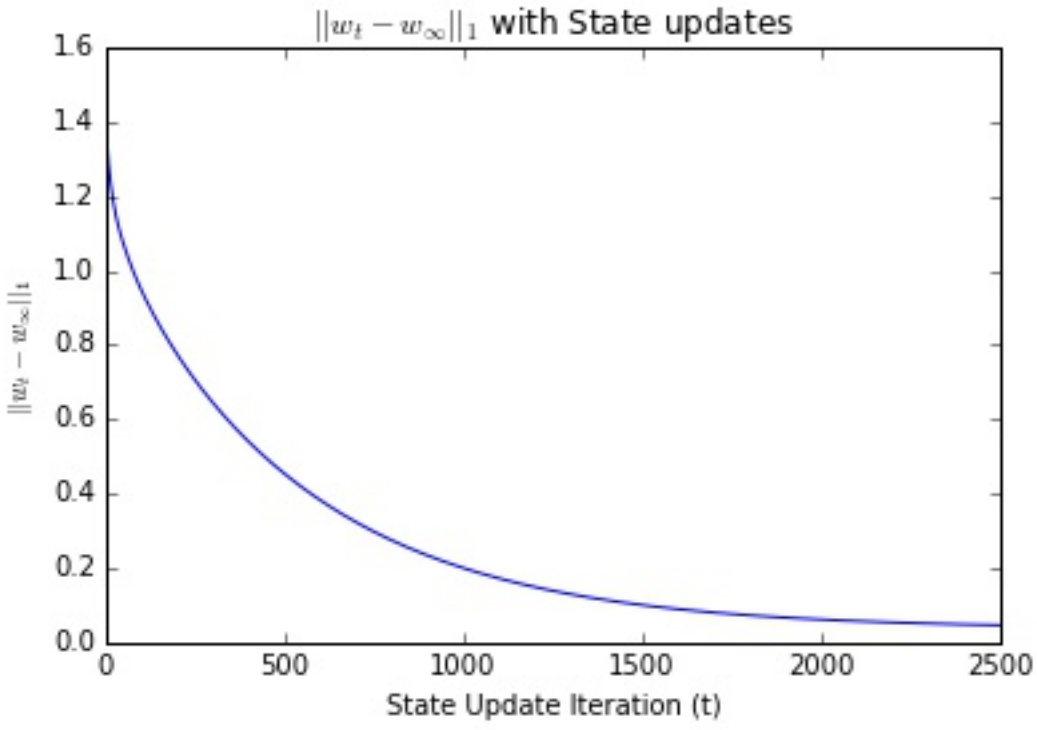
Machine Learning - Assignment #5

Maanit Mehra - mm4694@columbia.edu

May 2nd, 2016

Problem 1: Markov Chains



$$\boxed{||w_{2500} - w_\infty|| = 0.037541}$$

. Rank	T = 10 Weightage	. College
1	0.018045	Mount Union
2	0.013829	St Thomas
3	0.012406	NW Missouri St
4	0.012182	Alabama
5	0.009050	Marian IN
6	0.008820	Linfield
7	0.008256	UW-Whitewater
8	0.008000	Clemson
9	0.007498	Wabash
10	0.006757	UW-Oshkosh
11	0.006756	North Dakota St
12	0.006722	West Georgia
13	0.006596	Morningside
14	0.006502	Shepherd
15	0.006400	St Francis IN
16	0.006353	Wheaton
17	0.006190	Wesley
18	0.006119	Thomas More
19	0.006018	Ohio State
20	0.005925	Mary Hardin-Baylor
21	0.005879	Johns Hopkins
22	0.005565	Slippery Rock
23	0.005474	Stanford
24	0.005446	Amherst
25	0.005344	Grand Valley St

