



If the cookie had candy, then very few bites would have no candy.



If the cookie had no candy, then every bite would have no candy.



The probability of a no-candy bite, given a candy cookie, is 1/3.



The probability of a no-candy bite, given a no-candy cookie, is 1.

CS 1671/2071 Human Language Technologies

Session 3: Linear algebra, probability review

Michael Miller Yoder

January 15, 2025



School of Computing and Information

Overview: Linear algebra and probability review

- 1. Course logistics
- 2. JupyterHub setup and preprocessing activity
- 3. Probability review
- 4. Linear algebra review

Course logistics

- No class next Mon for MLK Day
- Next class is next Wed Jan 22
- <u>Homework 1</u> is **due next Thu Jan 23**

JupyterHub setup and activity

Set up Python virtual environment

- 1. Go to this <u>nbgitpuller link</u>
 - Log in with your Pitt username
 - Start a server with Teach 6 cores, 3 hours
 - This should pull a folder (cs1671_jupyterhub) into your JupyterLab
- 2. Open a terminal



Set up Python virtual environment

In a terminal, run
sh install_kernel.sh

	C cs1671_jupyt ((auto-8) - Jupyte × 🛛 🕂		~			×
÷	\rightarrow C O	음 ब्ª https://jupyter.c	rc. pitt.edu /user/mmyoder/lab/workspaces/auto-8/tree/cs1671_jupyterhub	☆	<u>*</u> হ	ב ב	≡
\bigcirc	File Edit View	Run Kernel Tabs S	ettings Help				
	+ 🗈	± C	mmyoder@teach-cpu-n1:~/c× +				°.
0	Filter files by na	ime Q	[mmyoder@teach-cpu-n1 cs1671_jupyterhub]\$ sh install_kernel.sh [InstallKernelSpec] WARNING Config option `kernel_spec_manager_class` not	recognize	d by `Inst	tallK	
	🖿 / cs1671_jupyte	erhub /	ernelSpec`. [InstallKernelSpec] Installed kernelspec cs1671 in /ihome/cs1671 2025s/mmvoc	der/.local.	/share/iur	ovter	Ť
:=	Name 🔺	Last Modified	/kernels/cs1671	,,	, J-F	.,	
	cs1671	2 minutes ago	[mmyoder@teach-cpu-n1 cs16/1_jupyterhub]\$				
	🖿 old	2 minutes ago					
	install_kernel.sł	h 2 minutes ago					
	README.md	2 minutes ago					
	• 🔳 session2_text_r	n 2 minutes ago					

Open Jupyter Notebook

- Double-click
 session2_text_normalization
 .ipynb on the left panel to open
 the notebook
- 2. From the top menu, click Kernel> Change Kernel...
- 3. Select **cs1671** as your kernel -
- 4. Run the first code cell under
 Test kernel and environment
 that imports pandas and nltk



Preprocessing Airbnb listings

Implementation

- Remove undesired text with regular expressions
- Lowercase
- Remove stopwords
- Tokenize with the NLTK package
- Stem the tokens with NLTK

