ELEN E4903: MACHINE LEARNING HOMEWORK 5

Problem 1:

a) After processing all the games to obtain matrix M according to the steps given in the question (using the formula and normalizing), setting w_0 to the uniform distribution, and using $w_{t+1} = w_t M$, the following are the rankings of the top 25 and their corresponding values in w_t for t = 10, 100, 1000, 10000.

For t = 10

Rank	Team	wt
1	MountUnion	0.018356191
2	MaryHardin-Baylor	0.012997553
3	StFrancisIN	0.011063820
4	UW-Oshkosh	0.009959080
5	TAMU-Commerce	0.009295249
6	BrockportSt	0.009168531
7	NorthDakotaSt	0.009008083
8	Alabama	0.008993839
9	Morningside	0.008795425
10	Georgia	0.008380292
11	JamesMadison	0.008075856
12	DelawareValley	0.007939080
13	Wartburg	0.007578923
14	IndianaPA	0.007222248
15	Clemson	0.007081443
16	MinnSt-Mankato	0.006954965
17	CentralFlorida	0.006753229
18	OhioState	0.006671468
19	Reinhardt	0.006639562
20	Wisconsin	0.006636598
21	StThomas	0.006016499
22	Oklahoma	0.005948072
23	FrostburgSt	0.005506446
24	Ashland	0.005367881
25	Assumption	0.005324518

ELEN E4903: MACHINE LEARNING HOMEWORK 5

For t = 100

Rank	Team	wt
1	MountUnion	0.065695673
2	Alabama	0.022969467
3	Georgia	0.019242680
4	UW-Oshkosh	0.017903447
5	OhioState	0.014861803
6	Clemson	0.014722849
7	Oklahoma	0.014208006
8	Wisconsin	0.013759670
9	CentralFlorida	0.013054591
10	MaryHardin-Baylor	0.012603659
11	TAMU-Commerce	0.012508443
12	Auburn	0.011962336
13	StFrancisIN	0.011092095
14	PennState	0.010704577
15	BrockportSt	0.009477647
16	NotreDame	0.008181064
17	DelawareValley	0.007975728
18	NorthDakotaSt	0.007956025
19	FrostburgSt	0.007860020
20	TCU	0.007819963
21	MinnSt-Mankato	0.007649788
22	CaseWesternReserve	0.007155505
23	JohnCarroll	0.006854621
24	MiamiFL	0.006548244
25	Harding	0.006454558

ELEN E4903: MACHINE LEARNING HOMEWORK 5

For t = 1000

Rank	Team	wt
1	Alabama	0.042291052
2	Georgia	0.034930371
3	MountUnion	0.033996686
4	OhioState	0.027736262
5	Clemson	0.026899454
6	Oklahoma	0.026182774
7	Wisconsin	0.025691547
8	CentralFlorida	0.023850533
9	Auburn	0.021937496
10	PennState	0.019866365
11	NotreDame	0.014941293
12	TCU	0.014249405
13	MiamiFL	0.011905362
14	MichiganSt	0.011774669
15	OklahomaSt	0.011291157
16	Iowa	0.010675694
17	Northwestern	0.010336568
18	IowaSt	0.010250483
19	LSU	0.009575158
20	NorthDakotaSt	0.009493707
21	SouthernCal	0.009480329
22	Washington	0.009286620
23	UW-Oshkosh	0.009229803
24	Stanford	0.008744015
25	MississippiSt	0.008543101