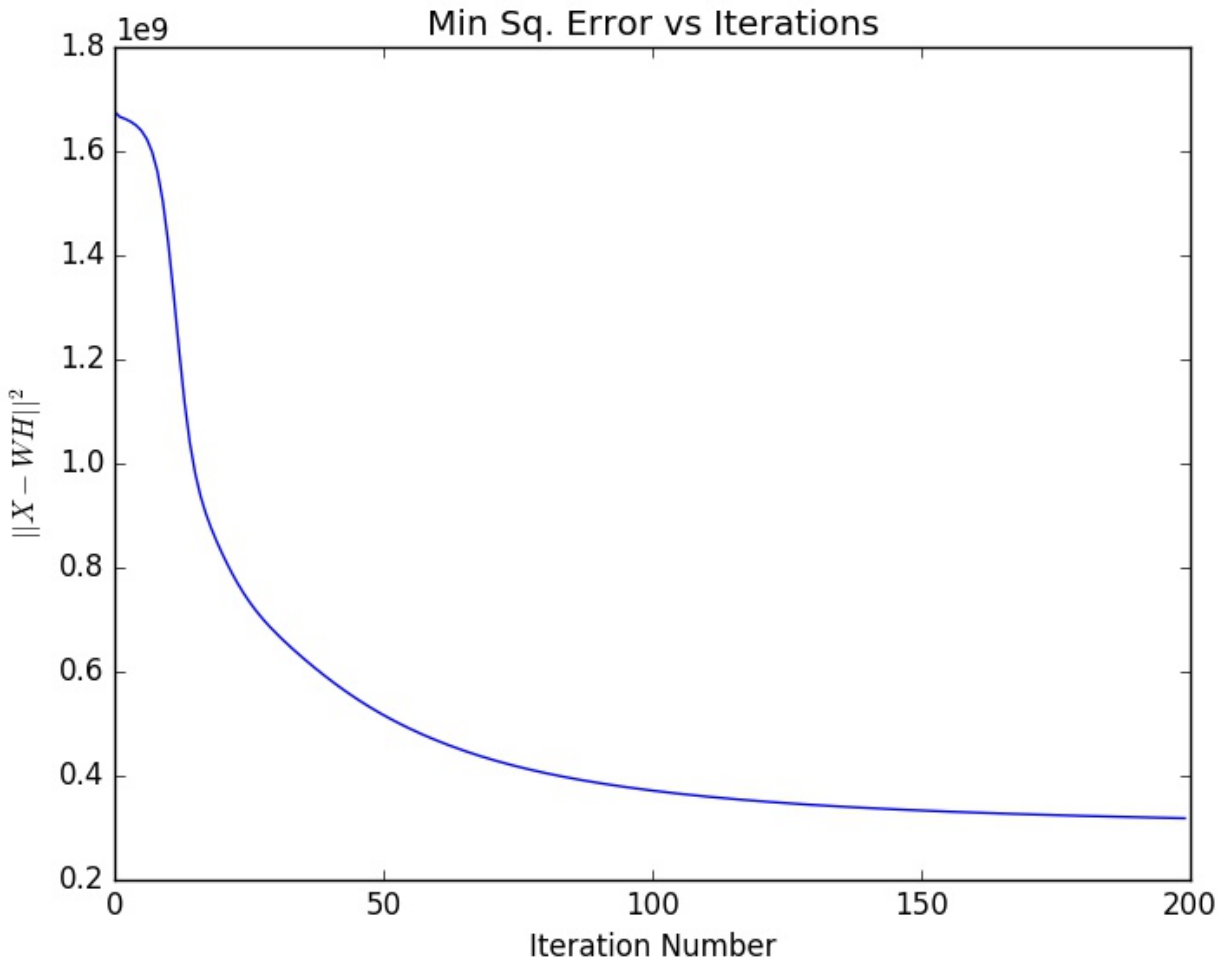
. Rank	T = 100 Weightage	. College
1	0.058296	Mount Union
2	0.032117	Alabama
3	0.030921	NW Missouri St
4	0.024293	St Thomas
5	0.016924	Clemson
6	0.013411	Mississippi
7	0.012833	Ohio State
8	0.011267	Oklahoma
9	0.011165	Marian IN
10	0.010525	Stanford
11	0.010306	Michigan St
12	0.010047	Houston
13	0.008691	St Francis IN
14	0.008561	Linfield
15	0.008313	Michigan
16	0.007969	Iowa
17	0.007964	TCU
18	0.007723	Wesley
19	0.007645	Emporia St
20	0.007611	Florida
21	0.007546	LSU
22	0.007282	Arkansas
23	0.007206	Notre Dame
24	0.006883	UW-Whitewater
25	0.006849	Florida St

