HW solutions at front

We'll have extra time at the end of lecture for questions

# Victory Lap

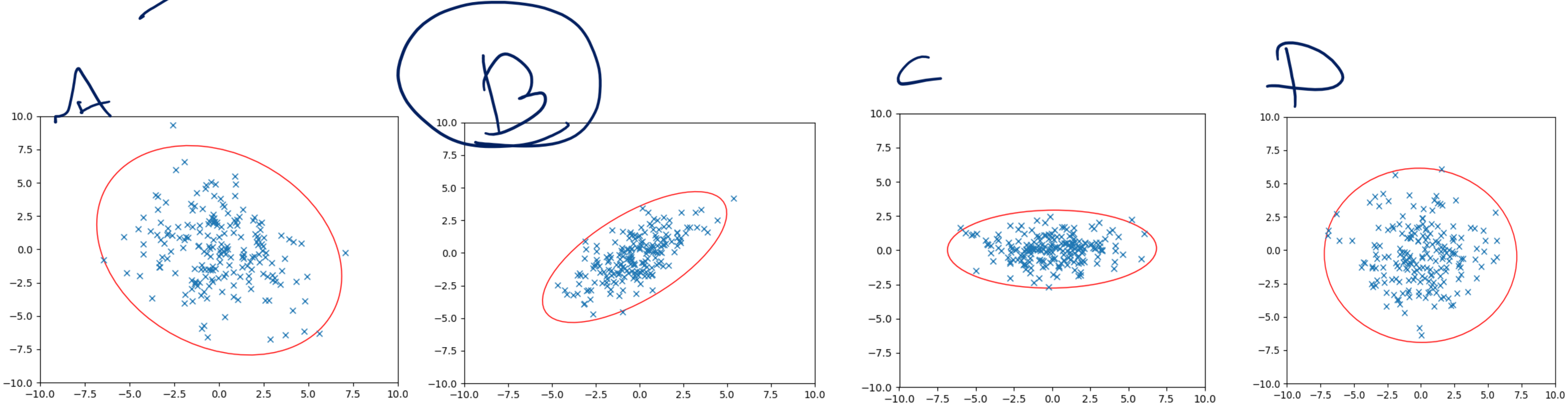
CSE 312 Winter 25
Lecture 27

# Dependence

Let's think about 2 dimensions.

Let $X = [X_1, X_2]^T$ where $\mathrm{Var}(X_1) = 3$, $\mathrm{Var}(X_2) = 3$ BUT $X_1$ and $X_2$ are dependent. $\mathrm{Cov}(X_1, X_2) = 2$

What is $\Sigma$? Which of these pictures are i.i.d. samples of $X$?



