CSDS 440: Machine Learning

Soumya Ray (he/him, sray@case.edu)

Olin 516

Office hours T, Th 11:15-11:45 or by appointment

Today

- Intro to CSDS 440

Class Meeting Protocol

- In person
  - Lecture slides and other materials will be on canvas
  - Quizzes will be in person
  - Masks may be worn or not, at your discretion
- If you have any COVID symptoms or are unable to make it in person:
  - Do not come to class
  - Let me know, and I will open a zoom session that you can join remotely
- If I develop any COVID symptoms, we will switch to zoom meetings until symptoms go away

What is this class about?

- ML is a subfield of AI
- This course is a graduate-level “deep dive” into ML theory, algorithms and experimental methods
- **Outcome**: understand key ideas and tradeoffs involved so you can
  - design your own (new) ML algorithms,
  - measure their performance correctly,
  - analyze their behavior.

Syllabus

- Part 1: Foundations, Algorithms, Evaluation
- Part 2: Theory and Extensions

What you should know

- Computer science concepts (algorithms, runtime/space, data structures etc)
- Good programming skills
- Strong foundations in probability and statistics, calculus and linear algebra

- Helpful
  - General exposure to Artificial Intelligence
  - Optimization

Alternative new class

- CSDS 340
  - Undergrad level
  - Especially suitable for learning how to apply ML methods
  - More “hands on” than CSDS440
  - Same schedule as CSDS 440 (this semester)

Mechanics: Canvas

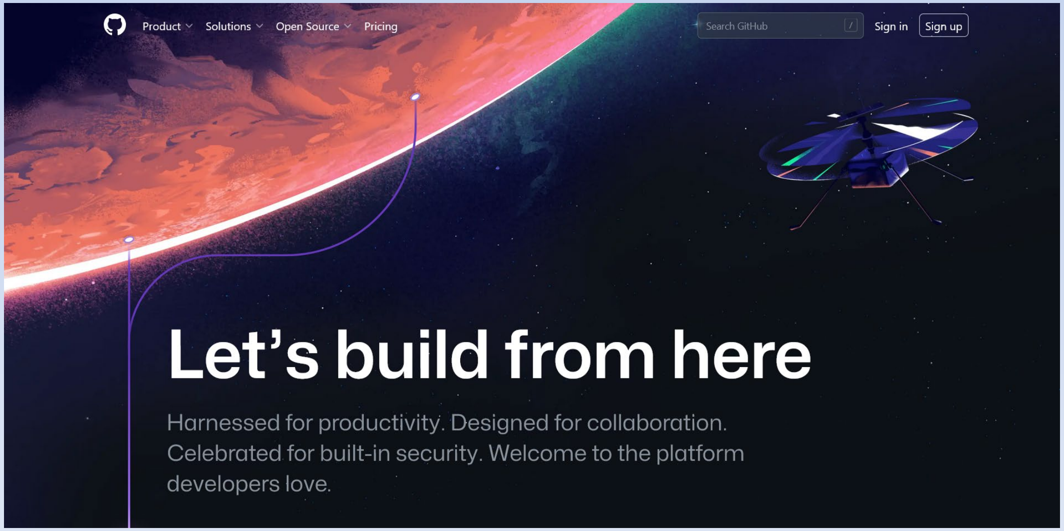
- “Files:” Lecture Notes and assignment descriptions

Mechanics: Textbooks

- No required text
- Some recommended:
  - Machine Learning by Tom Mitchell
    - Chapter excerpts on canvas
  - Pattern Recognition and Machine Learning by Christopher Bishop
  - Deep Learning by Goodfellow, Bengio, Courville
- Other material will be linked from course website as needed
- Papers will be cited as needed, you can ask me for copies

Mechanics: Github

- All assignments will be submitted through github (both programming and written)
- Please accept invite to “cwru-courses”
- I will go over how to setup git



Office hours and TAs

- T, Th 11:15-11:45
  - Or by appointment

- TAs: Ibrahim Berber and ??
  - Office hours TBA

Mechanics: Email/Slack

- You can send me email at sray@case.edu
- It will help if you prefix the subject line with "CSDS 440"
  - E.g. "CSDS 440: Question about homework 1"
- Slack group
  - Optional, but easier than email+ allows group discussion

Written Assignments

- ~3 written assignments
  - Due Friday 11:59pm
  - Solutions to be written in Github Markdown and submitted through Github
    - https://docs.github.com/en/get-started/writing-on-github
  - All questions can be answered from lecture slides, or additional reading as indicated
  - Ask for help if stuck!