Saving your work

C C Image: Section 2 Imag	> C																	
File Edit View Remet Table Help New Auscher 0 At 2000 0		jupyter.crc.pitt.edu	/user/gnowmik/l	/lab/work	spaces/a	uto-5/tre	e/Untitle	d7.ipynl	b		\$	6	•	9	m	-	Ð	
New Jaynb New Lauchter 9 KL Open from Taht Open from VBL Open from VBL Twee for Notebook New Consol for Notebook New New Consol for Notebook New Close Tab New Save All Tabs Image: Same All Tabs Dominad Image: Same All Tabs Dipulate Notebook Pint	File Edit Vie	ew Run Kernel Tal	bs Settings H	Help														
New Lancher • X L X • • • • • • • • • • • • • • • • • • •	New		> 17.ip	pynb	٠	+												
Open from URL Open from URL New Cwer for Notabook Close Tab X W Close Tab X W Close Tab Save All Tabe Save Notebook Save Notebook As Save Notebook As Save Notebook As Save Notebook Durlicate Notebook As Durlicate Notebook As Save Current Workspace As Print X P Hub Control Panel Log Out Downstab Durlicate Notebook. Dispyretmb Ist year Imatab 2 years ago Imatab	New Launch	her	°*L X	00	► =	C ++	Code	~	÷				Notebo	ook 🖸	ö	Pyth	on 3 (ip	yker
Open from URL * New Vew for Notebook New Ornel for Notebook Close and Shut Down Notebook ^ O Q Close and Shut Down Notebook ^ O Q Close and Shut Down Notebook Swe Notebook Mew Kotebook from Disk Revert Notebook Duplicate Notebook Duplicate Notebook As Swe and Export Notebook As Duplicate Notebook As Swe and Export Notebook As Swe and Export Notebook As Duplicate Notebook. Download Swe Current Workspace As Swe Current Workspace As Swe Current Workspace As Swe Current Workspace Print XP Hub Cottrol Panel Log Od Diptysterhab Ist year In matab 2 years ago In matab 2 wars ago In mataba 1 withitedal 1 days ago Yurithtodal 9 days ago Yurithtodal 9 days ago Yurithtodal 9 days ago	Open from P	Path																
New View for Notebook New Canada for Notebook Close Tab Close Tab Save Notebook Save Notebook Save Notebook Save All Redaa Notebook Save All Redaa Notebook Redaa Notebook Save All Redaa Notebook Save All Save Current Vorkspace Print Save Current Vorks	Open from L	URL	11											ſ	• ^	\downarrow	÷ '	Ŧ
New Console for Notebook Close Tab Close and Shut Down Notebook Cose and Tabs Save Notebook Save Notebook Save Notebook from Disk Rename Notebook. Rename Notebook. Download Save Current Workspace As Nettoold As Download Save Current Workspace As Save Current Workspace As	New View fo	or Notebook																
Close Tab V W Close and Shut Down Notebook	New Consol	le for Notebook																
Close and Shut Down Notebook ^ Close All Tabs Save Notebook * XS Save Notebook * XS Save Notebook * Save Notebook * Save All * Relead Notebook Tor Disk Revert Notebook. Duplicate Notebook. Duplicate Notebook. Duplicate Notebook Download Save and Export Notebook As Save Current Workspace As Save Current Workspace Print XP Hub Control Panel Log Out Dispyrentub Itary yar Mub Control Panel Dispyrentub Itary yar Dispy	Close Tab		35 W															
Close All Tabs Save Notebook X X S Save Notebook X O X S Save All Reload Notebook to Checkpoint Rename Notebook Duplicate Notebook C Download Save and Export Notebook As Save Current Workspace As	Close and S	hut Down Notebook	^ 0 0															
Save Notebook 45	Close All Tai	bs																
Save Notebook As Save All Reload Notebook To Disk Revert Notebook to Checkpoint Rename Notebook. Duplicate Notebook Download Save and Export Notebook As Save Current Workspace As Save Current Workspace As Save Current Workspace Print M P Hub Control Panel Log Out Dipyterhub I ast year matab 2 years ago matab 2 years ago matab 2 years ago M Unitided M P Unitided M P Unitided M P Unitided M P Mub Control Panel Log Out M Unitided M P M P M P M P M P M P M P M P	Cours Nataba																	
Save All Relada Notebook from Disk Revert Notebook to Checkpoint Rename Notebook Duplicate Notebook Duplicate Notebook As	Save Notebu	ook As	0 # 5															