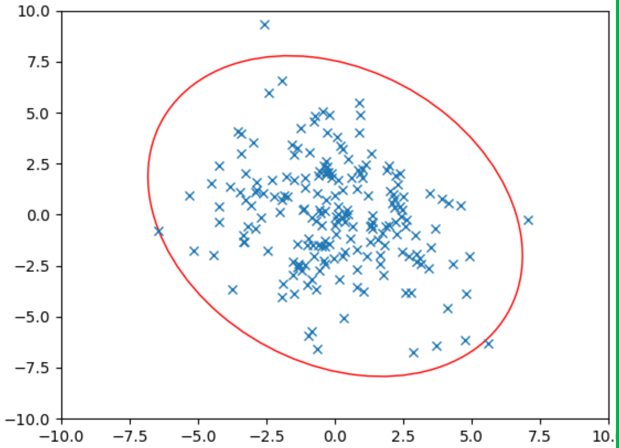
# Dependence

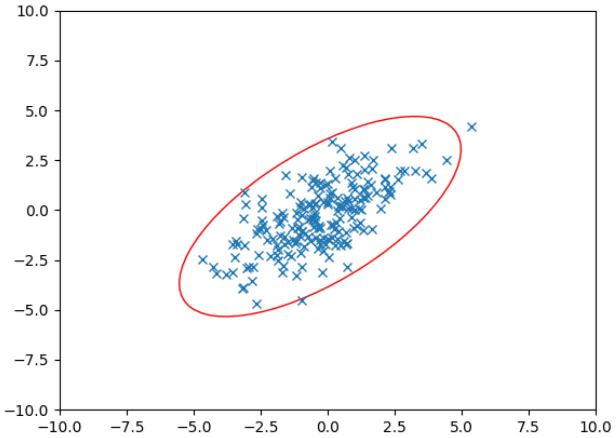
Let's think about 2 dimensions.

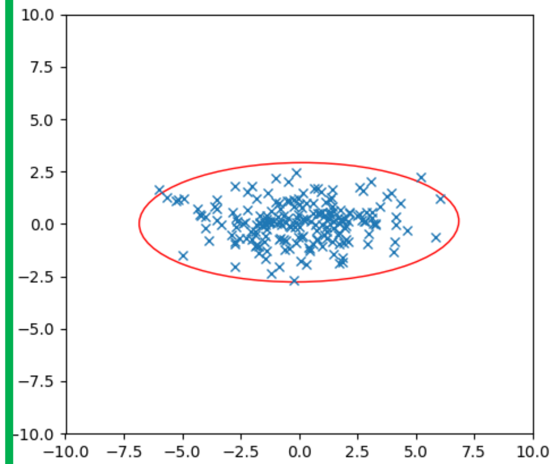
Let $X = [X_1, X_2]^T$ where $\mathrm{Var}(X_1) = 3$, $\mathrm{Var}(X_2) = 3$ BUT $X_1$ and $X_2$ are dependent. $\mathrm{Cov}(X_1, X_2) = 2$

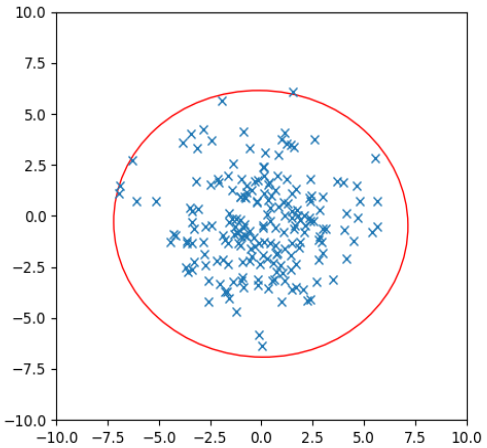
What is $\Sigma$? Which of these pictures are i.i.d. samples of $X$?

