

Problem 1 (Markov chains):

a) Use w_t to rank the teams by sorting in decreasing value according to this vector. List the top 25 team names (see accompanying file) and their corresponding values in w_t for $t = 10; 100; 1000; 10000$.

The top 25 $t=10,100,1000,10000$ are as shown below:

Top 25 teams for t = 10			Top 25 teams for t = 1000		
	Name	Value in w_t		Name	Value in w_t
0	NorthCentral	0.013785	0	LSU	0.084615
1	LSU	0.012375	1	Clemson	0.045633
2	Wheaton	0.012000	2	OhioState	0.043553
3	Morningside	0.010047	3	Georgia	0.032796
4	NorthDakotaSt	0.009649	4	Alabama	0.024736
5	UW-Whitewater	0.009641	5	Florida	0.023752
6	MountUnion	0.009357	6	Oregon	0.019311
7	MinnSt-Mankato	0.009314	7	Auburn	0.019081
8	OhioState	0.009227	8	Oklahoma	0.018747
9	Muhlenberg	0.009114	9	PennState	0.018016
10	WestFlorida	0.009005	10	Wisconsin	0.015201
11	Clemson	0.008778	11	NotreDame	0.014682
12	StJohnsMN	0.008640	12	Minnesota	0.014234
13	MaryHardin-Baylor	0.008331	13	Baylor	0.012734
14	FerrisSt	0.007726	14	Michigan	0.012073
15	MarianIN	0.007341	15	Utah	0.011730
16	Georgia	0.006735	16	Memphis	0.011691
17	GrandView	0.006681	17	Iowa	0.011032
18	SlipperyRock	0.006028	18	AppalachianSt	0.010949
19	ValdostaSt	0.006000	19	Texas	0.010211
20	Oregon	0.005919	20	NorthDakotaSt	0.009249
21	Wartburg	0.005709	21	Navy	0.009093
22	DelawareValley	0.005580	22	TexasA&M	0.008954
23	UnionNY	0.005528	23	BoiseSt	0.008326
24	Lenoir-Rhyne	0.005454	24	KansasSt	0.007352

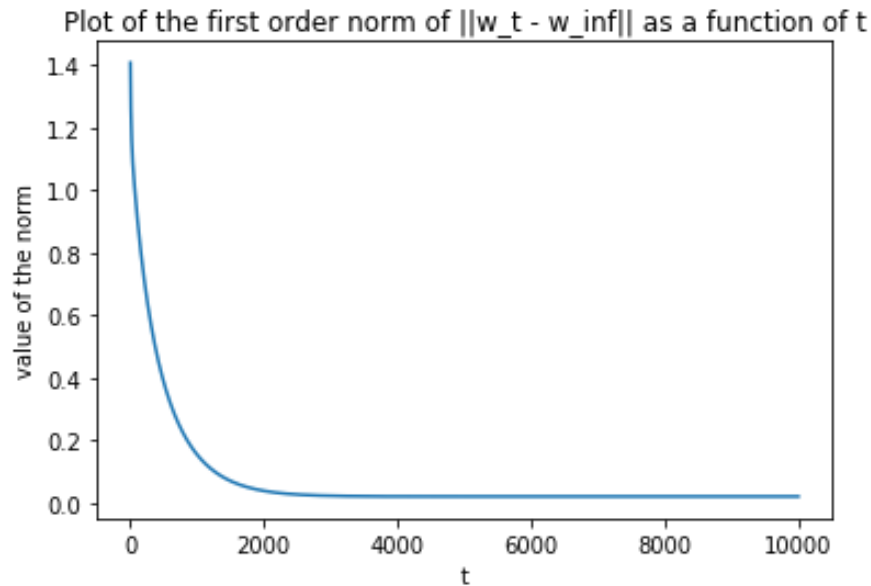
Top 25 teams for t = 100			Top 25 teams for t = 10000		
	Name	Value in w_t		Name	Value in w_t
0	LSU	0.043520	0	LSU	0.091674
1	Wheaton	0.026537	1	Clemson	0.049428
2	NorthCentral	0.025048	2	OhioState	0.047168
3	Clemson	0.023639	3	Georgia	0.035501
4	OhioState	0.022626	4	Alabama	0.026785
5	Morningside	0.021715	5	Florida	0.025711
6	Georgia	0.017115	6	Oregon	0.020849
7	StJohnsMN	0.015441	7	Auburn	0.020652
8	WestFlorida	0.013548	8	Oklahoma	0.020270
9	UW-Whitewater	0.013204	9	PennState	0.019494
10	Alabama	0.012882	10	Wisconsin	0.016447
11	Florida	0.012404	11	NotreDame	0.015874
12	Muhlenberg	0.011386	12	Minnesota	0.015399
13	Oregon	0.010500	13	Baylor	0.013763
14	Auburn	0.009992	14	Michigan	0.013067
15	Oklahoma	0.009982	15	Utah	0.012659
16	MinnSt-Mankato	0.009875	16	Memphis	0.012632
17	PennState	0.009508	17	Iowa	0.011934
18	MarianIN	0.009144	18	AppalachianSt	0.011832
19	ValdostaSt	0.009002	19	Texas	0.011040
20	Lenoir-Rhyne	0.008150	20	Navy	0.009818
21	Wisconsin	0.007996	21	NorthDakotaSt	0.009749
22	NotreDame	0.007843	22	TexasA&M	0.009690
23	MountUnion	0.007602	23	BoiseSt	0.008974
24	Minnesota	0.007510	24	KansasSt	0.007942