Programming Exercises

- 2 programming exercises, due Friday 11:59pm (may be Saturday 11:59pm if colliding with written)
  - Programming in Python
  - A framework will be provided to parse inputs etc
  - No external libraries except data structures and math (e.g. numpy, scipy, cvxopt and equivalent)
- To be turned in via github

Class Project

- Survey algorithms for an area of ML (I will specify)
  - Read at least two papers/person
  - Implement and evaluate at least one algorithm per person
- Implement "research extension"
  - More details after midterm
- Present your research in class
- Write technical report on your work (in Markdown)
- Code and report submitted through github

Tests and Quizzes

- 3 tests, math/theory questions
  - In class, ~45 minutes
  - Cumulative

- 5 short quizzes, short answer/fill in the blank
  - ~10-15 minutes
  - Will NOT be announced beforehand
  - Best 3 will be used for grade

Tentative Grading Scheme

- Written Problems: 15%
- Programming Exercises: 15%
- Project Report+presentation+code: 35%
- Tests: 25%
- Quizzes: 10%
- Grades are assigned on an "optimistic curve": 80-90 OR class average to average+1sd is B if class average is below 80

Bonuses to your grade

- Any written or programming assignment turned in a week or more in advance of due date: +10% to your score
- Partial credit is available for all assignments
- +2% possible at my discretion for quality participation in class
- +1% for filling in course evaluations at the end

Penalties to your grade

- One late day each week (up to Saturday 11:59 pm)
  - If Canvas marks your assignments as late, you will receive a 20% penalty to the score
- Submissions after Saturday 11:59pm for a week will NOT be graded
  - (If you have a genuine reason, such as illness, please email/slack or talk to me and ask for an extension)

Quality of Work

- Graduate level work is expected
  - Neatly formatted answers (in markdown)
  - Clear arguments, no steps omitted proofs
  - Well structured, clean and efficient code that is properly documented for your project

Collaboration policy

- All assignments can be done in groups of up to 3
  - Each group will share one repository to submit their assignments
  - Self signup enabled in Canvas
  - It is expected that everyone will pull their weight
    - Must contribute equally to all work done
      - Verified using git logs (contributions must be identifiable!)
    - If not, person doing less work will be penalized
    - For written assignments, names of everyone who contributed must appear on the assignment

Academic Integrity Policy

- You are free to talk to fellow students, TAs/mentors and me about assignments to clarify/refine your ideas
- But any submitted work MUST be substantially your own
  - Do not copy from any source <u>including any online sources</u>
    - Some questions may allow you to consult online sources, then it is ok
  - Do not put your name on an assignment where your partners did the work
  - Do not use content created by AI and submit it as your own
  - Any violation will be penalized up to failure in the class

Group Formation

- Ideally, everyone in a group has similar background/academic stage
- Must have common free time weekly
- Must commit to regular, responsive communication
- Groups are not immutable
  - But do let me, the TAs and your current group know if you are changing groups

Course Load

- I recommend setting aside 6-8 hours each week to work on this course (besides class hours)

- Do timely work

Questions?

Set up your git(hub) environment

- Accept invite from cwru-courses
- Once you have formed your group:
  - Choose one person from the group to do the following
  - Go to cwru-courses and make a **private** repository csds440-f24-n, where n is your group number
    - **Please do not modify this repository name in any way**
  - Add your group members (Settings)
  - Add me (sray-cwru) and the TAs as Admins

Set up your git(hub) environment

- Download and install Github Desktop
  - https://desktop.github.com/
- Start desktop and log in to github
  - https://docs.github.com/en/desktop/installing-and-authenticating-to-github-desktop/authenticating-to-github-in-github-desktop

- Choose to “Clone a repository”, then choose the repository you just created
- Remember the directory it is cloned into
  - All your homework, code etc should be placed in this directory

Todo list

- Form groups by Friday
  - If not I will assign
- Refresh python
- Refresh git
  - Accept cwru-courses invite
- Review probability slides
- Communicate with partners, set up schedule and allocate block of time for this class

Part 2

- Review of Probability

Probability Theory

- A language that associates facts with "degrees of belief," with associated mechanics for reasoning

I think it is 60% likely that it will rain tomorrow.

$RainTomorrow = true$	60%
(fact)	(degree of belief)

Random Variable (R.V.)

