

CIS532v2: Estimating Probability Distributions



Live Session 1

Ernest Green

Today's Live Session

CIS532v2 - Course Info

Bayes Optimal Classifier, MLE, Naive Bayes

Pre-project coding exercises

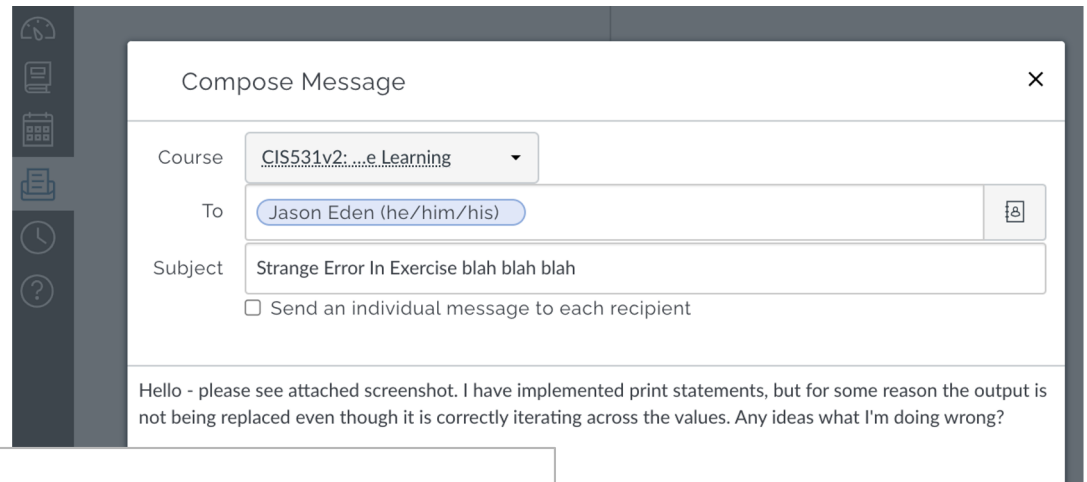
Reminder about avoiding loops whenever possible



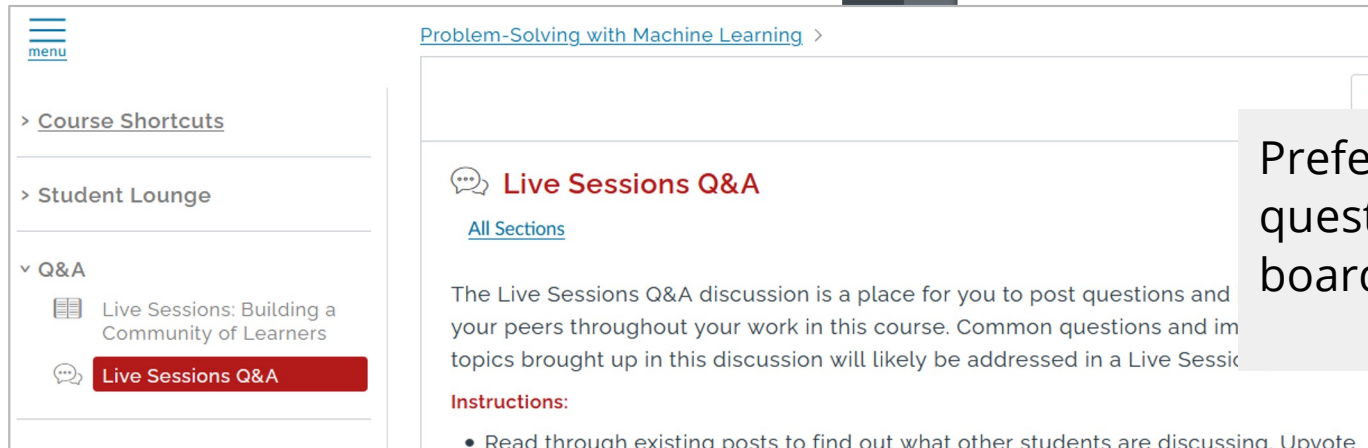
Ernest Green, Course Facilitator

How to get assistance:

- Private message me through Canvas



A screenshot of the Canvas LMS 'Compose Message' window. The window has a dark sidebar on the left with icons for home, messages, calendar, documents, and help. The main area is titled 'Compose Message' and contains the following fields: 'Course' (a dropdown menu showing 'CIS531v2: ...e Learning'), 'To' (a text field with 'Jason Eden (he/him/his)' and a user icon), and 'Subject' (a text field with 'Strange Error In Exercise blah blah blah'). Below these fields is a checkbox labeled 'Send an individual message to each recipient'. At the bottom, there is a text area with the message: 'Hello - please see attached screenshot. I have implemented print statements, but for some reason the output is not being replaced even though it is correctly iterating across the values. Any ideas what I'm doing wrong?'.