b) We saw that w_1 is related to the first eigenvector of M_T . That is, we can find w_1 by getting the first eigenvector and eigenvalue of M_T and post-processing

This is because $u_1^T * u_1 = 1$ by convention. Also, we observe that $\lambda_1 = 1$ for this specific matrix.

Plot $\|w_t - w_{\text{inf}}\|$ (first order norm) as a function of t for $t = 1; \dots; 10000$.

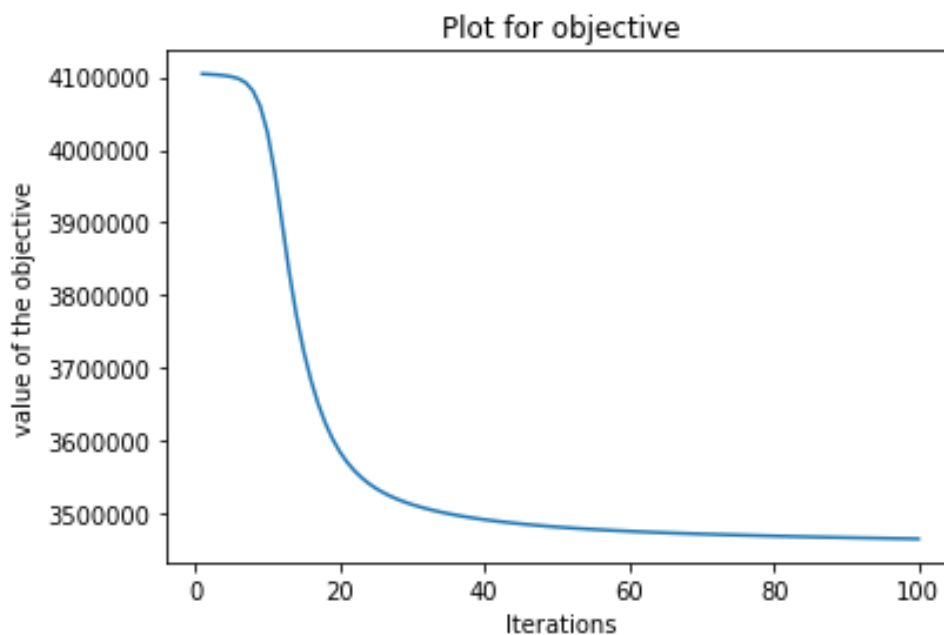
The plot for the first order norm of difference of w_t and w_{infinity} is as shown below:



Problem 2 (Nonnegative matrix factorization):

a) Implement and run the NMF algorithm on this data using the divergence penalty. Set the rank to 25 and run for 100 iterations. This corresponds to learning 25 topics. Plot the objective as a function of iteration.

The objective plotted as a function of the iterations is as shown below:



b) After running the algorithm, normalize the columns of W so they sum to one. For each column of W , list the 10 words having the largest weight and show the weight. The i th row of W corresponds to the i th word in the “dictionary” provided with the data. Organize these lists in a 5 x 5 table.