- A variable that refers to an uncertain fact
  – Has a domain that can be discrete or continuous

- For each value (or set of values), we can specify a *degree of belief* that shows how much we believe the stated fact---this is the *probability* associated with the fact
  – Denoted $\Pr(.)$

Example

- $RainTomorrow \in$ {True, False}
  - $\Pr(RainTomorrow = True) = 0.6$

- $Current\_X\_Position \in (-\infty, +\infty)$
  - $\Pr(-1 \leq Current\_X\_Position) = 0.2$

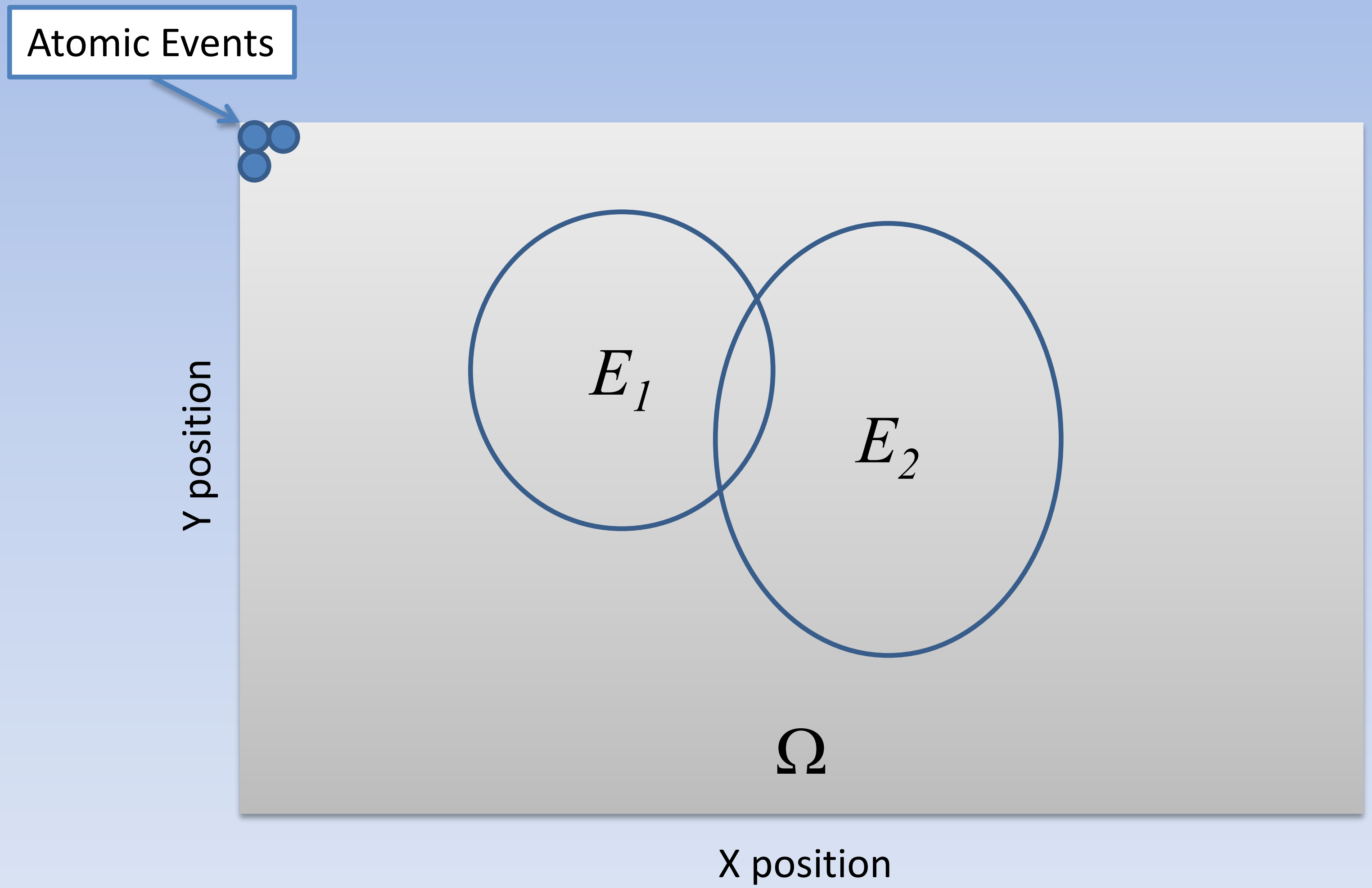
Atomic Event

- If the state of the world is described by n r.v.'s and we assign values to all of them, this defines an atomic event.