$$\Sigma = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$



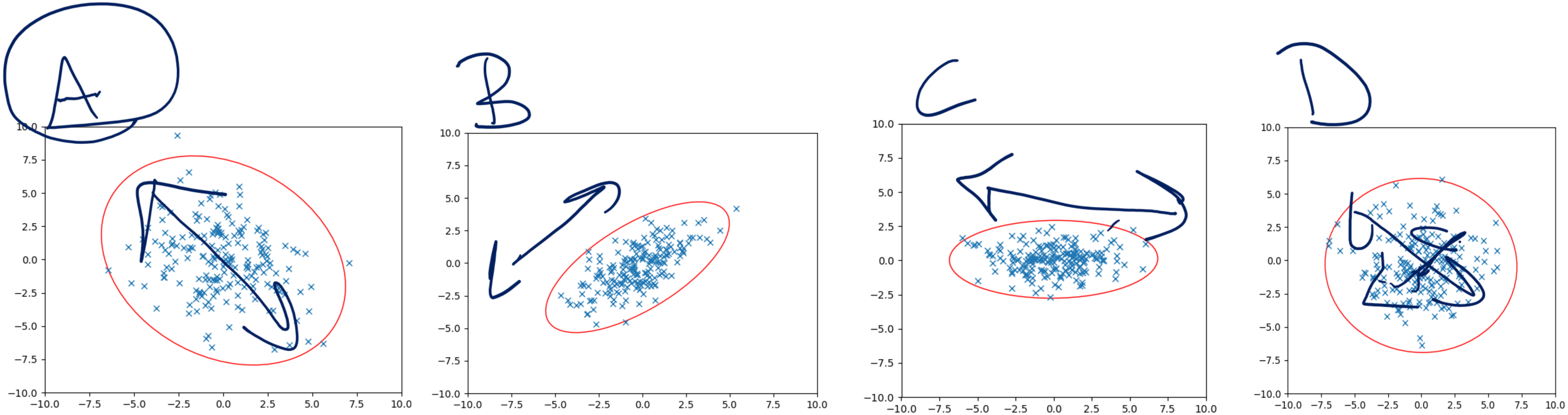



# Dependence

Let's think about 2 dimensions.

Let $X = [X_1, X_2]^T$ where $\mathrm{Var}(X_1) = 5$, $\mathrm{Var}(X_2) = 7$ BUT $X_1$ and $X_2$ are dependent. $\mathrm{Cov}(X_1, X_2) = -2$

What is $\Sigma$? Which of these pictures are i.i.d. samples of $X$?



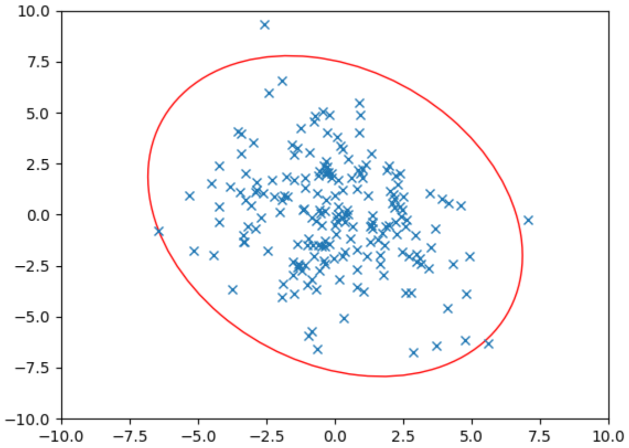
# Dependence

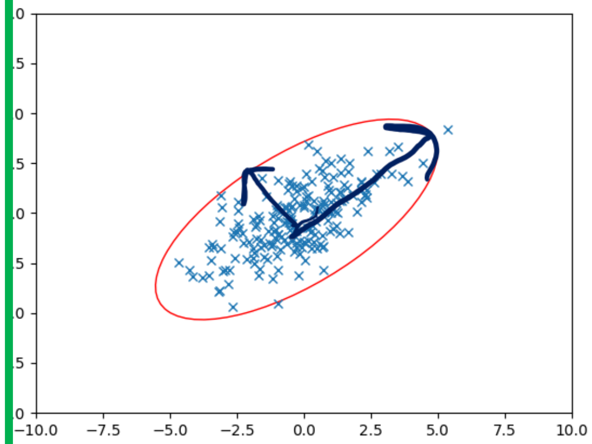
Let's think about 2 dimensions.

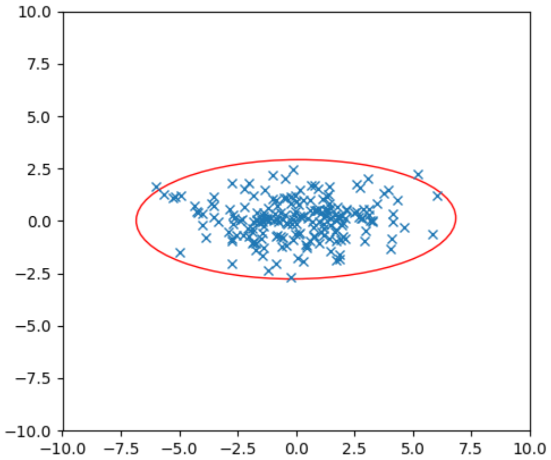
Let $X = [X_1, X_2]^T$ where $\mathrm{Var}(X_1) = 5$, $\mathrm{Var}(X_2) = 7$ BUT $X_1$ and $X_2$ are dependent. $\mathrm{Cov}(X_1, X_2) = -2$