ELEN E4903: MACHINE LEARNING HOMEWORK 5

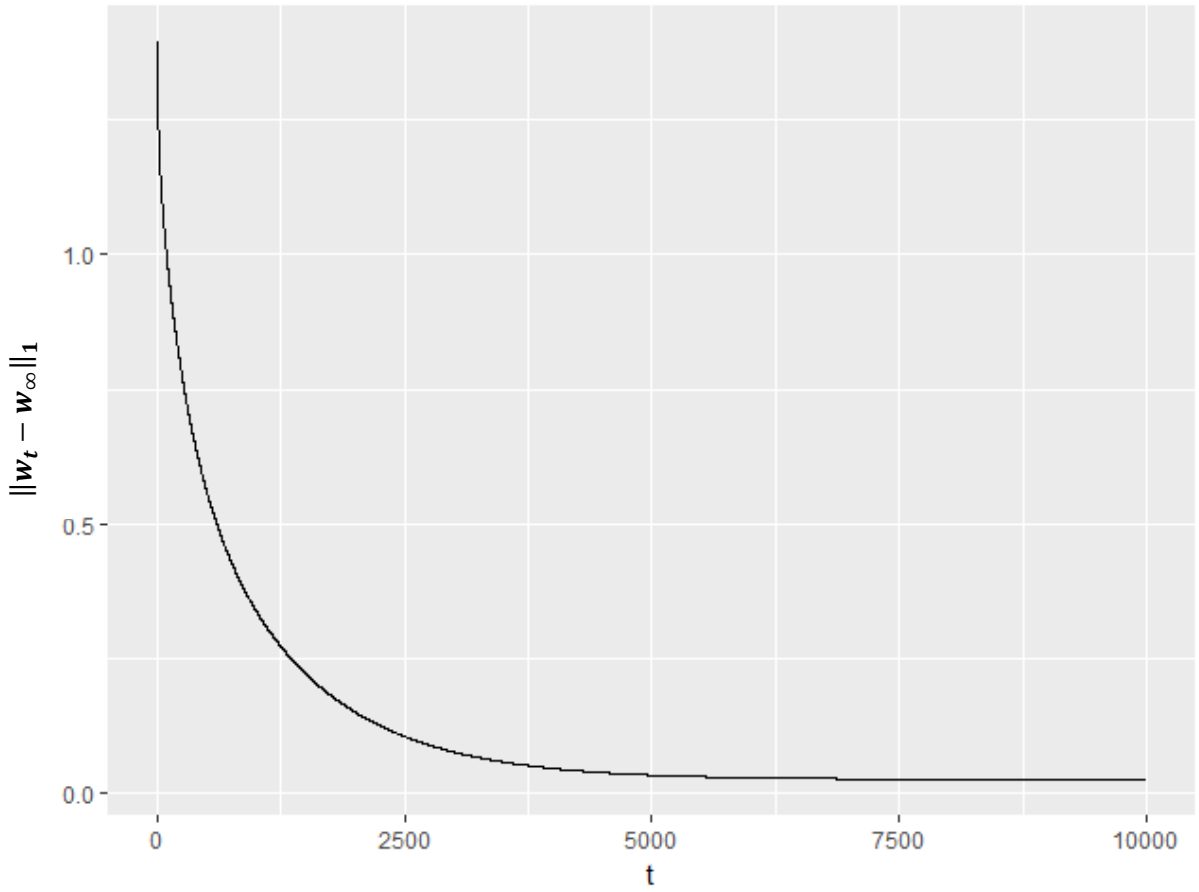
For t = 10000

Rank	Team	wt
1	Alabama	0.051116308
2	Georgia	0.042178218
3	OhioState	0.033565072
4	Clemson	0.032495655
5	Oklahoma	0.031655648
6	Wisconsin	0.031091265
7	CentralFlorida	0.028811193
8	Auburn	0.026507135
9	PennState	0.024030519
10	NotreDame	0.018050380
11	TCU	0.017212718
12	MiamiFL	0.014377678
13	MichiganSt	0.014240452
14	OklahomaSt	0.013636615
15	Iowa	0.012913296
16	Northwestern	0.012505383
17	IowaSt	0.012370671
18	LSU	0.011565219
19	SouthernCal	0.011439345
20	Washington	0.011191508
21	NorthDakotaSt	0.010939734
22	Stanford	0.010549292
23	MississippiSt	0.010317744
24	VirginiaTech	0.010060487
25	NorthCarolinaSt	0.009595815

ELEN E4903: MACHINE LEARNING HOMEWORK 5

b) The plot of $\|w_t - w_\infty\|_1$ as a function of t for this problem is shown below. Intuitively, as t increases, w_t should converge to w_∞ and $\|w_t - w_\infty\|_1$ approach 0 as the stationary distribution is reached. This is exactly what is depicted in the figure below.