Events and the Sample Space

- Atomic events are mutually exclusive and exhaustive
  - At most one can be the true state of affairs
  - The true state of affairs must be one of them
- An "event" is a collection of atomic events
  - Example: the event $\{x=2\}$ is the collection of atomic events $\{(x=2, y=1), (x=2, y=2), (x=2, y=3), \ldots\}$
- The "sample space" is the collection of all possible atomic events (Ω)

Picture



Joint Probability

- Just like we assign degrees of belief to single r.v.'s, we can do the same for groups of r.v.'s
  - $\Pr(RainTomorrow{=}Yes,\ CloudyTomorrow{=}Yes) = 0.99$
  - $\Pr(-1 \leq x,\ y \leq 1) = 0.2$
  - In particular, we can assign degrees of belief to atomic events

Axioms of Probability

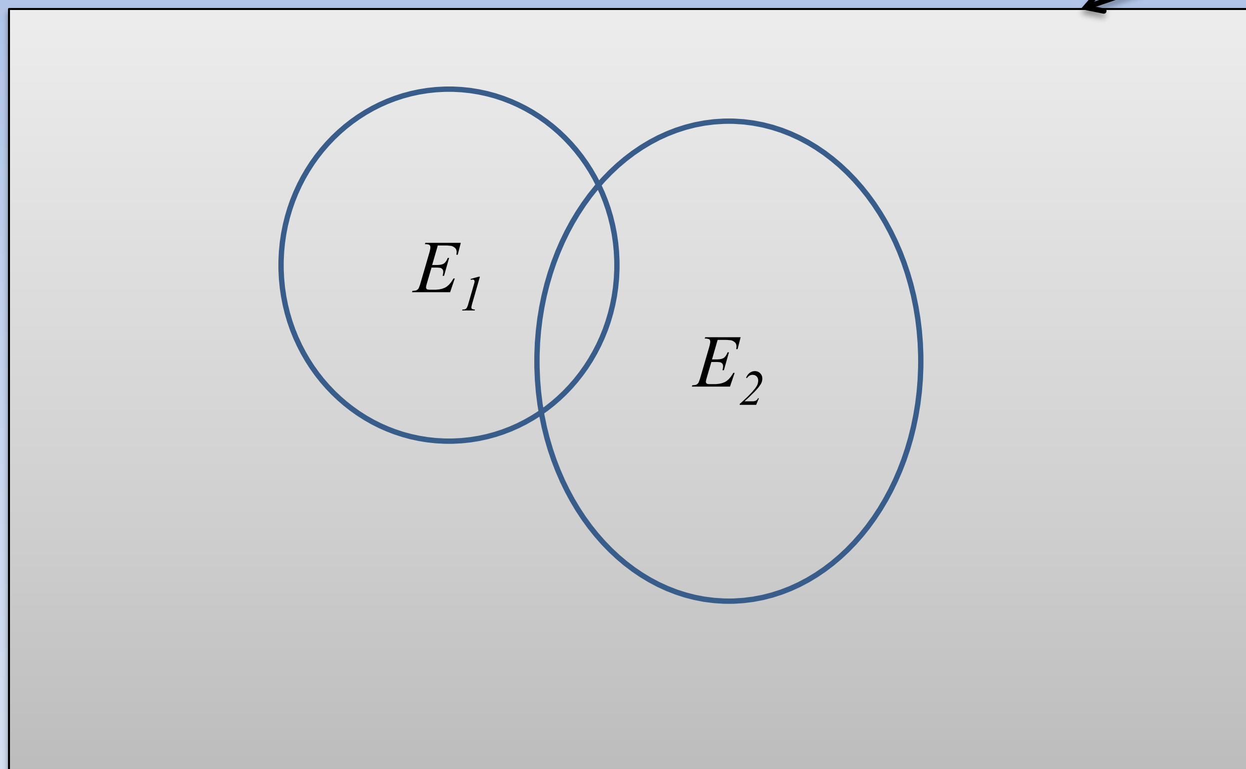
- For any event E, $0 \le \Pr(E) \le 1$
- $\Pr(\Omega)=1$
- For *mutually disjoint* events, the probability of the union is given by:

$$\Pr(\bigcup_{i=1} E_i) = \sum_{i=1} \Pr(E_i)$$

In particular this must apply to atomic events.

