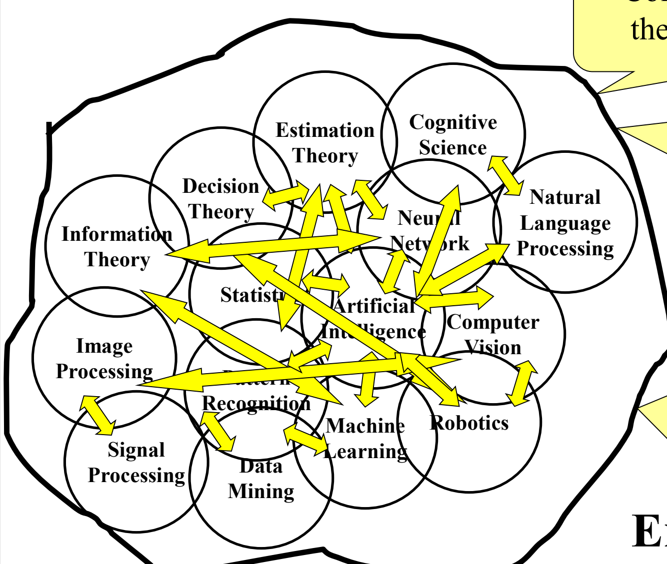
What is PAMI?

- Pattern Analysis and Machine Intelligence is a study for:
  - A classic 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

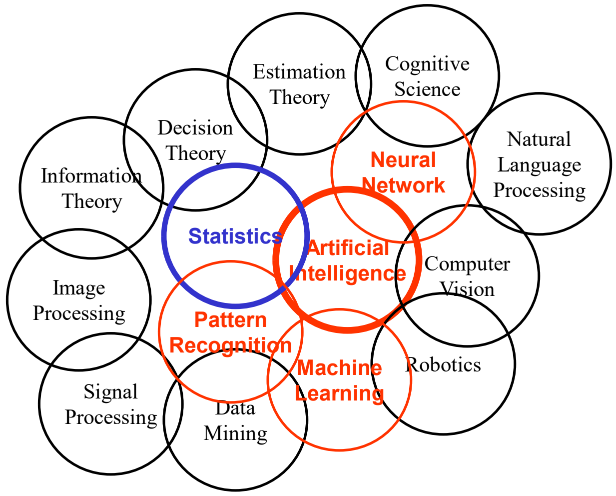


Well, which one should I use for my program…

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

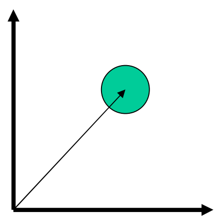
Review: Basic Concepts

- 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 X** : Variable indicating a single-valued entity
- **Vector X** : Variable indicating a multiple-valued entity

$\begin{pmatrix} x \\ y \end{pmatrix}^T = (x, y)$



x = $(x \;\; y)^T$ = $\begin{pmatrix} x \\ y \end{pmatrix}$
x = a : area
x = μ : angle
Dimension := number of coeffs

KR: Continuous vs. Discrete

- **Continuous Variable X**
  - indicates real-value entities
  - $x \in R$
  - $x = (x_1, \dots, x_n)^T \in R^n$

$\overbrace{R \times R \times R \times \cdots \times R}^{n}$

N-Dim vector

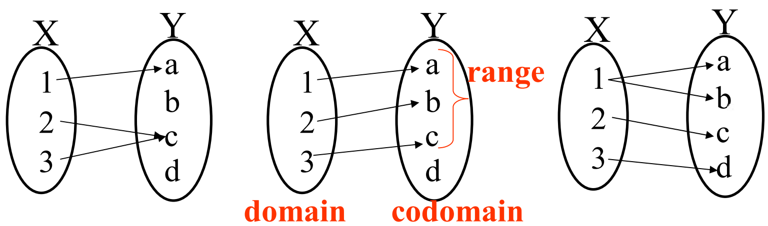
- **Discrete Variable X_i**
  - only take a set of predetermined discrete values

    $x_i \in N$ Natural number: $i = 1, 2, \dots$.