Graph of $\|w_t - w_\infty\|_1$ against t

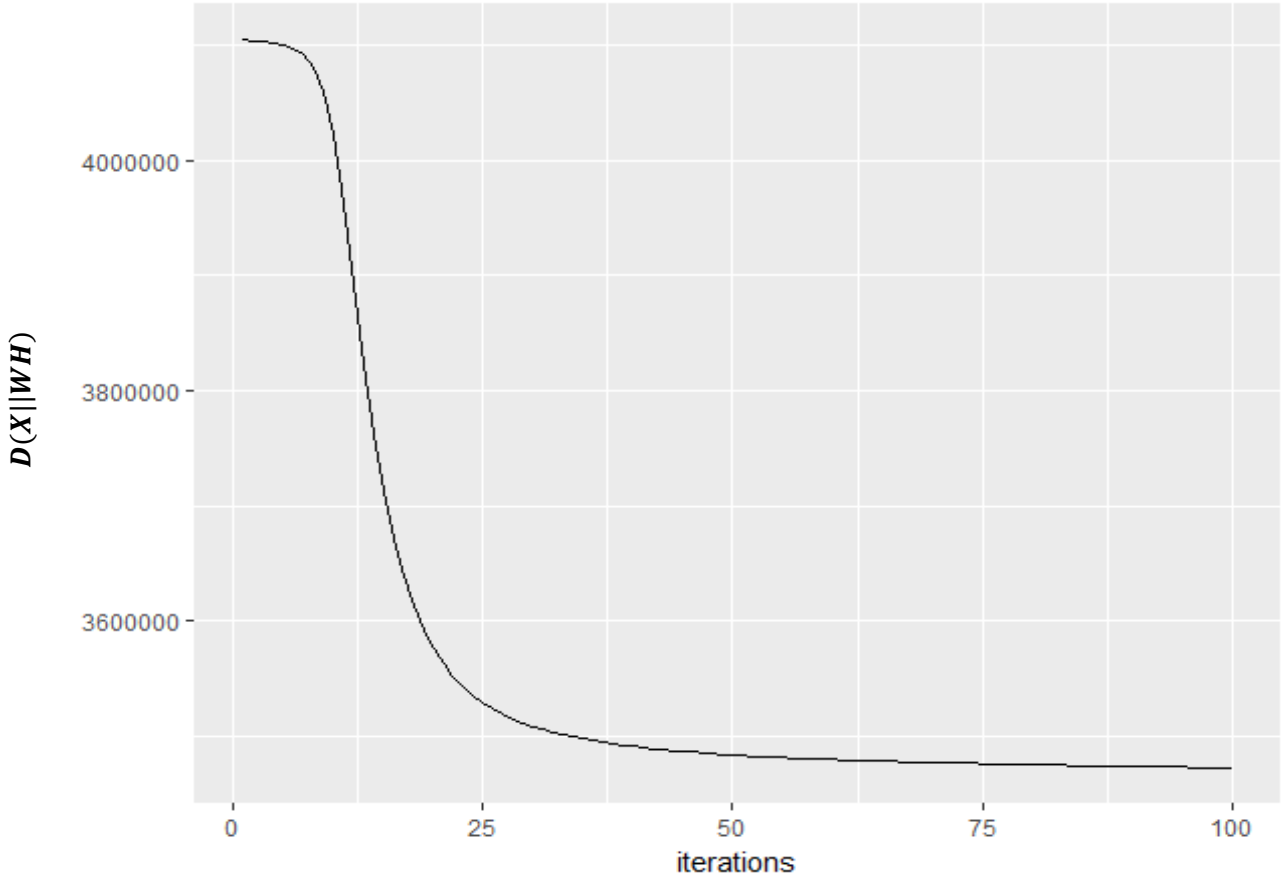


ELEN E4903: MACHINE LEARNING HOMEWORK 5

Problem 2:

a) Implementing the NMF algorithm which uses the divergence penalty gives the following objective function: $D(X||WH) = -\sum_i \sum_j [X_{ij} \ln(WH)_{ij} - (WH)_{ij}]$. To run the algorithm, H and W have to be randomly initialized with a non-negative value (from a Uniform (1,2) distribution), and the following equations have to be iterated, first for all values in H and then all in W; for 100 iterations as specified in the question: $H_{kj} \leftarrow H_{kj} \frac{\sum_i W_{ik} X_{ij}/(WH)_{ij}}{\sum_i W_{ik}}$ and $W_{ik} \leftarrow W_{ik} \frac{\sum_j H_{kj} X_{ij}/(WH)_{ij}}{\sum_j H_{kj}}$. The plot below shows the objective function as a function of iterations. As expected, the objective function, which we aim to minimize, decreases as the number of iteration increases and eventually converges (when change in the objective $D(X||WH)$ is small, we can stop the iterations).

Graph of objective function $D(X||WH)$ against iterations



ELEN E4903: MACHINE LEARNING HOMEWORK 5

b) The following are the 10 words with the largest weight and their corresponding weights for the 25 topics (since rank K = 25) learnt, organized in a 5x5 table:

	Words in topic 1	Weights 1	Words in topic 2	Weights 2	Words in topic 3	Weights 3	Words in topic 4	Weights 4	Words in topic 5	Weights 5
1	build	0.016840319	school	0.05585892	court	0.025429055	official	0.03355452	agreement	0.014089115
2	project	0.015342995	student	0.03899211	case	0.024054741	police	0.03302390	deal	0.013751006
3	site	0.014479102	father	0.02685911	lawyer	0.023524534	yesterday	0.02447986	agree	0.011107851
4	water	0.014393265	graduate	0.02540188	law	0.022123989	charge	0.02127591	plan	0.010543300
5	plant	0.012239304	mrs	0.02352848	judge	0.017589850	officer	0.02042564	decision	0.009183307
6	area	0.011929313	son	0.01878661	legal	0.014747226	report	0.01607384	issue	0.008407713
7	foot	0.010425646	president	0.01715032	rule	0.012603646	arrest	0.01420116	union	0.008315346
8	plan	0.010054201	daughter	0.01674853	trial	0.012016167	investigation	0.01392274	negotiation	0.007637541
9	land	0.009782466	teacher	0.01630876	state	0.011653410	spokesman	0.01293298	reach	0.007329925
10	construction	0.008885682	parent	0.01581189	file	0.009863885	drug	0.01193824	negotiate	0.007316092

	Words in topic 6	Weights 6	Words in topic 7	Weights 7	Words in topic 8	Weights 8	Words in topic 9	Weights 9	Words in topic 10	Weights 10
1	letter	0.013787244	book	0.014801782	editor	0.013783657	man	0.03226861	thing	0.017098041
2	write	0.010555356	art	0.012016001	study	0.012918980	woman	0.02865272	lot	0.013294652
3	article	0.010485598	artist	0.007929940	doctor	0.011146390	life	0.02299058	guy	0.011214295
4	tell	0.009857261	write	0.007612046	health	0.010792851	family	0.02124740	play	0.011133668
5	ask	0.009614290	history	0.007557323	patient	0.009583302	child	0.01770497	feel	0.011111392
6	question	0.009524839	collection	0.007172140	research	0.008686151	death	0.01766003	really	0.010539320
7	public	0.009019390	world	0.006420973	medical	0.008548427	die	0.01456538	job	0.009781890
8	report	0.008956212	century	0.005903412	drug	0.008488382	young	0.01337033	tell	0.009069591
9	news	0.008494818	museum	0.005836762	cause	0.008365834	mother	0.01133763	ask	0.008895797
10	newspaper	0.008397360	view	0.005665445	treatment	0.007799642	friend	0.01098834	team	0.008240713