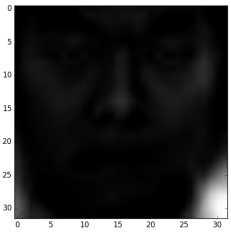
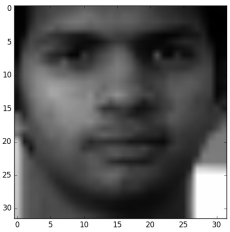
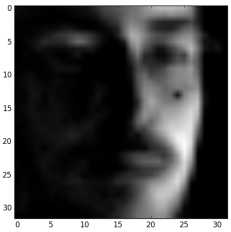
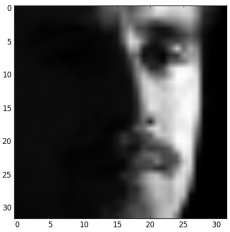
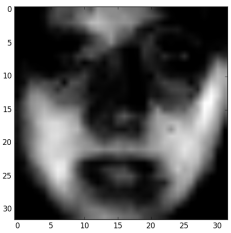
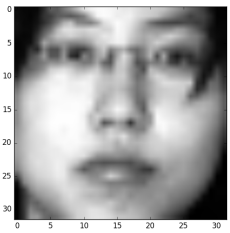
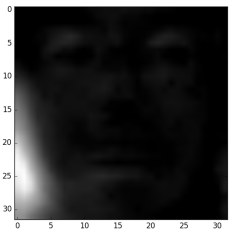
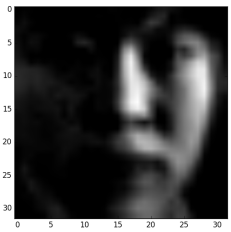
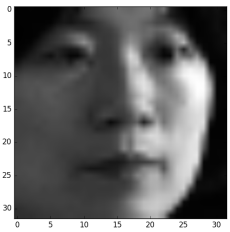
. Rank	T = 1000 Weightage	. College
1	0.060949	Alabama
2	0.031893	Clemson
3	0.025395	Mississippi
4	0.024252	Ohio State
5	0.020878	Oklahoma
6	0.019605	Stanford
7	0.019349	Michigan St
8	0.019005	Houston
9	0.015638	Michigan
10	0.014618	TCU
11	0.014613	Iowa
12	0.014440	Florida
13	0.014259	LSU
14	0.013727	Arkansas
15	0.013527	Notre Dame
16	0.012918	NW Missouri St
17	0.012813	Florida St
18	0.012628	Tennessee
19	0.012445	Mount Union
20	0.011187	Baylor
21	0.011006	Utah
22	0.010757	Navy
23	0.010739	Oklahoma St
24	0.010680	Northwestern
25	0.010112	North Carolina

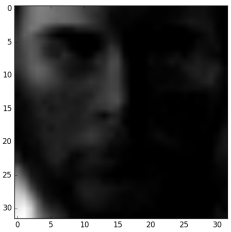
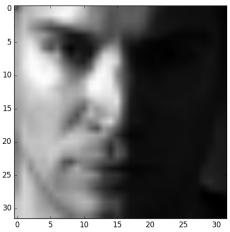
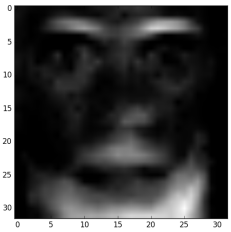
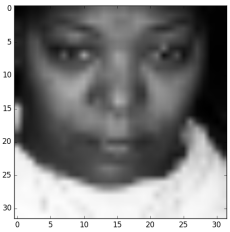
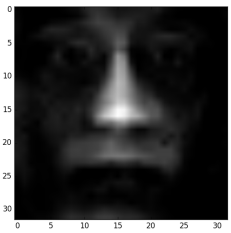
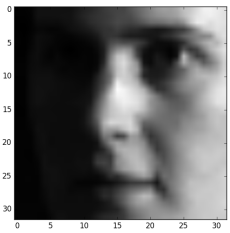
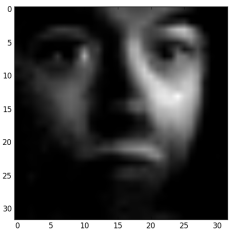
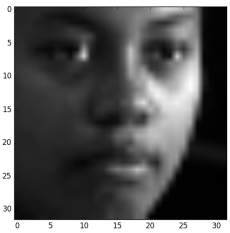
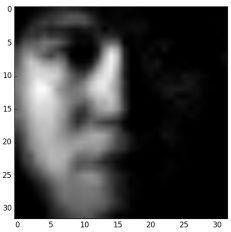
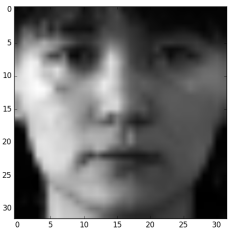
. Rank	T = 2500 Weightage	. College
1	0.066742	Alabama
2	0.034911	Clemson
3	0.027803	Mississippi
4	0.026556	Ohio State
5	0.022822	Oklahoma
6	0.021446	Stanford
7	0.021177	Michigan St
8	0.020812	Houston
9	0.017118	Michigan
10	0.015968	Iowa
11	0.015968	TCU
12	0.015812	Florida
13	0.015609	LSU
14	0.015024	Arkansas
15	0.014806	Notre Dame
16	0.014019	Florida St
17	0.013820	Tennessee
18	0.012215	Baylor
19	0.012036	Utah
20	0.011767	Navy
21	0.011726	Oklahoma St
22	0.011682	Northwestern
23	0.011056	North Carolina
24	0.010742	Wisconsin
25	0.010660	Oregon

