COMS W4721: Machine Learning for Data Science

Homework 5

jc4609

1.

a). ‘w10’ represent the values of state vector at the 10th iteration. So on and so forth for ‘w100’,’w1000’,’w10000’.

‘w\_final’ represent the state vector calculated by using the eigenvalue.

‘team’ below each state vector represent the team name corresponding to the state vector.

w10: [[ 0.01475209 0.01315112 0.01241204 0.01102911 0.00992525 0.00817769

0.00816794 0.00809767 0.00775471 0.00742784 0.0069544 0.00691117

0.00683891 0.00672165 0.00654897 0.00636897 0.00629096 0.00619634

0.00613562 0.00576391 0.00556592 0.00553723 0.00546441 0.00545738

0.0054416 ]]

team: ['Mary Hardin-Baylor' 'NW Missouri St' 'Alabama' 'Clemson' 'UW-Oshkosh'

'St Thomas' 'UW-Whitewater' 'St Francis IN' 'Mount Union' 'James Madison'

'Reinhardt' 'John Carroll' 'Shepherd' 'California PA' 'North Dakota St'

'Baker' 'Alfred' 'Grand Valley St' 'North Alabama' 'Harding' 'Marian IN'

'North Central' 'Trinity CT' 'Wheaton' 'Morningside']

w100: [[ 0.03825784 0.03336442 0.03298808 0.03088284 0.01684449 0.01215973

0.01044446 0.01026125 0.01024589 0.00994073 0.00942623 0.00933187

0.00889067 0.0087886 0.00871304 0.0081987 0.00807673 0.00804134

0.00800668 0.00756942 0.00728989 0.0069545 0.00690421 0.00654385

0.00644855]]

team: ['NW Missouri St' 'Alabama' 'Clemson' 'Mary Hardin-Baylor' 'UW-Oshkosh'

'UW-Whitewater' 'Florida St' 'Grand Valley St' 'John Carroll' 'Ohio State'

'Ferris St' 'St Thomas' 'Washington' 'Virginia Tech' 'Mount Union'

'Michigan' 'Pittsburgh' 'Oklahoma' 'Emporia St' 'Penn State' 'Wisconsin'

'Louisville' 'Miami FL' 'North Carolina' 'Southern Cal']

w1000: [[ 0.06071896 0.06040796 0.03049481 0.01881362 0.01798335 0.01588941

0.01582415 0.01482737 0.01465114 0.01439026 0.01367182 0.01308813

0.01261349 0.01242035 0.01158945 0.01141893 0.01129224 0.01125102

0.01081326 0.01081239 0.00997832 0.0095472 0.00952065 0.00951113

0.00868321]]

team: ['Alabama' 'Clemson' 'NW Missouri St' 'Florida St' 'Ohio State'

'Washington' 'Virginia Tech' 'Michigan' 'Pittsburgh' 'Oklahoma'

'Penn State' 'Wisconsin' 'Louisville' 'Miami FL' 'Southern Cal'

'North Carolina' 'Oklahoma St' 'Florida' 'Tennessee' 'Auburn'

'Georgia Tech' 'LSU' 'Western Michigan' 'Stanford' 'James Madison']

w10000: [[ 0.06728211 0.06698283 0.02083524 0.01992155 0.0175791 0.01752392

0.01642561 0.01634207 0.01623476 0.01592243 0.01514424 0.01448896

0.0139769 0.01375278 0.01283024 0.01261239 0.01248801 0.01245617

0.01197327 0.01197299 0.01105016 0.01054168 0.01053654 0.01052444

0.00952351]]

team: ['Alabama' 'Clemson' 'Florida St' 'Ohio State' 'Washington' 'Virginia Tech'

'Michigan' 'NW Missouri St' 'Pittsburgh' 'Oklahoma' 'Penn State'

'Wisconsin' 'Louisville' 'Miami FL' 'Southern Cal' 'North Carolina'

'Oklahoma St' 'Florida' 'Tennessee' 'Auburn' 'Georgia Tech' 'LSU'

'Western Michigan' 'Stanford' 'Texas A&M']

w\_final: [ 0.0681792 0.06787593 0.02111304 0.02018717 0.01781349 0.01775757

0.01664462 0.01655996 0.01645122 0.01613473 0.01534617 0.01468215

0.01416326 0.01393616 0.01300131 0.01278055 0.01265452 0.01262225

0.01213291 0.01213263 0.0111975 0.01068223 0.01067703 0.01066477

0.00965049]