1 ----- sell: 100.42 sale: 83.96 company: 75.81 store: 73.8 buy: 71.32 price: 53.74 business: 49.84 cost: 44.42 industry: 44.2 customer: 43.21 -----	2 ----- school: 138.31 student: 116.27 child: 54.82 program: 53.1 class: 52.97 education: 48.93 college: 48.6 group: 41.97 teacher: 39.86 community: 39.48 -----	3 ----- police: 133.86 kill: 82.39 officer: 78.95 man: 68.22 charge: 65.89 crime: 58.33 arrest: 56.51 death: 53.22 official: 50.63 victim: 43.75 -----	4 ----- father: 168.49 son: 140.5 mrs: 132.3 mother: 116.82 daughter: 109.4 graduate: 86.54 family: 78.77 marry: 74.42 receive: 62.38 retire: 62.06 -----	5 ----- cause: 51.33 problem: 43.61 damage: 40.02 plant: 32.74 water: 32.35 official: 31.1 report: 29.46 scientist: 25.6 safety: 24.67 system: 24.2 -----
6 ----- official: 79.64 military: 61.02 american: 59.95 states: 55.48 war: 47.56 force: 44.26 meeting: 40.9 government: 39.59 nations: 35.12 agreement: 34.84 -----	7 ----- pay: 83.68 money: 79.29 state: 71.15 budget: 65.68 bill: 65.65 tax: 55.84 program: 52.42 cut: 50.56 plan: 48.91 cost: 47.21 -----	8 ----- food: 70.85 fresh: 38.82 serve: 37.25 pound: 33.54 eat: 33.15 restaurant: 31.11 taste: 30.28 dry: 29.74 cook: 28.58 fish: 28.24 -----	9 ----- music: 98.86 play: 84.17 performance: 63.97 perform: 43.39 stage: 42.94 production: 42.57 theater: 42.27 dance: 42.16 song: 40.64 audience: 40.57 -----	10 ----- team: 144.6 player: 96.86 season: 94.9 game: 92.25 play: 74.78 coach: 61.93 baseball: 47.94 league: 44.9 contract: 36.55 football: 35.85 -----
11 ----- building: 95.75 build: 94.39 project: 84.73 city: 81.13 area: 55.93 plan: 52.32 space: 47.24 construction: 46.1 property: 41.4 community: 41.22 -----	12 ----- percent: 133.06 price: 82.2 market: 82.05 rate: 76.16 rise: 69.36 fall: 49.39 stock: 46.62 economy: 45.37 increase: 43.4 low: 42.73 -----	13 ----- health: 61.7 drug: 56.96 doctor: 56.91 medical: 48.93 patient: 46.09 treatment: 42.2 study: 40.05 information: 39.11 care: 34.51 hospital: 32.41 -----	14 ----- write: 94.48 life: 92.65 book: 80.27 man: 64.17 story: 57.77 woman: 55.45 editor: 52.17 child: 43.16 writer: 42.45 love: 41.91 -----	15 ----- woman: 66.93 black: 63.01 man: 60.09 live: 58.95 wear: 56.14 white: 53.05 door: 51.26 street: 47.16 house: 37.9 dress: 36.66 -----
16 ----- art: 87.03 artist: 53.89 history: 39.55 century: 38.36 image: 34.16 american: 30.58 culture: 30.36 world: 29.91 exhibition: 29.77 photograph: 29.37 -----	17 ----- design: 48.8 color: 39.82 wall: 37.81 display: 35.93 room: 28.11 small: 25.86 glass: 25.55 light: 25.24 red: 25.2 open: 24.65 -----	18 ----- country: 97.67 government: 81.19 political: 79.29 american: 55.61 policy: 53.68 states: 51.91 nation: 51.45 leader: 50.81 power: 49.76 economic: 43.5 -----	19 ----- television: 82.47 network: 31.32 video: 30.73 medium: 29.95 computer: 28.42 program: 27.47 movie: 27.33 screen: 26.83 create: 26.05 different: 25.14 -----	20 ----- mile: 65.86 travel: 50.31 hour: 45.26 town: 45.06 car: 44.43 city: 42.22 trip: 40.18 road: 37.23 train: 36.48 drive: 35.81 -----
21 ----- company: 194.15 executive: 104.83 business: 76.22 president: 71.84 chief: 64.58 share: 53.75 chairman: 49.77 yesterday: 43.07 announce: 42.86 financial: 42.44 -----	22 ----- thing: 101.93 feel: 76.18 ask: 74.74 lot: 72.33 tell: 65.15 really: 63.1 little: 53.1 happen: 51.74 put: 49.9 keep: 47.02 -----	23 ----- case: 95.76 court: 83.91 lawyer: 82.83 law: 74.9 judge: 59.39 legal: 51.73 issue: 46.66 rule: 43.11 charge: 41.33 state: 40.81 -----	24 ----- win: 111.75 second: 96.51 play: 76.74 game: 76.04 victory: 70.46 score: 63.41 third: 61.07 point: 57.18 lose: 48.33 final: 46.58 -----	25 ----- campaign: 99.97 vote: 80.72 political: 68.0 election: 67.26 candidate: 66.51 party: 64.36 republican: 61.84 democratic: 48.42 voter: 42.98 support: 42.82 -----