Picture

Sample Space: Total “area”=1



$Pr(E)$=area under E

Why?

- Could there be other ways of representing uncertainty?
- But probability theory has a major positive result: suppose someone's degrees of belief for some set of events does NOT obey the axioms of probability. Then there is a way to bet against them such that they will always lose money (utility) over time. (Bruno de Finetti 1931)

Probability Density Functions

- Earlier we defined probabilities associated with r.v.'s: $\Pr(RainTomorrow=Yes)=0.8$
- A function that maps *every* value of an r.v. to a probability is called a probability density function (p.d.f.)

$$p_{RainTomorrow}(x)=\begin{cases}0.8 \text{ if } x = Yes\\ 0.2 \text{ if } x = No\end{cases}$$

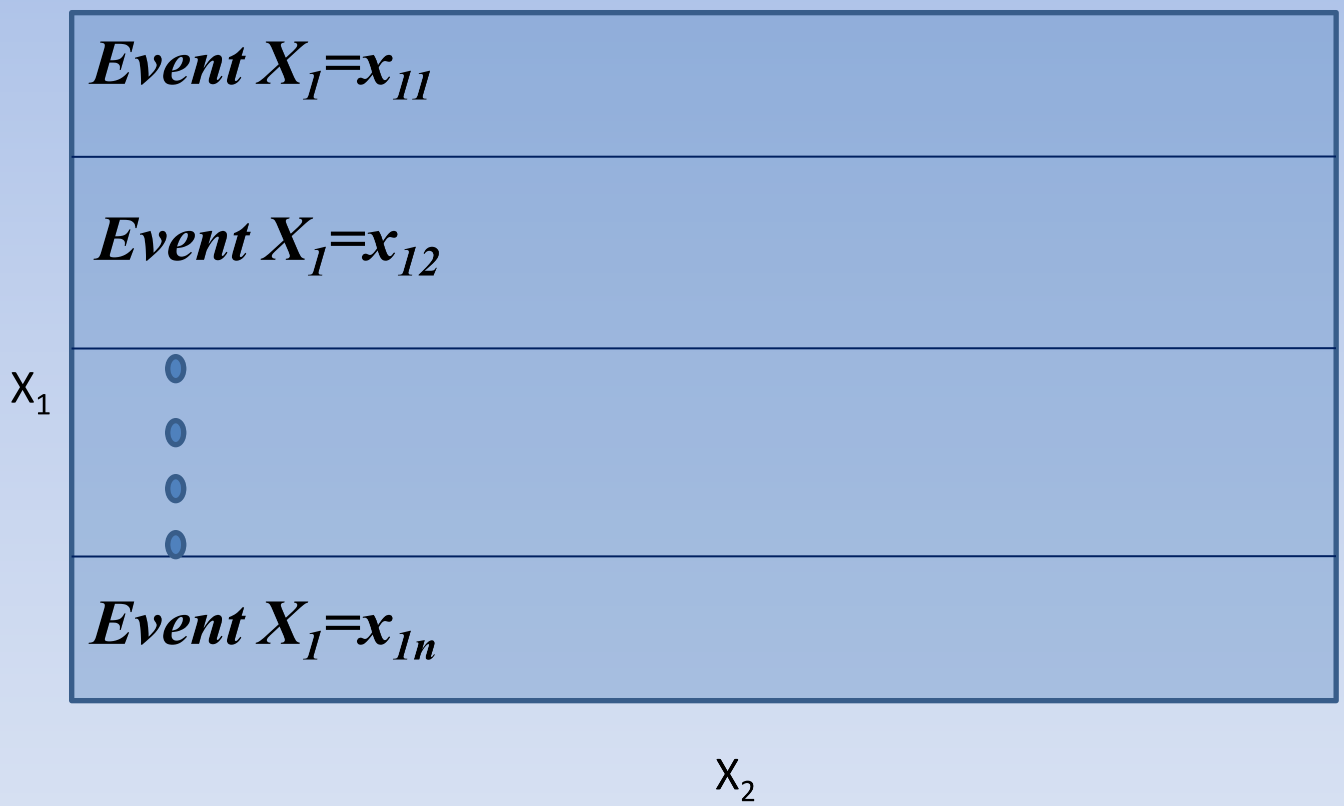
$$p_X(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\ x\in\{-\infty,+\infty\}$$