    $x_i \in \{MO, TU, WE, TH, FR, SU, SA\}$

KR: Function

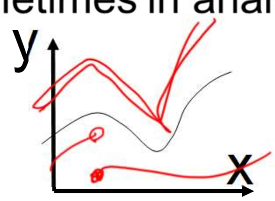
- Deterministic dependence of two quantities/sets, associating input X to output Y by a binary relation



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

KR: Function Properties

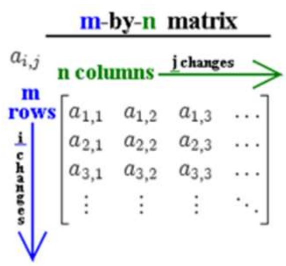
- Rules of f described in a graphical plot or sometimes in analytic formula when known



$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$$
$$\boldsymbol{x} \in \boldsymbol{X}$$
$$\boldsymbol{y} \in \boldsymbol{Y}$$

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

KR: Matrix

A = 

Vector as a Matrix

$x = \begin{pmatrix} x \\ y \end{pmatrix} = (x \ \ y)^T$

$x^T = (x \ \ y)$

- Product does not commute: $\mathbf{AB} \neq \mathbf{BA}$
- Transpose: $\mathbf{A}^T : \mathbf{a}_{ij} \leftarrow \mathbf{a}_{ji}$, (swapping rows & columns): $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$
- Symmetric matrix $\mathbf{A}$: $\mathbf{A}^T = \mathbf{A} : \mathbf{a}_{ij} = \mathbf{a}_{ji}$
- Inverse matrix of $\mathbf{A}^{-1}$: $\mathbf{A}\,\mathbf{A}^{-1} = \mathbf{A}^{-1}\,\mathbf{A} = \mathbf{I}_n$
- Orthogonal matrix $\mathbf{A}$: $\mathbf{A}^T = \mathbf{A}^{-1}$: $\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I}_n$
- Outer and inner product

$xx^T \qquad xx^T = \begin{pmatrix} x^2 & xy \\ xy & y^2 \end{pmatrix}$

$x^Tx \qquad x^Tx = x^2 + y^2 = tr(xx^T)$

PF: Matrix Equations

- Linear equations

$$ax_1 + bx_2 = e$$
$$cx_1 + dx_2 = f$$

⇒ $Ax = b; A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}; b = \begin{pmatrix} e \\ f \end{pmatrix}$

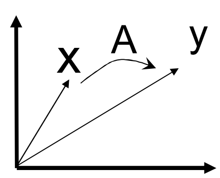
$x = A^{-1}b$

- Linear transform

$y = Ax$

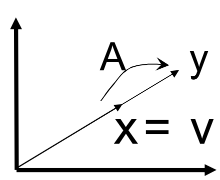
$x = A^{-1}y$

$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ ⇒



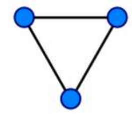
- Eigen-values & vectors

$\lambda v = Av; \lambda \in R$ ⇒

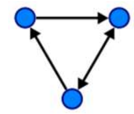


KR: Graph

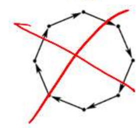
- Undirected graph: G=(V,E)
  - vertices and edges (no direction)



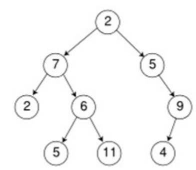
- Directed graph: G=(V,A)
  - vertices and arrows



- 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

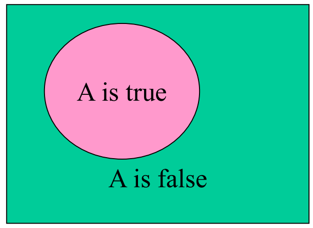
- **A** is a Boolean variable if it indicates two-valued system, a statement or event
  - e.g., indicator variable **A** = **{Yes, No}**
  - e.g., **A** = My name is George
  - e.g., **A** = I teach CSC872
- Some event has intrinsic degree of **uncertainty** as to whether **A** occurs
  - e.g., **A** = There will be an earthquake tomorrow P(A) = 0.0001
  - 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"