A screenshot of the Canvas LMS course page for 'Problem-Solving with Machine Learning'. The page has a left sidebar with a 'menu' icon and three sections: 'Course Shortcuts', 'Student Lounge', and 'Q&A'. The 'Q&A' section is expanded, showing 'Live Sessions: Building a Community of Learners' and 'Live Sessions Q&A' (which is highlighted in red). The main content area is titled 'Live Sessions Q&A' and includes a link to 'All Sections'. Below this, there is a paragraph of text: 'The Live Sessions Q&A discussion is a place for you to post questions and your peers throughout your work in this course. Common questions and im topics brought up in this discussion will likely be addressed in a Live Session'. Below the paragraph is an 'Instructions:' section with a bullet point: 'Read through existing posts to find out what other students are discussing. Upvote'.

Preferred: You can also post questions to the Q&A discussion board

Additional Resources

[Lecture 7 "Estimating Probabilities from Data: Maximum Likelihood Estimation" -Cornell CS4780 SP17](#)

<https://www.youtube.com/watch?v=RIawrYLVdlw>

[Machine Learning Lecture 8 "Estimating Probabilities from Data: Naive Bayes" -Cornell CS4780 SP17](#)

<https://www.youtube.com/watch?v=pDHEX2usCS0>

[Machine Learning Lecture 9 "Naive Bayes continued" -Cornell CS4780 SP17](#)

<https://www.youtube.com/watch?v=VDK0nkjFh5U>

[Machine Learning Lecture 10 "Naive Bayes continued" -Cornell CS4780 SP17](#)

<https://www.youtube.com/watch?v=rqB0XWoMreU>

[Machine Learning Lecture 11 "Logistic Regression" -Cornell CS4780 SP17](#)

<https://www.youtube.com/watch?v=GnkDzIOxfzl>

Avoid the Appearance of Plagiarism

Key points:

- **Show your work** - comment out, do not delete, mistakes, print statements, test cells, etc. Messy notebook that shows original thought is better than pristine code that happens to look like someone else's work
 - **Reference helpful websites** - put links to discussion boards, Q&A forums (Stack Overflow, etc.) that contained information you found useful in developing your code
 - **When in doubt, ask your facilitator!**
-

Statistics Refreshers

- Bayes' Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- Chain Rule

$$P(A, B, C, D) = P(A)P(B|A)P(C|A, B)P(D|A, B, C)$$

Bayes Basics

[The quick proof of Bayes' theorem](#)

https://www.youtube.com/watch?v=U_85TaXbelo

[Bayes theorem](#)

<https://www.youtube.com/watch?v=HZGCoVF3YvM>

[Naive Bayes, Clearly Explained!!!](#)

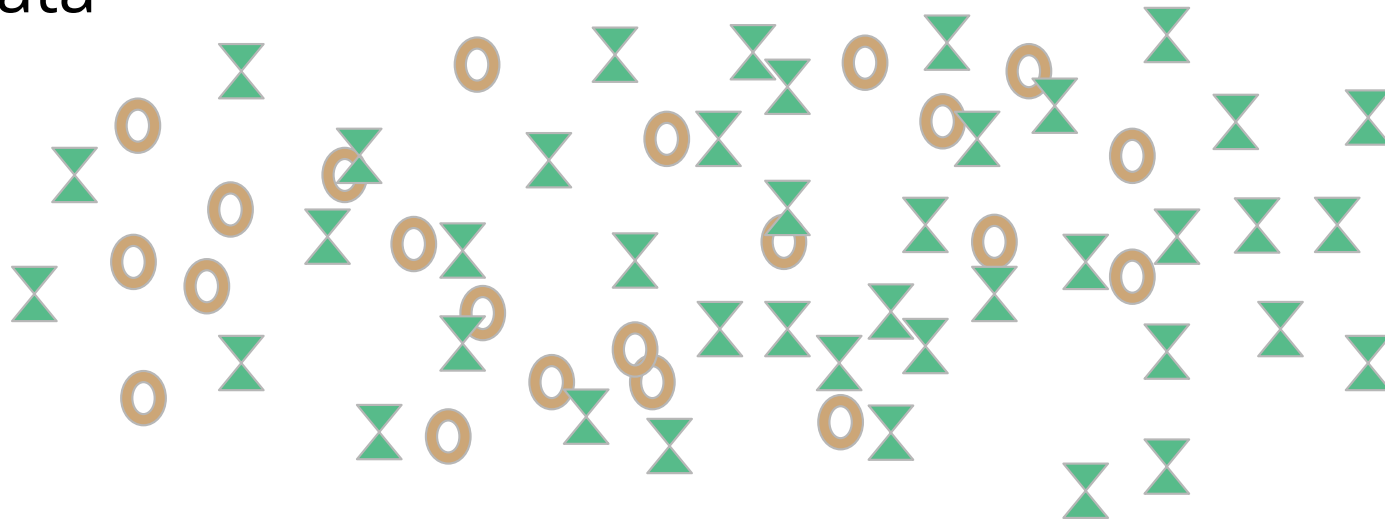
<https://www.youtube.com/watch?v=O2L2Uv9pdDA>