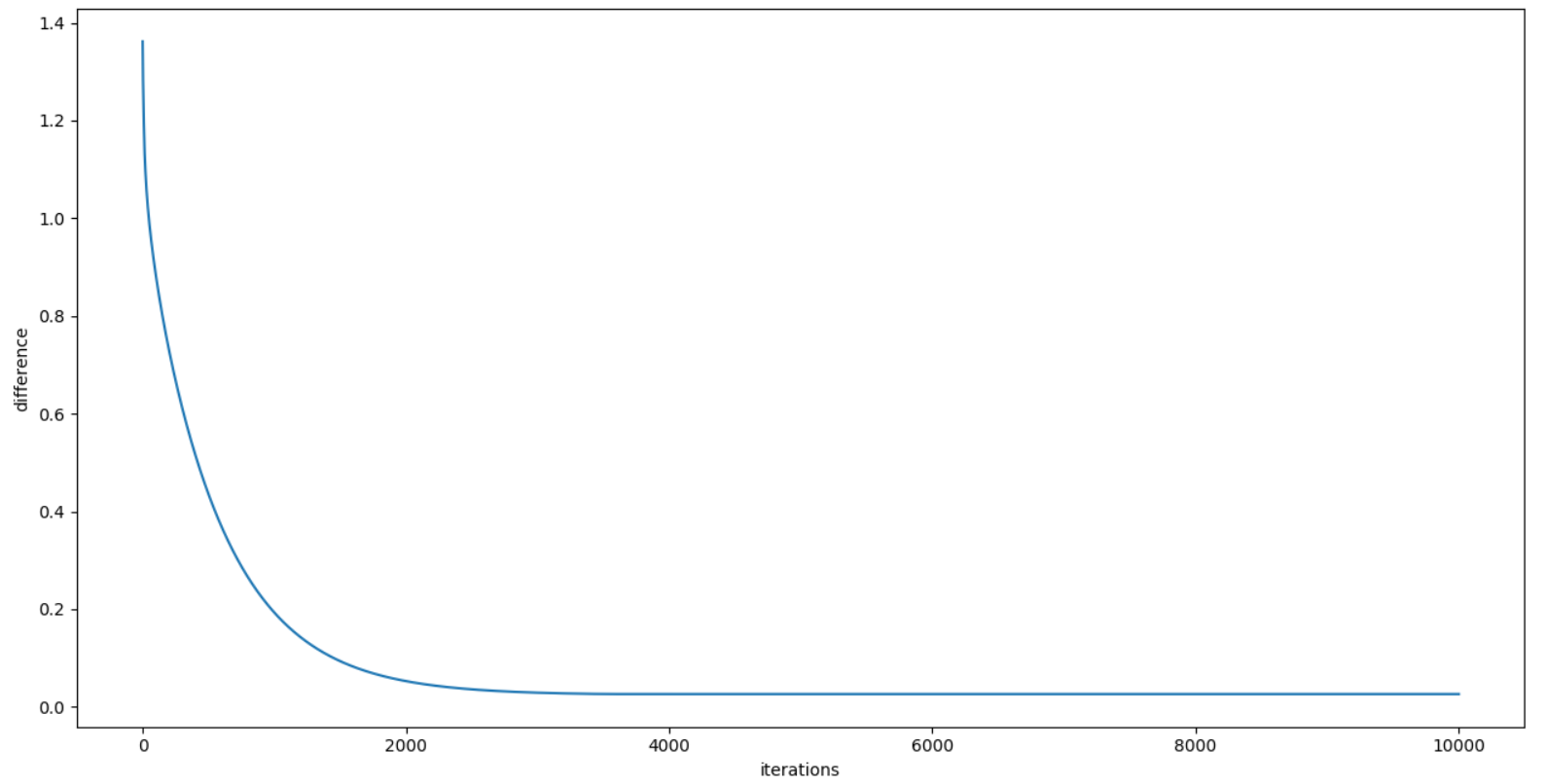
team: ['Alabama' 'Clemson' 'Florida St' 'Ohio State' 'Washington' 'Virginia Tech'

'Michigan' 'NW Missouri St' 'Pittsburgh' 'Oklahoma' 'Penn State'