Reload Notebook from Disk Revert Notebook to Checkpoint Rename Notebook Duplicate Notebook Download Save and Export Notebook As Save Current Workspace As Save Current Workspace As Save Current Workspace As Print Dipyternub Ist year matiabcr Ist year matiabcr Ist year matiabcr Ist year Nor envords Untitled.l 14 days ago 14 Untitled.l 9 days ago 14 Untitled.l 9 days ago	Save All	www.codin																
Revert Notebook to Checkpoint Renam Notebook. Duplicate Notebook. Download Save and Export Notebook As , Save Current Workspace As Save Current Workspace As Save Current Workspace Print X P Hub Control Panel Log Out Dipyterhub last year Inatlab 2 years ago Current Workspace Dipyterhub last year Inatlab 2 years ago Chuttled.i 14 days ago Curtel 14 days ago Muttled.i 14 days ago	Delevel	haali faan Diala	-															
Remain Notebook Duplicate Notebook Download Save and Export Notebook As Save Current Workspace As Save Current Workspace Print X P Hub Control Panel Log Out Diggenrou Digge	Revert Note	book from Disk																
Duplicate Notebook Duplicate Notebook Save and Export Notebook As	Rename Not	tehook																
Download Save and Export Notebook As Save Current Workspace As Deploymentoc Deploymentoc Imatab 2 years ago Dematab 2 years ago Dematab 2 years ago Muttidelp 1ast year Demover 1bour ago Muttidelp 14 days ago M Untiteds1 14 days ago M Untiteds1 9 days ago M Untiteds1 9 days ago M Untiteds1 9 days ago	Duplicate No	otebook																
Save Current Workspace As Save Current Workspace As Save Current Workspace Print X P Hub Control Panel Log Out Dipuperhub last year Dipuperhub last y	Download		_															
Save Current Workspace As Save Current Workspace Print X P Hub Control Panel Log Out jupyterhub last year matlab 2 years ago matlab 2 years ago matlab 2 years ago matlab.cr last year noverworks 1 hour ago MutitedJ 14 days ago MutitedJ 14 days ago MutitedJ 14 days ago MutitedJ 9 days ago MutitedLi 9 days ago MutitedLi 9 days ago	Save and Ex	port Notebook As	,															
Save Current Workspace As Save Current Workspace As Save Current Workspace As Save Current Workspace As Print Print Dipptentub D	Cause Current	Washanaan An																
Print X P Hub Control Panel Log Out Dippremou Iow Dippremou Iow Dimatab_cr Iast year Dimatab_cr Iast year<	Save Curren	t Workspace																
Film K P Hub Control Panel Log Out Log Out Isst year Dipyremou Isst year Dimatab_cr Isst	Delet	it monopuee	-															
Hub Control Panel Log Out paymentations paymentations paymentations matabe	Print		AP															
Ligg Out pupyrembu Iow pupyrembu Iast year matlab_cr Iast year matlab_cr Iast year method age Mutitled.ip 14 days age Mutitled.i 14 days age Mutitled.i 14 days age Mutitled.i 14 days age Mutitled.i 9 days age Mutitled.i 9 days age	Hub Control	I Panel																
Importentub Iast year Immatibab_cr Id days ago Immatibab_cr Immatibab_cr Immatibab_cr Immatibab_cr <	Log Out		_															
Imataba 2 years ago Imataba_cr Isst year Imataba_cr Isst year Imataba_cr Ihour ago	🗅 jupyterhub.	last year																
D matlab_cr last year D new-works 1 hour ago R Unitided.ip 14 days ago R Unitided.i 9 days ago R Unitided.i 9 days ago R Unitided.i 9 days ago	🗅 matlab	2 years ago																
New-works 1 hour ago R Untitled.ip 14 days ago M Untitled.i 9 days ago M Untitled.i 9 days ago	matlab_cr	last year																
Munitided.j 14 days ago Munitided.l 9 days ago Munitided.l 9 days ago Munitided.l 9 days ago Munitided.l 9 days ago	C new-works	1 hour ago																
Image: Construction of the state of the	 Untitled.ip. 	14 days ago																
Imitted 2.1 14 days ago Imitted 3.1 14 days ago Imitted 4.1 9 days ago Imitted 5.1 9 days ago Imitted 5.1 9 days ago	Untitled1.i.	14 days ago																
Muturbust Hutited4.i 9 days ago Muturbust 9 days ago Muturbust 9 days ago Muturbust	Untitled2.i.	14 days ago																
UnititedS.i 9 days ago IunititedS.i 10 hour ago	Untitled4 i	9 days ago																
Lutitledsi	Untitled5.i																	
	-																	