Joint PDF

- Using joint probability, we can define joint density functions for collections of random variables

$$p_{R,C}(R = x, C = y) = \begin{cases} 0.5 \text{ if } x = Yes, y = Yes \\ 0.2 \text{ if } x = No, y = Yes \\ 0.2 \text{ if } x = Yes, y = No \\ 0.1 \text{ if } x = No, y = No \end{cases}$$

All PDFs must sum to 1



Example

Joint Probability Density Function

CloudyTomorrow	RainTomorrow	WetGrass	Probability
No	No	No	0.4
No	No	Yes	0.01
No	Yes	No	0
No	Yes	Yes	0.01
Yes	No	No	0.15
Yes	No	Yes	0.02
Yes	Yes	No	0.01
Yes	Yes	Yes	0.4

Atomic Event

Event

Sample Space

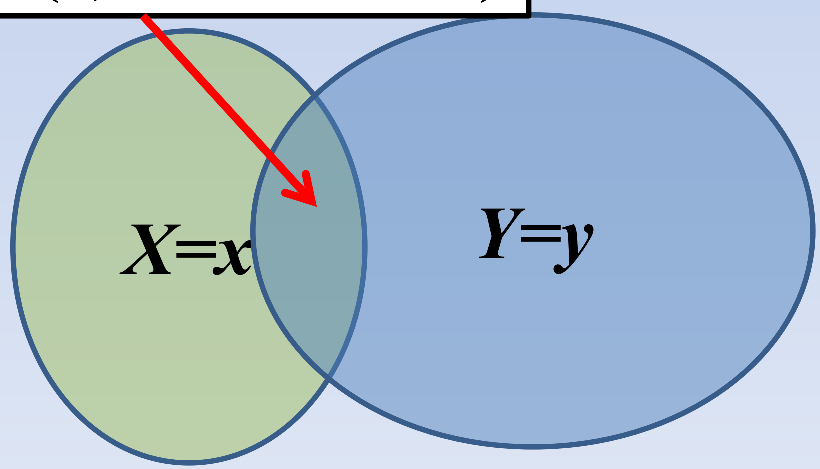
Terminology and Results

Conditional Probability

- The conditional probability of X **given** Y is:

$$p_{X|Y}(X=x \mid Y=y) = \frac{p_{X,Y}(X=x, Y=y)}{p_Y(Y=y)}$$

$X=x$, $Y=y$ (“,” means AND)



Product Rule

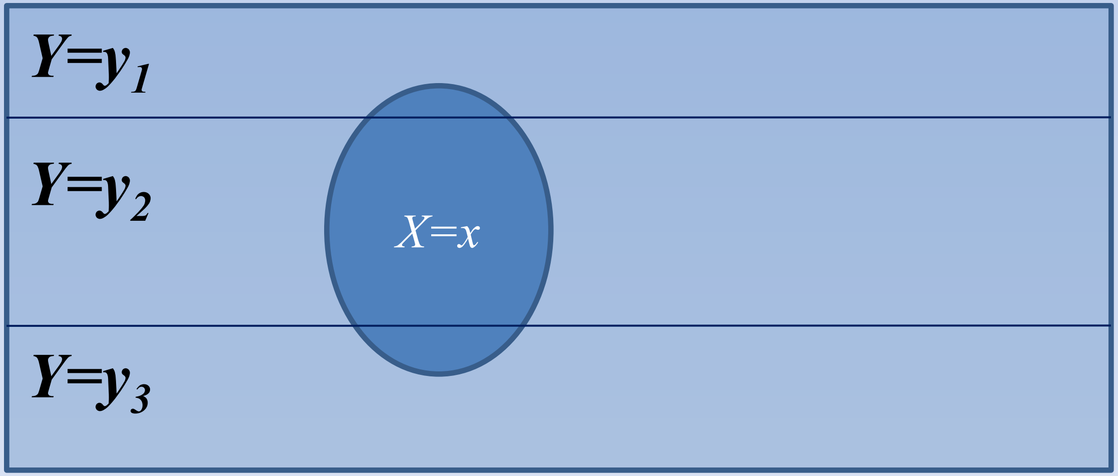
- From the definition of conditional probability:

$$p_{X,Y}(X = x, Y = y) =$$
$$p_Y(Y = y)\, p_{X|Y}(X = x \mid Y = y)$$

Marginalization

- For any two random variables X and Y:

$$p_X(X=x)=\sum_y p_{X,Y}(X=x,Y=y)$$



Conditioning

$$p(X = x) = \sum_y p(X = x, Y = y)$$

Marginalization

$$= \sum_y p(X = x \mid Y = y) p(Y = y)$$

Product Rule

Bayes' Rule

(Rev. Thomas Bayes 1763)

$$p(C=c \mid E=e) = \frac{p(C=c, E=e)}{p(E=e)}$$

Def. of Conditional Prob.