Distributions

- Gaussian (a.k.a. Normal)
 - Continuous outcome variables
 - Bell curve, symmetric at center
 - Values for mean = mode = median
 - Binomial
 - Two discrete outcome variables (ex: coin flips)
 - Other distribution types
- <https://www.unf.edu/~cwinton/html/cop4300/s09/class.notes/DiscreteDist.pdf>
-

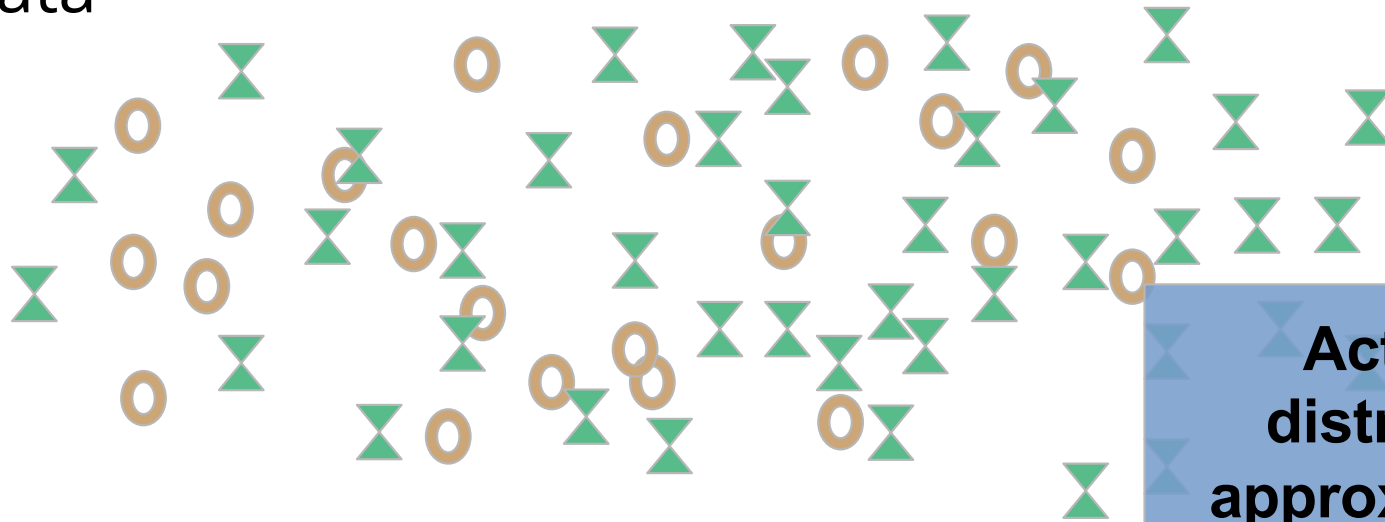
Estimating Distributions from Data

- **Bayes Optimal Classifier** - Bayes Theorem applied, but only works if you know the actual distribution of all data



Estimating Distributions from Data

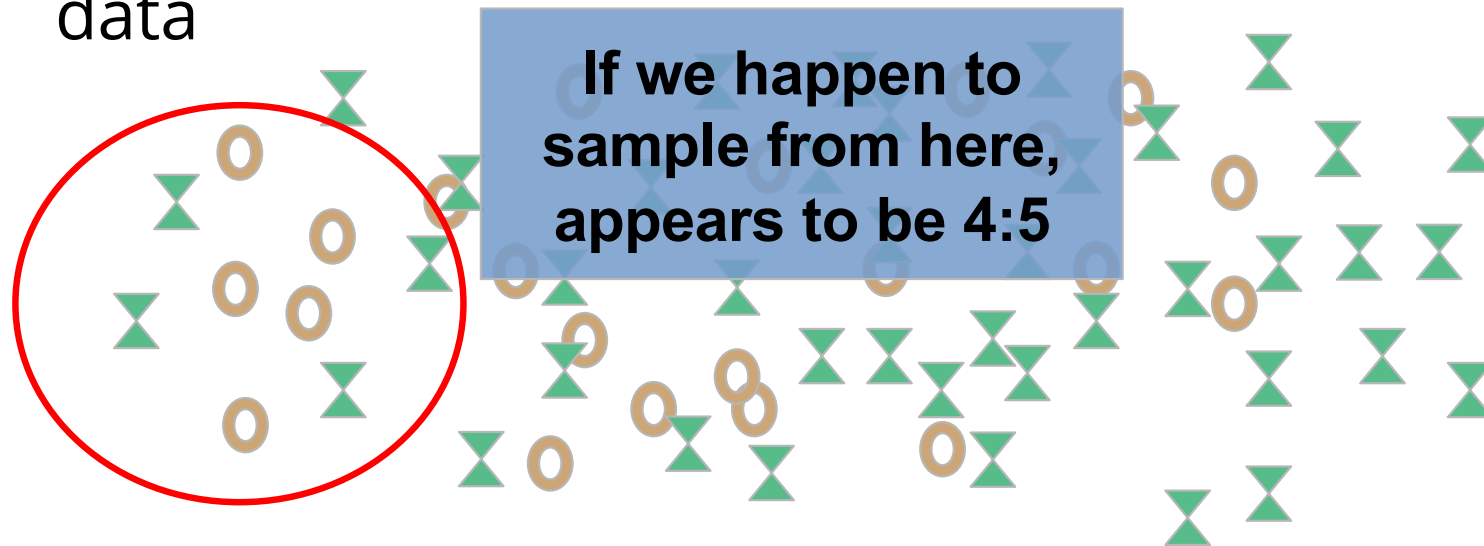
- **Bayes Optimal Classifier** - Bayes Theorem applied, but only works if you know the actual distribution of all data