C Hetitled? isush (sute 5) - is

Ending your session

Be sure to save your work before ending the session

- Select File > Hub Control Panel
- 2. Click Stop My Server



Now	Kull Kerner	Tawa	Jetting		neip								
New Launcher		0	₩L,	37.1	pynb	ŕ	Þ	:	+ C	**	Code	~	B
Open from Path													-
Open from URL				1									
New View for Not	ebook			L									
New Console for	Notebook			Ł									
Close Tab	b laashli		w 7	L									
Close and Shut D Close All Tabs	own Notebook	~	οQ	L									
Save Notebook			×s	1									
Save Notebook A	S	0	×s	L									
Save All													
Reload Notebook	from Disk			L									
Revert Notebook	to Checkpoint			L									
Duplicate Notebo	ok			L									
Download				1									
Save and Export 1	Notebook As		,	1									
Save Current Wor	kspace As			1									
Save Current Wor	kspace			L									
Print			¥ P	1									
Hub Control Pane	el .												
Log Out	z seconus ago	_		J									
🗅 jupyterhub	last year	1											
🗅 matlab	2 years ago												
matlab_cr	last year												
C new-works	1 hour ago												
Untitled.ip	14 days ago												
Untitled1.i	14 days ago												
Untitled2.i	14 days ago												
Untitled4.i	9 days ago												
Untitled5.i	9 days ago												
Untitled6.i	1 hour ago												
• 🖪 Untitled7.i	10 seconds ago												

17

Probability review



FOUNDATIONS OF STATISTICAL NATURAL LANGUAGE PROCESSING

CHRISTOPHER D. MANNING AND HINRICH SCHÜTZE

Probability

- Probability of an event *a* occurring
- *P*(*a*)
 - For example, *a* could be a die showing a 2 out of {1, 2, 3, 4, 5, 6}
- Estimate *P*(*a*) as $\frac{\text{count}(a)}{\text{count}(all events)}$
 - Relative frequency or maximum likelihood estimate (MLE)

Probability distributions



Random variables

- **Random variable:** a mapping from a domain of possible outcomes in a sample space to a range of measurable space, such as counts
 - Typically the "result of an experiment"
 - For example, flipping a coin multiple times (possible outcomes {H, T}) and recording the result as 0 for tails and 1 for heads
- Distribution of a random variable X
 - *P*(*X*) is a probability distribution over all possible values in the sample space. Probability mass function
 - P(X = x) is the probability that the random variable X has the value x
 - \circ *P*(*X* = *heads*), where *X* is the random variable of a coin flip



Figure 7.1: $\mathbf{P}(k \text{ heads})$ in 30 tosses, success prob 1/3.

Joint probability

- Probability of 2 events both occurring $P(A \cap B)$ P(A,B)
- When rolling 2 dice, what's the probability of getting two 5s?

Let D_1 be dice 1, D_2 be dice 2. These events are independent, so:

$$P(D_1 = 5, D_2 = 5) = P(D_1 = 5) \cdot P(D_2 = 5)$$

 $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$ since there are 36 different possible combinations

Conditional probability

- Probability distributions sometimes change if you know another event has occurred or not occurred
- Conditional probability of an event *a* occurring given that another event, *b*, has already occurred
 - \circ P(a|b)
- Assume
 - X is the outcome of rolling a die once
 - F is the event X = 6
 - \circ *E* is the event *X* > 4
- Die is rolled and we are told that *E* has occurred
- What is P(F|E)?

Conditional probability



Figure 4.1: Events on the dart board

• Assume a very bad dart thrower (maybe Michael)

$$\mathbf{P}(A) = \frac{\mathbf{area}(A)}{\mathbf{area}(\Omega)}$$

Conditional probability



Figure 4.1: Events on the dart board

- You don't see the throw, but somebody tells you that the dart landed in *B* (so *B* occurred)
- What is the formula for P(A|B)?

Linear algebra review



An array of numbers with D dimensions

[23]

Can be represented as a point in *D*-dimensional space



Dot product: vector · vector

Sum of the products of each vector dimension





A matrix is an array of numbers



Two rows, three columns.

It's Easy to Multiple a Matrix by a Scalar



Dot product: vector · matrix



Let a_1 and a_2 be the row vectors of matrix A and b_1 and b_2 be the column vectors of a matrix B. Find C = AB

$$\begin{bmatrix} 1 & 7 \\ 2 & 4 \end{bmatrix} \cdot \begin{bmatrix} 3 & 3 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} a_1 \cdot b_1 & a_1 \cdot b_2 \\ a_2 \cdot b_1 & a_2 \cdot b_2 \end{bmatrix} = \begin{bmatrix} 38 & 17 \\ 26 & 14 \end{bmatrix}$$

A must have the same number of rows as B has columns.

Questions?

No class next Mon for MLK Day. Will see you again on Wed. Take a look at HW1