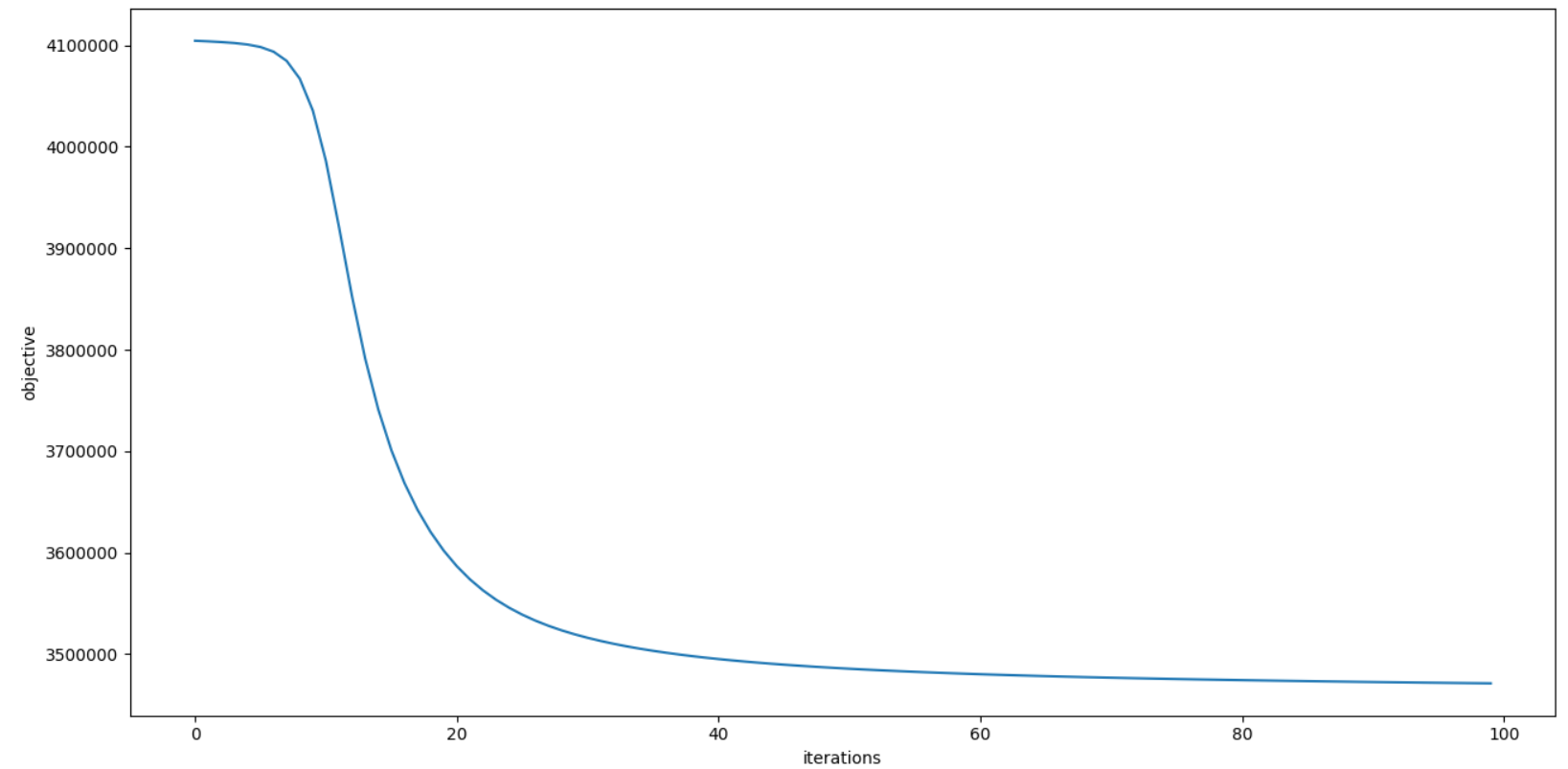
'Wisconsin' 'Louisville' 'Miami FL' 'Southern Cal' 'North Carolina'

'Oklahoma St' 'Florida' 'Tennessee' 'Auburn' 'Georgia Tech' 'LSU'

'Western Michigan' 'Stanford' 'Texas A&M']

b).

2.

a). 

b).