**Actual data
distribution is
approximately 2:1**

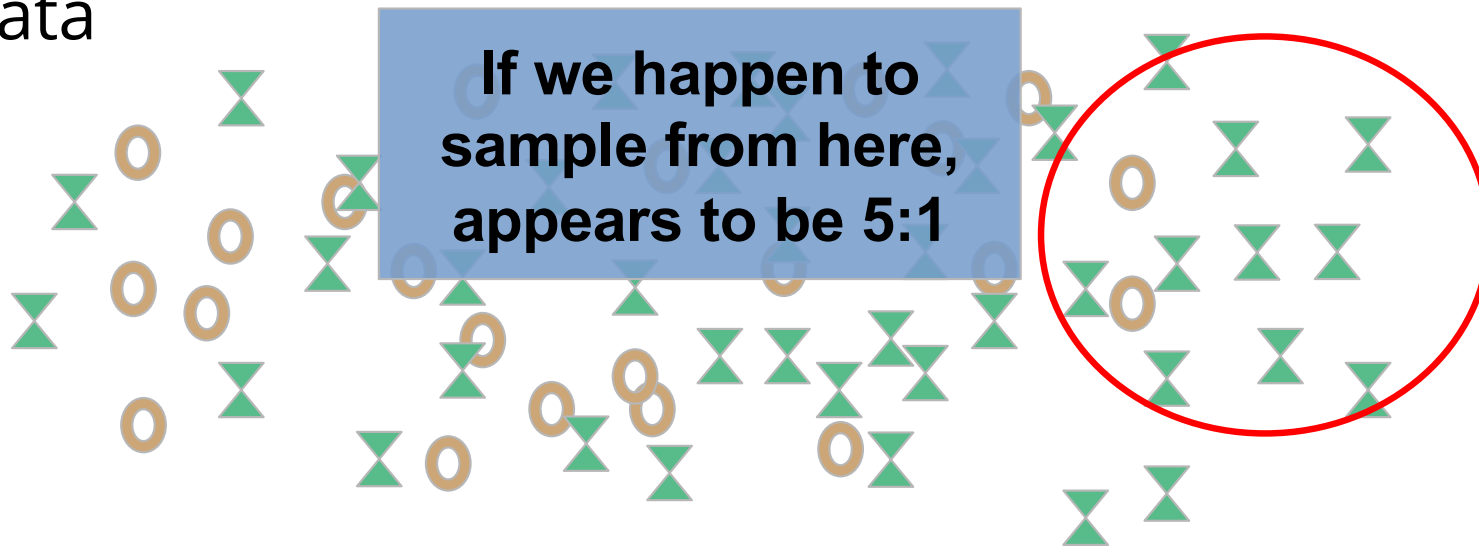
Estimating Distributions from Data

- **Bayes Optimal Classifier** - Bayes Theorem applied, but only works if you know the actual distribution of all data