Problem 2: Non-negative Matrix Factorization

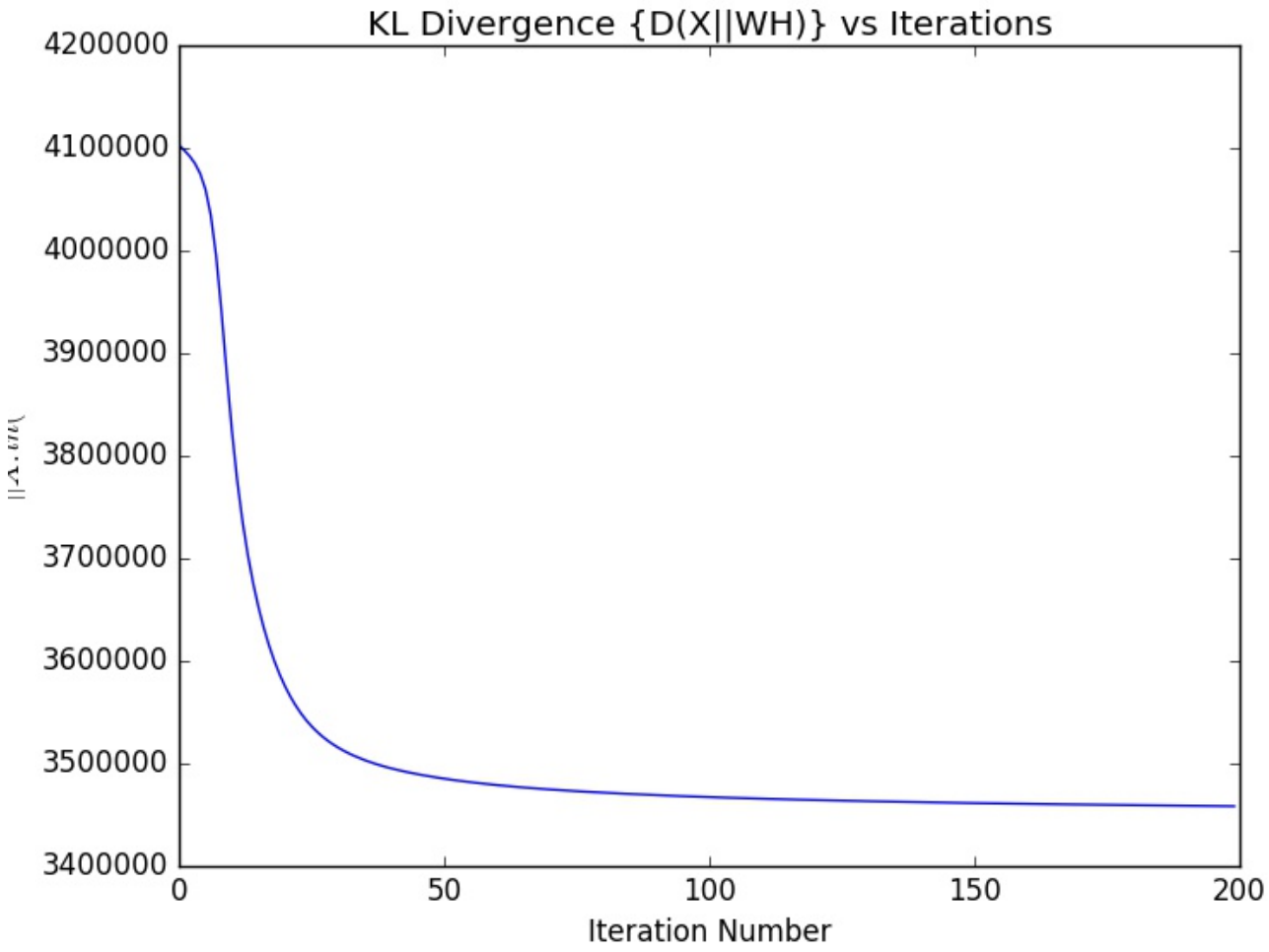
`Part 1 (Faces)`



W	Image(w)	x	Image(x)
0		210	
3		159	
5		264	
7		999	
12		359	

W	Image(w)	x	Image(x)
14		304	
17		109	
18		140	
20		995	
22		960	

Part 2: New York Times Data



Rank	Topic Number 2 Weight	Word
0	0.009584	room
1	0.009316	wear
2	0.008249	thing
3	0.007663	sit
4	0.006973	ask
5	0.006603	stand
6	0.006306	night
7	0.006236	tell
8	0.006079	walk
9	0.006012	little

Rank	Topic Number 15 Weight	Word
0	0.013677	play
1	0.013008	film
2	0.012010	character
3	0.011844	love
4	0.011243	movie
5	0.009741	man
6	0.009690	story
7	0.009158	life
8	0.008204	television
9	0.007817	star

Rank	Topic Number 11 Weight	Word
0	0.015860	art
1	0.010275	artist
2	0.009816	design
3	0.009122	collection
4	0.008520	museum
5	0.008516	image
6	0.007756	photograph
7	0.007378	exhibition
8	0.006998	century
9	0.006562	color

Rank	Topic Number 7 Weight	Word
0	0.027581	music
1	0.014055	performance
2	0.013461	play
3	0.012608	perform
4	0.011638	dance
5	0.011507	song
6	0.008932	sing
7	0.008689	concert
8	0.008565	band
9	0.008429	program

Rank	Topic Number 6 Weight	Word
0	0.035541	team
1	0.025418	player
2	0.022444	season