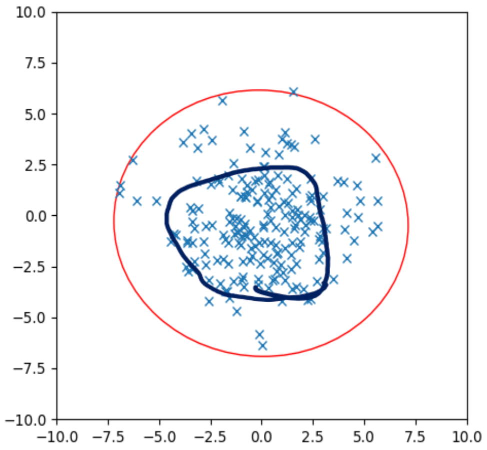
What is $\Sigma$? Which of these pictures are i.i.d. samples of $X$?

$$\Sigma = \begin{bmatrix} 5 & -2 \\ -2 & 7 \end{bmatrix}$$



# Using the Covariance Matrix

What were those ellipses in those datasets?

How do we know how many standard deviations from the mean a 2D point is, for the independent, variance 1 ones

Well $(x_1 - \mathbb{E}[X_1])$ is the distance from $x$ to the center in the $x$-direction.

And $(x_2 - \mathbb{E}[x_2])$ is the distance from $x$ to the center in the $y$-direction.

So the number of standard deviations is $\sqrt{(x_1 - \mathbb{E}[X_1])^2 + (x_2 - \mathbb{E}[x_2])^2}$

That's just the distance!

In general, the major/minor axes of those ellipses were the eigenvectors of the covariance matrix. And the associated eigenvalues tell you how the directions should be weighted.

# Victory Lap

# What Have We Done?

Well let’s look back…

# Content

## Combinatorics (*fancy* counting)

Permutations, combinations, inclusion-exclusion, pigeonhole principle

## Formal definitions for Probability

Probability space, events, conditional probability, independence, expectation, variance

## Common patterns in probability

Equations and inequalities, "zoo" of common random variables, tail bounds

## Continuous Probability

pdf, cdf, sample distributions, central limit theorem, estimating probabilities

## Applications

Across CS, but with some focus on ML.

# Themes

## Precise mathematical communication

Both reading and writing dense statements.

## Probability in the "real world"

A mix of CS applications

And some actual "real life" ones.

## Refine your intuition

Most people have some base level feeling of what the chances of some event are.

We're going to train you to have better gut feelings.

# Use Your Powers Wisely

We've seen probability can be used in the real world!

But also that it:

Can be counter-intuitive/hard to explain (Bayes Rule/Real World)

Probability estimates can depend on the model you're using (Real World)

Can be used to analyze ML applications, and think about the impacts of using them.

# How (not to) lie with statistics

You now know a lot of the tools that people use to lie with statistics. (See also: INFO 270)

Some patterns to watch out for:

My smoke alarm is going off, please pay for my new house! (analogy from Matt Parker)

Make a model, find that an event that occurred had small probability/fails some statistical test, claim that the **only** explanation is something nefarious occurred.

Better response: could the model be wrong? Is this statistical test appropriate? Once in 100 year events do happen…about once in every hundred years, is this just the one?

# How (not to) lie with statistics

See a story about testing?

Remember from Bayes' Rule that you need three numbers to understand a test. (3 of prior, posterior, false positive rate, false negative rate).

Headlines usually give you one number, that often isn't even one of the ones you need for Bayes ("this test is less accurate than a coin flip!").

The article itself, if you're lucky, might give you one or two of the numbers for Bayes – don't forget the prior!