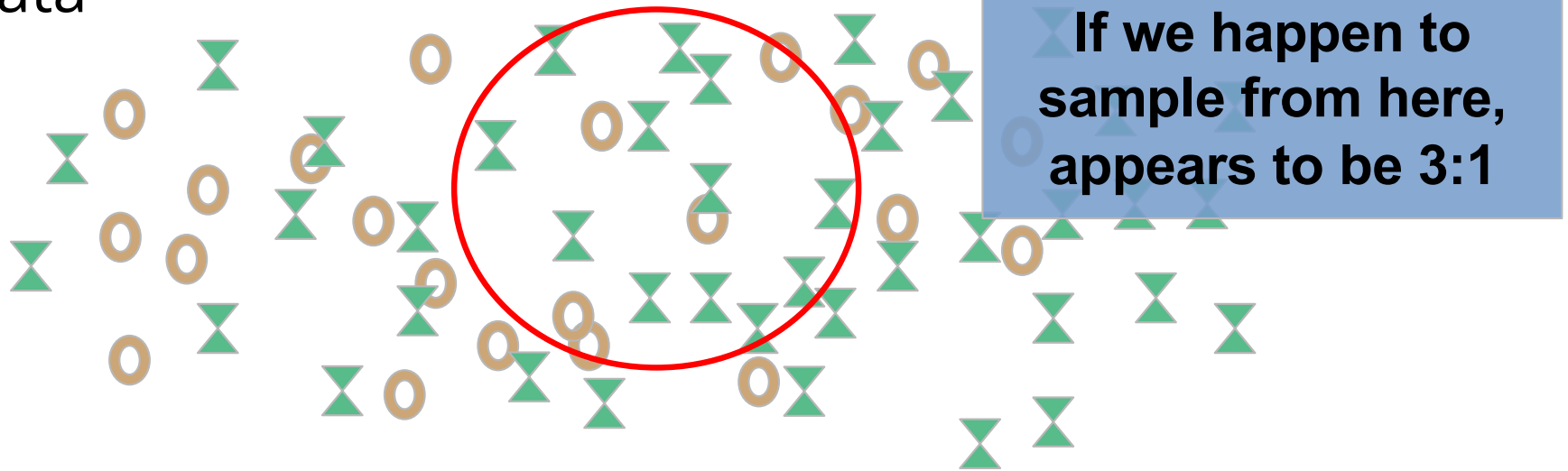
Estimating Distributions from Data

- **Bayes Optimal Classifier** - Bayes Theorem applied, but only works if you know the actual distribution of all data



Estimating Distributions from Data

- **Bayes Optimal Classifier** - Bayes Theorem applied, but only works if you know the actual distribution of all data



Estimating Distributions from Data

- **Bayes Optimal Classifier** - Bayes Theorem applied, but only works if you know the actual distribution of all data
 - ...and you never actually know the underlying distribution of your data



Maximum Likelihood Estimation

Step 1: Estimate the distribution of your data based on known parameters

Coin flip: it's either heads or tails, so binomial distribution

Take the data you observe, tweak those parameters until what you observed becomes as likely as possible (becomes maximized)

Estimating Theta

In a binomial distribution, theta is the likelihood of a target outcome

The probability of theta for given data approximately equals theta to the power of an observed outcome times 1-theta to the power of the other outcome

$$P_{\theta}(Data) \approx \theta^h (1 - \theta)^t$$

Simplifying Theta

But first... a note about log values



The Case for Log Values - Simplification

Log (natural log) is the inverse of Exponential

Has some interesting properties that can be used to simplify calculations:

$$\log(a * b) = \log(a) + \log(b)$$

$$\log(a / b) = \log(a) - \log(b)$$

$$\theta^h (1 - \theta)^t == h \log(\theta) + t \log(1 - \theta)$$

Simplifying Theta - First Derivative

To maximize a function, take the first derivative and set it equal to zero

$$h \log(\theta) + t \log(1 - \theta) \rightarrow \frac{h}{\theta} + \frac{t}{1 - \theta}(-1) = 0$$

If you're rusty on derivatives of logs (note - log and natural log are used interchangeably in portions of the course material, and `numpy.log()` refers to natural log):

<https://www.youtube.com/watch?v=Dp9sgIvaKPk>

Simplifying Theta - Solve for Theta

$$\frac{h}{\theta} + \frac{t}{1-\theta}(-1) = 0 \quad \rightarrow \quad \theta = \frac{h}{t+h}$$