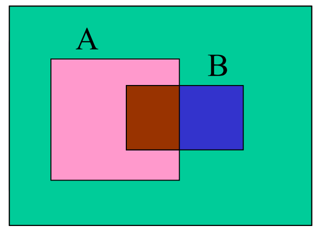
Event space of all possible worlds

Its area is 1



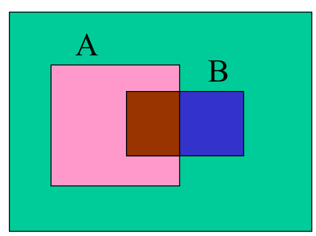
P(A) is the area of the pink circle

- The axioms of probability !!!
  - **0 ≤ P(A) ≤ 1**
  - **P(True) = 1**
  - **P(False) = 0**
  - **P(A or B) = P(A) + P(B) – P(A and B)**



KR: Probability Cont.

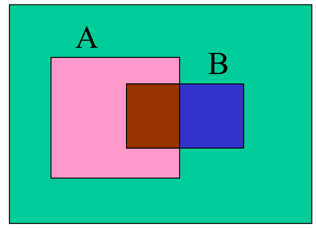
- Joint Probability
  - Probability of two events in conjunction
  - **P(A and B) := P(A ^ B) := P(A, B)**



- Marginal Probability
  - Probability of one event (A) regardless of the other events (B)
  - Obtained by summing (integrating) a joint probability over the event space Ω for unwanted events (B)
  - **P(A) =** $\sum_{v \in \Omega}$ **P(A ^ B=v)**
    = P(A ^ B) + P(A ^ ~B)
  - **P(not A) = P(~A)**
  - **P(A) + P(~A) = 1** total probability theorem

KR: Probability Cont.

- Conditional Probability
  - Probability of an event (A) given other event (B)
  - **P(A|B) = {area of A and B} / {area of B}**
  - **P(A|B) = P(A ^ B) / P(B)**
  - **P(Yes|Math)** ? = P(Yes,M) / P(Math) = 2/8 / 4/8 = 2/8 · 8/4 = 1/2



X	Y
Math	Yes
History	No
CS	Yes
Math	No
Math	No
CS	Yes
History	No
Math	Yes

X = College Major

Y = Likes "XBOX"

- Product Rule
  - Joint probability can be written as a product of a conditional and a marginal
  - **P(A ^ B) = P(A|B)P(B)**
    = P(B|A)P(A)
- Statistical Independence satisfies
  - **P(A ^ B) = P(A)P(B)**
  - P(A|B) = P(A)
  - P(B|A) = P(B)

35

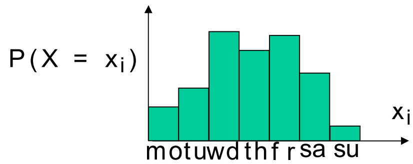
KR: Beyond Boolean Events

- When more than one state (over a discrete variable):
  - e.g., **X = day, Ω := {Mon,..,Sun}**
  - **Discrete Random Variable**
  - $P(X = v_i \text{ and } X = v_j) = 0 \text{ if } i \,!= j$ **(mutually exclusive)**
  - $P(X = v_1, .., X = v_k) = \sum_{i=1}^{k} P(X = v_i) = 1$ **(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 = temperature of SF**

36

KR: Probability Distribution

- For a discrete random variable X
  - Probability Mass Function



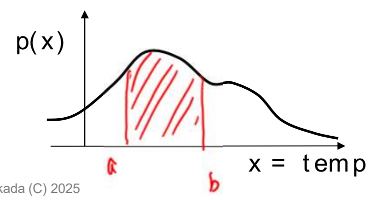
$\sum_{i=1}^{k} P(X = x_i) = \mathbf{1}$

total probability theorem

- For a real-valued random variable X
  - Probability Density Function

$\int_{-\infty}^{+\infty} p(x)\,dx = 1$

$P(a \le X \le b) = \int_a^b p(x)dx$



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
  - $E[X] = \int_\Omega xP(y)dy$

$E[X] = \mu_x \quad E[Y] = \mu_y$

- Linearity
  - E[aX+Y] = aE[X]+E[Y] $= a\mu_x + \mu_y$

PF: Statistics

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

- 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 …