	Words in topic 11	Weights 11	Words in topic 12	Weights 12	Words in topic 13	Weights 13	Words in topic 14	Weights 14	Words in topic 15	Weights 15
1	company	0.024967744	food	0.016664234	music	0.020103685	percent	0.016063208	owner	0.02148656
2	percent	0.019320791	serve	0.008848411	play	0.017864495	number	0.015529071	property	0.01876871
3	market	0.017315986	restaurant	0.008328310	performance	0.011028178	change	0.009357321	house	0.01598681
4	price	0.014066373	pound	0.007900921	song	0.008619708	increase	0.009069760	buy	0.01510483
5	stock	0.012386414	taste	0.007542390	sound	0.008598347	result	0.008296631	home	0.01460580
6	share	0.011338923	fresh	0.007537350	stage	0.008511771	level	0.007992549	housing	0.01427547
7	business	0.010185954	add	0.007095029	dance	0.008284913	large	0.007952995	building	0.01369427
8	sell	0.009420743	white	0.006745965	audience	0.007725037	far	0.007488290	estate	0.01361290
9	sale	0.009231808	red	0.006700368	perform	0.007717975	grow	0.007074588	real	0.01292198
10	executive	0.009116811	eat	0.006622674	production	0.007144750	recent	0.006234119	family	0.01205831

	Words in topic 16	Weights 16	Words in topic 17	Weights 17	Words in topic 18	Weights 18	Words in topic 19	Weights 19	Words in topic 20	Weights 20
1	campaign	0.01970956	war	0.025147347	company	0.020866650	city	0.03518033	game	0.022532692
2	vote	0.01922132	military	0.020909731	computer	0.017032283	street	0.02419121	win	0.018317646
3	political	0.01890106	force	0.016661719	sell	0.010846568	resident	0.01633453	play	0.017825451
4	candidate	0.01394093	attack	0.016086205	technology	0.010414596	neighborhood	0.01507618	team	0.017514043
5	election	0.01383144	american	0.012718934	customer	0.010090898	town	0.01417507	second	0.015890477
6	republican	0.01362891	weapon	0.011809019	store	0.010047203	open	0.01318056	season	0.015480651
7	party	0.01317009	troop	0.011124288	service	0.009906300	building	0.01197025	player	0.012599564
8	democratic	0.01158692	kill	0.010623965	product	0.009719829	house	0.01134970	victory	0.011315598
9	leader	0.01014683	soldier	0.009853705	system	0.008613377	local	0.01100306	score	0.010504761
10	support	0.00955415	fight	0.008347728	business	0.008593761	door	0.01094256	point	0.009928435

Table continued next page

ELEN E4903: MACHINE LEARNING HOMEWORK 5

	Words in topic 20	Weights 20	Words in topic 21	Weights 21	Words in topic 22	Weights 22	Words in topic 23	Weights 23	Words in topic 24	Weights 24	Words in topic 25	Weights 25
1	game	0.022532692	money	0.01808275	television	0.031610982	thing	0.009444756	car	0.027429746	country	0.023535930
2	win	0.018317646	pay	0.01591546	film	0.019228566	wear	0.008278309	driver	0.013302772	government	0.021151694
3	play	0.017825451	program	0.01557591	movie	0.018366459	little	0.007331130	train	0.012039811	states	0.018679099
4	team	0.017514043	state	0.01514188	event	0.015570471	feel	0.006656007	vehicle	0.011274516	american	0.015349636
5	second	0.015890477	budget	0.01473291	ticket	0.011753180	ask	0.006651497	truck	0.011138566	leader	0.013488208
6	season	0.015480651	tax	0.01319281	program	0.011191864	keep	0.006518082	worker	0.010596844	political	0.012840005
7	player	0.012599564	bill	0.01307772	feature	0.010780158	turn	0.006197909	gas	0.010382818	foreign	0.011104670
8	victory	0.011315598	cut	0.01141422	series	0.009596312	tell	0.006154725	drive	0.009884812	international	0.010461661
9	score	0.010504761	cost	0.01069502	network	0.008721325	put	0.005886362	accident	0.009697296	peace	0.009518011
10	point	0.009928435	government	0.01051896	schedule	0.008598295	friend	0.005360742	company	0.009626525	official	0.008917471