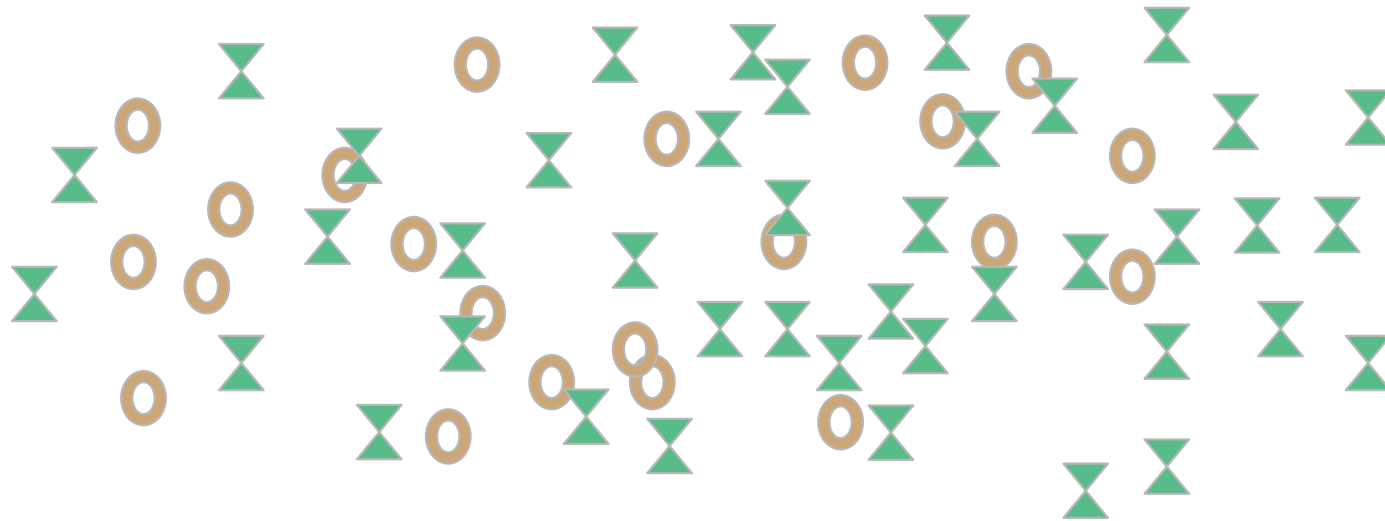
Simplifying Theta - Solve for Theta

$$\frac{h}{\theta} + \frac{t}{1 - \theta}(-1) = 0 \quad \rightarrow \quad \theta = \frac{h}{t + h}$$

...which makes intuitive sense, but now we have the mathematical underpinning.

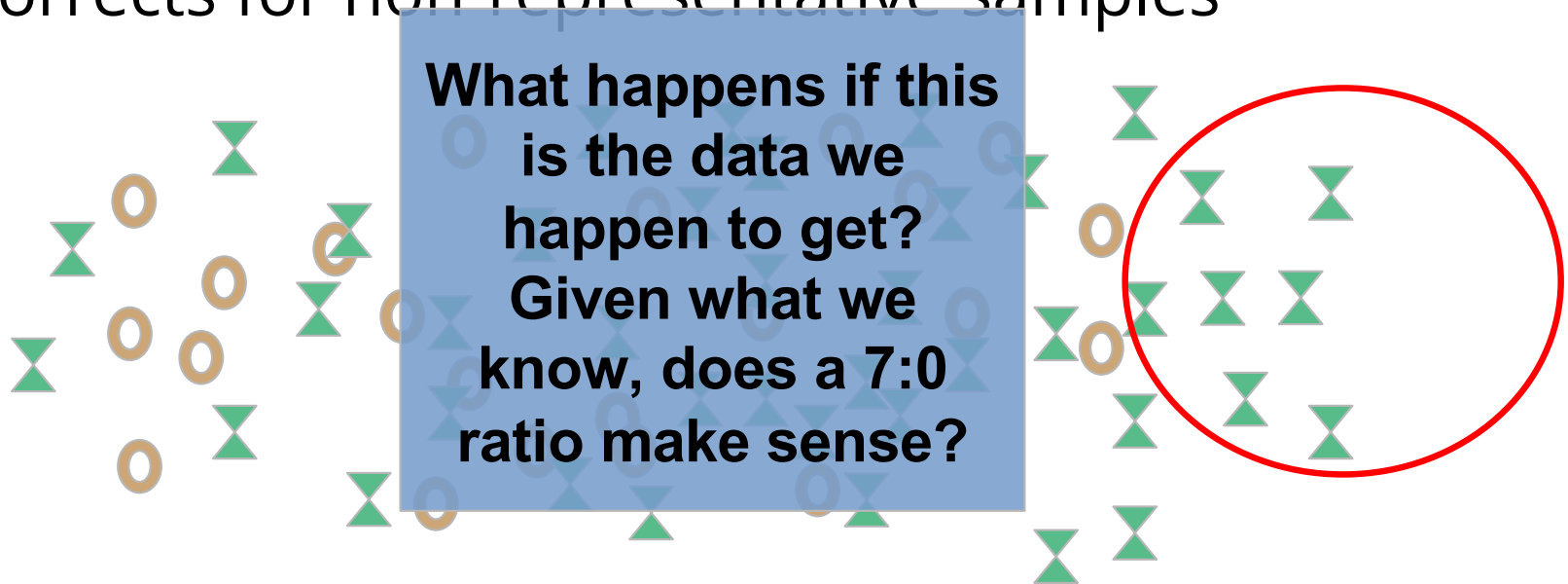
Smoothing - Why?