# How (not to) lie with statistics

Before being impressed with a number, make sure you understand what it means.

Recent example for Robbie:

In baseball umpires decide whether a pitch is a strike or a ball (whether it goes through an (invisible) rectangle when thrown to the hitter)

There are camera systems built into stadiums that track the ball, and figure out where it went

In an infamous game an umpire missed 12% of the calls, according to an unofficial analysis of the data.

Or possibly 4% of the calls, according to the official analysis of the data.

# How (not to) lie with statistics

We can apply our knowledge to the real world!

But if you're applying in a new domain, get information from domain experts, don't instantly assume because you know Bayes' Rule that you know better than domain experts.

Don't hesitate to use these tools to understand new domains better!

But do keep in mind some things can't be quantified and just because we can use an algorithm doesn't mean we always should.

# What to take next?

ML (CSE 446) using probability, linear algebra, and other techniques to extract patterns from data and make predictions.

CSE 421 designing algorithms – very little direct probability, but the combinatorics we did at the beginning will be useful.

We also have a graduate level course in randomized algorithms, but it has a few more prereqs

CSE 447 Natural Language Processing

CSE 426 Cryptography

CSE 422 Modern Algorithms

Other things!

# Conditional Expectation Practice

# Practice with conditional expectations

Consider of the following process:

Flip a fair coin, if it's heads, pick up a 4-sided die; if it's tails, pick up a 6-sided die (both fair)

Roll that die independently 3 times. Let $X_1, X_2, X_3$ be the results of the three rolls.

What is $\mathbb{E}[X_2]$? $\mathbb{E}[X_2 | X_1 = 5]$? $\mathbb{E}[X_2 | X_3 = 1]$?

# Using conditional expectations

Let $F$ be the event "the four sided die was chosen"

$\mathbb{E}[X_2] = \mathbb{P}(F)\mathbb{E}[X_2|F] + \mathbb{P}(\bar{F})\mathbb{E}[X_2|\bar{F}]$

$= \frac{1}{2} \cdot 2.5 + \frac{1}{2} \cdot 3.5 = 3$

$\mathbb{E}[X_2|X_1 = 5]$ event $X_1 = 5$ tells us we're using the 6-sided die.

$\mathbb{E}[X_2|X_1 = 5] = 3.5$

$\mathbb{E}[X_2|X_3 = 1]$ We aren't sure which die we got, but…is it still 50/50?

# Setup

*Let* $E$ *be the event "*$X_3 = 1$*"*

$\mathbb{P}(E) = \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{4} = \frac{5}{24}$

$\mathbb{P}(F|E) = \frac{\mathbb{P}(E|F) \cdot \mathbb{P}(F)}{\mathbb{P}(E)}$

$= \frac{\frac{1}{4} \cdot \frac{1}{2}}{5/24} = \frac{3}{5}$

$\mathbb{P}(\bar{F}|E) = \frac{\mathbb{P}(E|\bar{F}) \cdot \mathbb{P}(\bar{F})}{\mathbb{P}(E)} = \frac{\frac{1}{6} \cdot \frac{1}{2}}{5/24} = \frac{2}{5}$ (we could also get this with LTP, but it's good confirmation)

# Analysis

$\mathbb{E}[X_2|X_3 = 1] = \mathbb{P}(F|X_3 = 1)\mathbb{E}[X_2|X_3 = 1 \cap F] + \mathbb{P}(\bar{F}|X_3 = 1)\mathbb{E}[X_2|X_3 = 1 \cap \bar{F}]$

Wait what?

This is the LTE, applied in the space where we've conditioned on $X_3 = 1$.

**Everything** is conditioned on $X_3 = 1$. Beyond that conditioning, it's LTE.

$= \frac{3}{5} \cdot 2.5 + \frac{2}{5} \cdot 3.5 = 2.9.$

A little lower than the unconditioned expectation. Because seeing a 1 has made it ever so slightly more probable that we're using the 4-sided die.