25 topics:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ['food' 'water' 'restaurant' 'serve' 'fresh' 'dry' 'taste' 'eat' 'pound'  'small'] | ['food' 'water' 'restaurant' 'serve' 'fresh' 'dry' 'taste' 'eat' 'pound'  'small'] | ['drug' 'case' 'trial' 'doctor' 'patient' 'health' 'medical' 'treatment'  'evidence' 'lawyer'] | ['city' 'building' 'area' 'build' 'house' 'resident' 'town' 'street' 'home'  'live'] | ['official' 'states' 'american' 'military' 'plan' 'government' 'war'  'weapon' 'force' 'effort'] |
| ['team' 'play' 'game' 'win' 'player' 'season' 'second' 'coach' 'victory'  'point'] | ['art' 'design' 'artist' 'collection' 'image' 'color' 'photograph' 'museum'  'paint' 'form'] | ['book' 'write' 'editor' 'article' 'writer' 'page' 'publish' 'author'  'read' 'life'] | ['school' 'student' 'father' 'child' 'son' 'mother' 'mrs' 'graduate'  'daughter' 'parent'] | ['event' 'states' 'world' 'information' 'open' 'ticket' 'american' 'tour'  'start' 'free'] |
| ['keep' 'problem' 'far' 'bad' 'line' 'put' 'point' 'turn' 'hard' 'thing'] | ['program' 'money' 'state' 'cost' 'pay' 'budget' 'plan' 'provide' 'cut'  'increase'] | ['company' 'computer' 'sell' 'store' 'product' 'sale' 'customer' 'buy'  'technology' 'industry'] | ['music' 'performance' 'play' 'dance' 'song' 'perform' 'sing' 'sound'  'band' 'concert'] | ['country' 'political' 'american' 'government' 'states' 'leader' 'nation'  'foreign' 'group' 'americans'] |
| ['television' 'film' 'movie' 'character' 'play' 'audience' 'star' 'story'  'director' 'direct'] | ['hit' 'game' 'baseball' 'season' 'play' 'pitch' 'third' 'second' 'team'  'start'] | ['percent' 'report' 'increase' 'number' 'accord' 'rate' 'result' 'study'  'figure' 'decline'] | ['law' 'court' 'rule' 'pay' 'lawyer' 'case' 'union' 'contract' 'legal'  'file'] | ['vote' 'campaign' 'political' 'election' 'republican' 'candidate' 'party'  'state' 'leader' 'democratic'] |
| ['police' 'kill' 'officer' 'man' 'arrest' 'attack' 'force' 'official'  'death' 'soldier'] | ['company' 'executive' 'president' 'business' 'share' 'chief' 'chairman'  'firm' 'financial' 'vice'] | ['price' 'market' 'percent' 'rise' 'economy' 'rate' 'stock' 'fall' 'trade'  'economic'] | ['wear' 'woman' 'morning' 'night' 'man' 'hour' 'car' 'room' 'sit' 'front'] | ['official' 'office' 'comment' 'yesterday' 'ask' 'member' 'report'  'investigation' 'public' 'tell'] |

The vector of most common words in 25 topics and their weight:

[ 0.01611121 0.01012514 0.00906367 0.00863727 0.00794746 0.00794469

0.00772574 0.00743134 0.00721507 0.00698475]

['food' 'water' 'restaurant' 'serve' 'fresh' 'dry' 'taste' 'eat' 'pound'

'small']

[ 0.01564152 0.0123503 0.01165059 0.01142983 0.01041116 0.00898607

0.00861843 0.00795659 0.00772418 0.00719333]

['thing' 'tell' 'life' 'feel' 'ask' 'really' 'lot' 'old' 'friend' 'man']

[ 0.01823632 0.01754119 0.01349211 0.01205523 0.01050871 0.00991052

0.00988223 0.00964325 0.00829472 0.00808897]

['drug' 'case' 'trial' 'doctor' 'patient' 'health' 'medical' 'treatment'

'evidence' 'lawyer']

[ 0.02162285 0.0211242 0.01286046 0.01254512 0.01200488 0.01198264

0.00974087 0.00961521 0.00917747 0.00896647]

['city' 'building' 'area' 'build' 'house' 'resident' 'town' 'street' 'home'

'live']

[ 0.01541006 0.01220171 0.00981678 0.00917083 0.00878802 0.00833375

0.00674531 0.00661924 0.00632314 0.00585578]

['official' 'states' 'american' 'military' 'plan' 'government' 'war'

'weapon' 'force' 'effort']

[ 0.02508528 0.02252449 0.02153747 0.01880173 0.01747192 0.01524304

0.01343998 0.01195734 0.01171887 0.01123884]

['team' 'play' 'game' 'win' 'player' 'season' 'second' 'coach' 'victory'

'point']

[ 0.0141977 0.01061208 0.00984212 0.00935662 0.00916019 0.0081257

0.00787436 0.00727795 0.00720053 0.00658333]

['art' 'design' 'artist' 'collection' 'image' 'color' 'photograph' 'museum'

'paint' 'form']

[ 0.02363274 0.02171424 0.0213997 0.01361882 0.01081438 0.00978703

0.00965015 0.00963701 0.00950108 0.00868633]

['book' 'write' 'editor' 'article' 'writer' 'page' 'publish' 'author'

'read' 'life']

[ 0.04418923 0.03650321 0.03431373 0.03415217 0.0281505 0.02791707

0.02439328 0.0225976 0.02228094 0.02078499]