- Avoids “divide by 0” or “multiply by 0” errors
- Corrects for non-representative samples



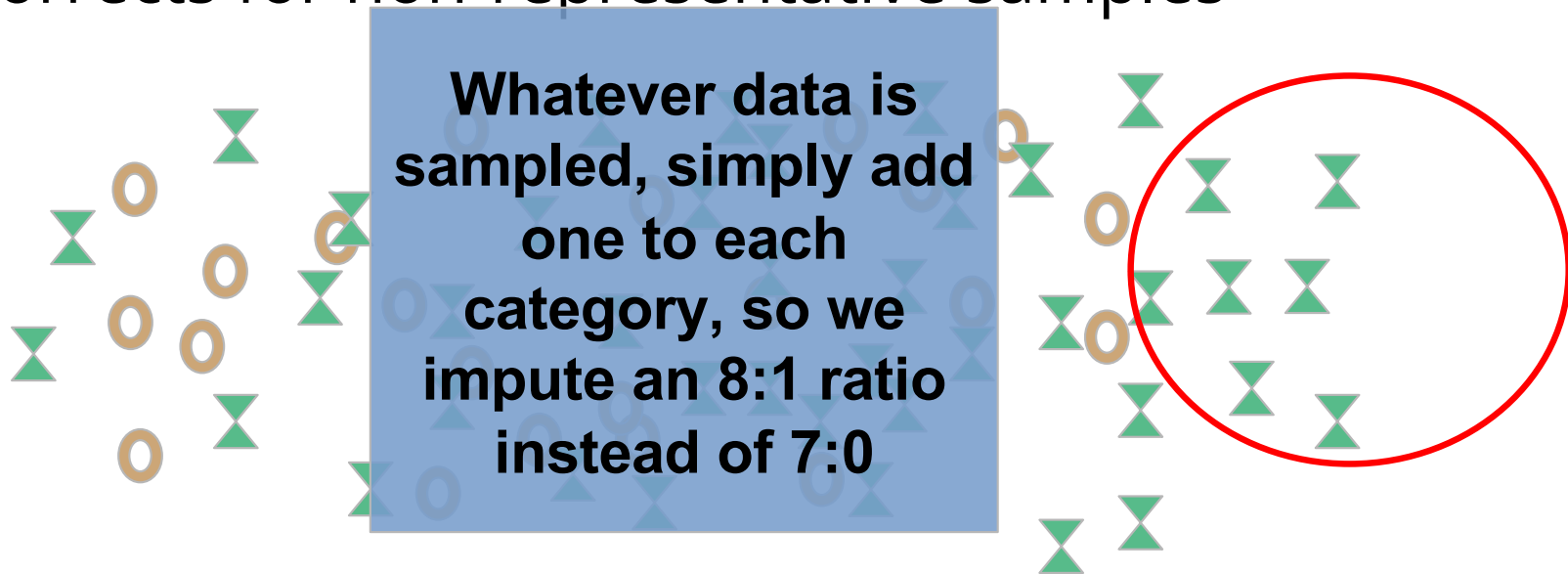
Smoothing - Accounting for Unlikely Samples

- Avoids “divide by 0” or “multiply by 0” errors
- Corrects for non-representative samples



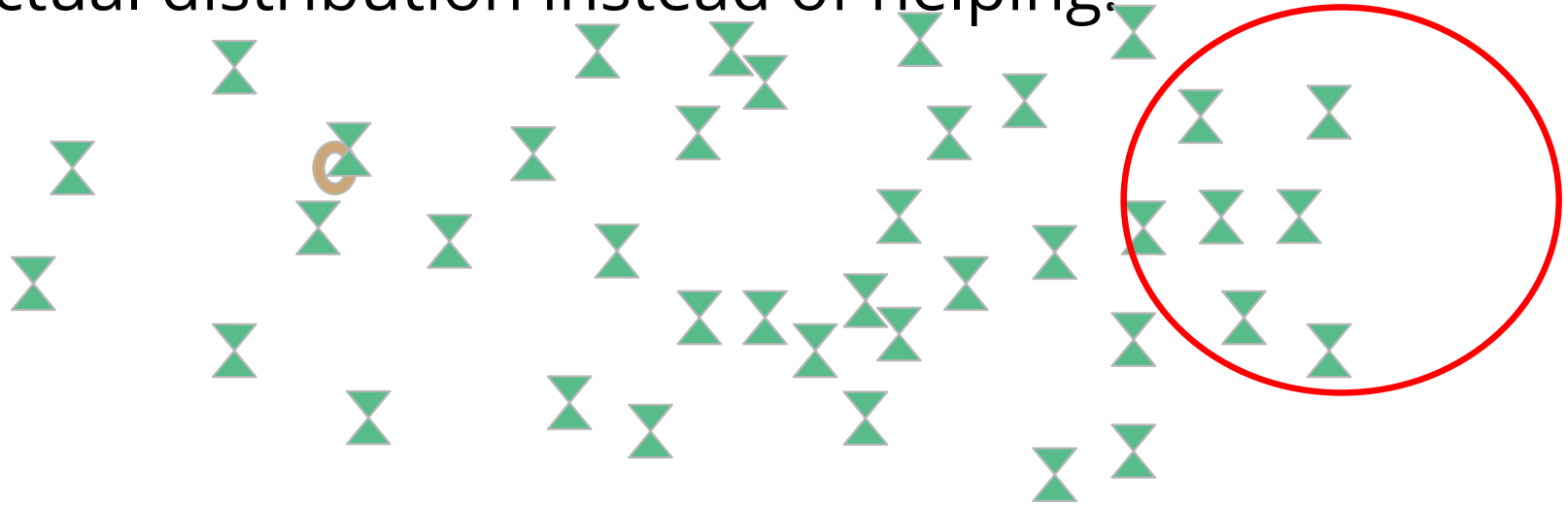
Smoothing - Laplace (+1)