3	0.021871	game
4	0.019741	play
5	0.016593	coach
6	0.010451	sport
7	0.010277	baseball
8	0.010070	league
9	0.008910	football

Rank	Topic Number 9 Weight	Word
0	0.073029	father
1	0.060898	son
2	0.057345	mrs
3	0.047420	daughter
4	0.036435	graduate
5	0.032287	marry
6	0.032021	mother
7	0.024576	retire
8	0.024477	receive
9	0.021552	degree

Rank	Topic Number 17 Weight	Word
0	0.042855	company
1	0.022085	executive
2	0.018717	business
3	0.014747	president
4	0.012518	chief
5	0.011910	sell
6	0.011853	share
7	0.011074	sale
8	0.009768	buy
9	0.009504	chairman

Rank	Topic Number 18 Weight	Word
0	0.022727	charge
1	0.017343	official
2	0.013847	investigation
3	0.011361	case
4	0.011239	report
5	0.010904	office
6	0.009564	yesterday
7	0.008188	information
8	0.008149	prosecutor
9	0.007817	investigate

Rank	Topic Number 3 Weight	Word
0	0.036990	family
1	0.036012	child
2	0.021059	life
3	0.019865	woman
4	0.017409	live
5	0.015995	home
6	0.015120	friend
7	0.013385	wife
8	0.013359	young
9	0.012339	old

Rank	Topic Number 16 Weight	Word
0	0.030290	book
1	0.026576	write
2	0.026272	editor
3	0.016210	article
4	0.013828	writer
5	0.011994	read
6	0.011957	page
7	0.011704	publish
8	0.011010	author
9	0.008503	newspaper