['school' 'student' 'father' 'child' 'son' 'mother' 'mrs' 'graduate'

'daughter' 'parent']

[ 0.03029569 0.01225788 0.01153041 0.01115057 0.01082854 0.01061057

0.00980824 0.00962742 0.00907891 0.0090572 ]

['event' 'states' 'world' 'information' 'open' 'ticket' 'american' 'tour'

'start' 'free']

[ 0.0105877 0.010079 0.00938405 0.00855389 0.00805988 0.0075832

0.00746961 0.00744967 0.00738965 0.00729666]

['keep' 'problem' 'far' 'bad' 'line' 'put' 'point' 'turn' 'hard' 'thing']

[ 0.01682945 0.01355916 0.01253777 0.01242028 0.01178232 0.00982428

0.00922794 0.00858133 0.00847229 0.00807779]

['program' 'money' 'state' 'cost' 'pay' 'budget' 'plan' 'provide' 'cut'

'increase']

[ 0.02024599 0.01596052 0.01584048 0.01461999 0.01160536 0.01145576

0.01057338 0.00969335 0.00950232 0.00841077]

['company' 'computer' 'sell' 'store' 'product' 'sale' 'customer' 'buy'

'technology' 'industry']

[ 0.02897855 0.01430728 0.01381177 0.01224297 0.01210572 0.01192192

0.00974496 0.00958773 0.00915304 0.00914105]

['music' 'performance' 'play' 'dance' 'song' 'perform' 'sing' 'sound'

'band' 'concert']

[ 0.01791672 0.01327704 0.01267055 0.01088387 0.01068914 0.01015078

0.00730657 0.00705686 0.00643427 0.00639371]

['country' 'political' 'american' 'government' 'states' 'leader' 'nation'

'foreign' 'group' 'americans']

[ 0.02803678 0.02135847 0.01803368 0.01548127 0.01490686 0.01057113

0.01042051 0.01039374 0.00999966 0.00915225]

['television' 'film' 'movie' 'character' 'play' 'audience' 'star' 'story'

'director' 'direct']

[ 0.03646655 0.02455579 0.01626845 0.01543662 0.01194354 0.01170186

0.01143213 0.01115914 0.01114143 0.01078914]

['hit' 'game' 'baseball' 'season' 'play' 'pitch' 'third' 'second' 'team'

'start']

[ 0.0496214 0.0375324 0.0202279 0.01922036 0.01849954 0.01508376

0.01405591 0.01145256 0.01128226 0.01101025]

['percent' 'report' 'increase' 'number' 'accord' 'rate' 'result' 'study'

'figure' 'decline']

[ 0.02131729 0.02018173 0.016219 0.01516371 0.01503053 0.01415446

0.01397963 0.01191749 0.01166824 0.01154177]

['law' 'court' 'rule' 'pay' 'lawyer' 'case' 'union' 'contract' 'legal'

'file']

[ 0.01930383 0.0189281 0.01462443 0.01406949 0.01363037 0.0129398

0.01216801 0.01159704 0.01121698 0.01057665]

['vote' 'campaign' 'political' 'election' 'republican' 'candidate' 'party'

'state' 'leader' 'democratic']

[ 0.02782849 0.01864969 0.01462347 0.01445305 0.01141666 0.01117586

0.01045234 0.00870195 0.00845995 0.00815738]

['police' 'kill' 'officer' 'man' 'arrest' 'attack' 'force' 'official'

'death' 'soldier']

[ 0.04204682 0.02345548 0.01732024 0.01670929 0.01353541 0.0130969

0.01127248 0.01112022 0.01069268 0.01033958]

['company' 'executive' 'president' 'business' 'share' 'chief' 'chairman'

'firm' 'financial' 'vice']

[ 0.02486727 0.02320489 0.01862431 0.01366159 0.01358498 0.01279303

0.01078869 0.01061422 0.01036816 0.00907194]

['price' 'market' 'percent' 'rise' 'economy' 'rate' 'stock' 'fall' 'trade'

'economic']

[ 0.01156774 0.01014317 0.00944118 0.0088008 0.00842612 0.00786581

0.00781802 0.00762565 0.00716788 0.00664719]

['wear' 'woman' 'morning' 'night' 'man' 'hour' 'car' 'room' 'sit' 'front']

[ 0.01966843 0.01125185 0.01046068 0.01037042 0.00977778 0.00915346

0.00897201 0.00889934 0.00862229 0.00839001]

['official' 'office' 'comment' 'yesterday' 'ask' 'member' 'report'

'investigation' 'public' 'tell']