- Avoids “divide by 0” or “multiply by 0” errors
- Corrects for non-representative samples



Smoothing - Beware of Assumptions

- If our data ratio really was 40:1 or similar, our +1 smoothing would hinder our progress in getting to the actual distribution instead of helping!



Additional Reading

- Maximum A Posterior (MAP)
 - <https://wiseodd.github.io/techblog/2017/01/01/mle-vs-map/>
- Basically: MLE makes distribution assumptions, whereas MAP assumes prior knowledge is used to determine theta value



Classification with the Naive Bayes Algorithm

- MLE Curse of Dimensionality
 - The higher d , lower chance of there being an exact match
 - Thus, lots of cases where no probability exists
 - Naive Bayes Assumption
 - “Naively” assume that all data points conditionally independent, even if you know they’re not
 - Allows us to use chain rule to determine probabilities - exact match not required for labeling
-

Naive Bayes in Action

Binary features:

C = "Guy wears a cape."

M = "Guy wears a mask."

U = "Guy wears his underwear outside his pants"

Labels:

G = "Guy is good"

B = "Guy is bad"

Class Prior Probabilities

$P(G)$

$P(B)$

Probabilities of Features for "Good" Class

$P(C|G)$

$P(M|G)$

$P(U|G)$

Probabilities of Features for "Bad" Class

$P(C|B)$

$P(M|B)$

$P(U|B)$

Naive Bayes in Action

Question 1: Estimate the probability that Suzie's date wears a mask and cape, given that he is good, using the Naive Bayes assumption.

$$P(M, C | G)$$

Multiply $P(M | G)$ by $P(C | G)$

Question 2: Before Suzie meets her date, what's the probability he is wearing a mask and cape (she does not know if he is good or bad)? (Use NB Assumption)

$$P(M, C) = P(M, C | G)P(G) + P(M, C | B)P(B)$$

Naive Bayes in Action - Question 3

Using Naive Bayes assumption and Bayes' Rule, predict the probability that he is good if he wears a mask and a cape. Hint: $P(G|M,C)$

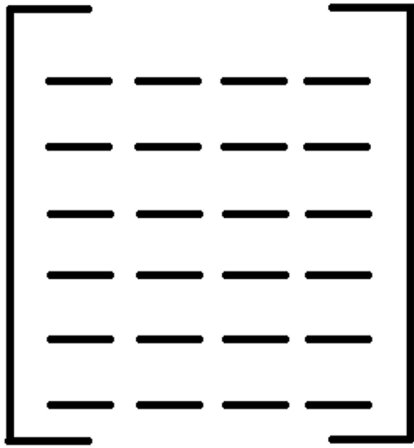
$$P(G|M,C) = \frac{P(M,C|G) P(G)}{P(M,C)}$$

[Machine Learning Lecture 9 "Naive Bayes continued" -Cornell CS4780 SP17](#)

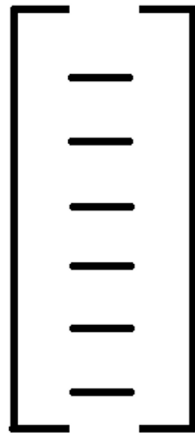
<https://www.youtube.com/watch?v=VDK0nkjFh5U>

Reminder: Avoid Loops, Utilize Indexing and Slicing

Matrix X



Matrix Y



```
ind = X[condition]  
subset_Y = Y[ind]
```

```
I, D = findknn(xTr, xTe, k)
```

```
yTe = yTr[I]
```

```
ind = Y[condition]  
subset_X = X[ind]
```

Questions?



Thank You For Attending!

End of Live Session 1