$$= \frac{p(E=e \mid C=c)\,p(C=c)}{p(E=e)}$$

Product Rule

$$= \frac{p(E=e \mid C=c)\,p(C=c)}{\sum_{c'} p(E=e \mid C=c')\,p(C=c')}$$

Conditioning

The importance of Bayes' Rule

- Let C be a random variable with values that are possible "causes"
- Let E denote a random variable with values that are possible effects of each cause
- It is often easy to specify $p(E=e|C=c)$, much harder to specify $p(C=c|E=e)$
- Bayes' Rule therefore allows us to **reason backwards over uncertain events**---fundamental to *learning*

Statistical Independence

- Two r.v.'s X and Y are statistically independent if

$$p_{X,Y}(X=x,Y=y)=p_X(X=x)p_Y(Y=y)$$

- If so, we can reason separately about x and y and then combine results---key factor in gaining efficiency (later)

Consequence

$$p_{X|Y}(X = x \mid Y = y) = p_X(X = x)$$

$$= \frac{p_X(X = x)\, p_Y(Y = y)}{p_Y(Y = y)}$$

$$= p_X(X = x)$$

Conditional Independence

- Two r.v.'s X and Y are conditionally independent given a third, R, if

$$p_{X,Y|R}(X = x, Y = y \mid R = r) =$$
$$p_{X|R}(X = x \mid R = r)\, p_{Y|R}(Y = y \mid R = r)$$

I.I.D. random variables

- A collection of r.v.'s is I.I.D. if they are
  - *independent* and
  - *identically distributed* (the density functions are the same)

Summarizing a PDF

- A PDF is a large table of numbers
- But generally, we don’t need to know the entire thing; often the “highlights” are enough
  - *Expectation* and *Variance*
  - (statistics)

Expectation

- The expectation of r.v. X is defined as:

$$E(X) = \sum_x x p_X(x)$$

- The “average value” of X under $p_X(x)$

Expectation example

- A coin has 0.99 probability of showing heads. You get \$0 if the coin shows heads, and \$10 else. How much do you *expect to get* if I toss the coin?

$$E(X) = \sum_x x p_X(x) = (0 * 0.99 + 10 * 0.01) = \$0.1$$

Expectation of a function

$$E(f(X)) = \sum_{x} f(x) p_X(x)$$

Variance

- The variance of r.v. X is defined as:

$$V(X) = E([X - E(X)]^2)$$

$$= \sum_x (x - E(X))^2 p_X(x)$$

- The "average spread" of values of a r.v. around the average of the r.v.

Variance example

- A coin has 0.99 probability of showing heads. You get \$0 if the coin shows heads, and \$10 else. What is the variance of your takings?

$$E(X) = \sum_x x p_X(x) = (0*0.99 + 10*0.01) = \$0.1$$

$$\begin{aligned} V(X) &= E([X - E(X)]^2) \\ &= (0-0.1)^2 * 0.99 + (10-0.1)^2 * 0.01 \\ &= 0.99 \end{aligned}$$

Variance example

- A coin has 0.99 probability of showing heads. You get $10 if the coin shows heads, and $0 else. What is the variance of your takings?

$$E(X) = \sum_x x p_X(x) = (10 * 0.99 + 0 * 0.01) = \$9.9$$

$$\begin{aligned} V(X) &= E([X - E(X)]^2) \\ &= (10 - 9.9)^2 * 0.99 + (0 - 9.9)^2 * 0.01 \\ &= 0.99 \end{aligned}$$

Variance example 3

- A coin has 0.5 probability of showing heads. You get $0 if the coin shows heads, and $10 else. What is the variance of your takings?

$$E(X) = \sum_x x p_X(x) = (0 * 0.5 + 10 * 0.5) = \$5$$

$$\begin{aligned} V(X) &= E([X - E(X)]^2) \\ &= (0-5)^2 * 0.5 + (10-5)^2 * 0.5 \\ &= 25 \end{aligned}$$