

### Exercise 6.3 - Distance Computation with Word Embeddings

For each of the following tasks, please calculate the similarity / distance scores for Law2Vec [5] and compare them with the scores of GloVe [4]:

A) Compute the top 5 most similar words to each of the 20 selected words in

Table 1 of the Law2Vec paper [3] by using the cosine similarity

➤ (article, ..., motion)

Solution:-

**Table 1** Top-5 similar words for a set of 20 selected words based on cosine similarity between the associated word embeddings

article	convention, section, articles, clause, provisions
act	statute, provision, mccarranferguson, irca, tupa
action	suit, actions, lawsuit, claim, proceeding
crime	offense, murder, crimes, felony, violent
felony	offense, misdemeanor, felonies, offenses, convicted
punishment	penalty, punishments, sentencing, sentence, imprisonment
security	social, health, administration, retirement
fraud	fraudulent, theft, deceit, misrepresentation, bribery
privacy	confidentiality, communications, liberty, freedom, freedoms
intellectual	copyrights, patents, copyright, trademark, wipo
terrorism	terrorist, trafficking, counter-terrorism, violent, laundering
immigrant	immigrants, nonquota, alien, asylum, citizenship
illegal	unlawful, corrupt, improper, illicit, fraudulent
drugs	drug, narcotic, addicts, psychotropic, medicines
appeal	appeals, review, hearing, appellate, appealed
abuse	violence, sexual, self-destructive, assault, mistreatment
alcohol	liquor, spirits, intoxicating, beer, vinous
complaint	grievance, allegations, allegation, complaints, counterclaim
indictment	conviction, summary, imprisonment, indictable, triable
motion	motions, petition, dismiss, leave, cross-motion

For GloVe embeddings:-

```
{'abuse': [('sexual', 0.7119625806808472),  
          ('abuses', 0.6575741767883301),  
          ('allegations', 0.656885027885437),  
          ('abused', 0.644381046295166),  
          ('sex', 0.6409667730331421)],  
'act': [('acts', 0.7794305086135864),
```

('law', 0.6375415921211243),  
('laws', 0.6352546215057373),  
('legislation', 0.6244657635688782),  
('enacted', 0.6225273609161377)],  
'action': [('actions', 0.7189486026763916),  
('response', 0.6410176753997803),  
('any', 0.6229310035705566),  
('move', 0.6129212379455566),  
('measures', 0.6109905242919922)],  
'alcohol': [('drinking', 0.6793702840805054),  
('drink', 0.6587716341018677),  
('alcoholic', 0.6550531983375549),  
('liquor', 0.6263242959976196),  
('addiction', 0.616674542427063)],  
'appeal': [('appeals', 0.8120466470718384),  
('appealed', 0.7374923229217529),  
('appealing', 0.7363486886024475),  
('court', 0.6937041282653809),  
('decision', 0.6561456918716431)],  
'article': [('articles', 0.7287136316299438),  
('published', 0.6664239168167114),  
('page', 0.6382156610488892),  
('publication', 0.6227012872695923),  
('magazine', 0.6194932460784912)],  
'complaint': [('filed', 0.7600173950195312),  
('alleging', 0.7556746602058411),  
('lawsuit', 0.7532499432563782),  
('complaints', 0.7340501546859741),  
('alleges', 0.6849454641342163)],  
'crime': [('criminal', 0.7245837450027466),  
('crimes', 0.6970728635787964),

('murder', 0.6458415985107422),  
('trafficking', 0.6441540718078613),  
('murders', 0.6189703345298767)],  
'drugs': [('drug', 0.8692162036895752),  
('medications', 0.7425538897514343),  
('medicines', 0.702671229839325),  
('medication', 0.697799801826477),  
('prescription', 0.6924887895584106)],  
'felony': [('misdemeanor', 0.8802327513694763),  
('felonies', 0.771780788898468),  
('counts', 0.716042160987854),  
('larceny', 0.6874004602432251),  
('offenses', 0.6866480708122253)],  
'fraud': [('fraudulent', 0.7209619283676147),  
('embezzlement', 0.6988704800605774),  
('charges', 0.6957302093505859),  
('irregularities', 0.6874820590019226),  
('evasion', 0.6713991761207581)],  
'illegal': [('illegally', 0.7032275199890137),  
('smuggling', 0.7011212110519409),  
('trafficking', 0.6657689213752747),  
('immigrants', 0.6572180390357971),  
('illicit', 0.6378949880599976)],  
'immigrant': [('immigrants', 0.8401675224304199),  
('migrant', 0.6586581468582153),  
('immigration', 0.6520180702209473),  
('undocumented', 0.6261692643165588),  
('migrants', 0.6165194511413574)],  
'indictment': [('indictments', 0.7474280595779419),  
('indicted', 0.6797527074813843),  
('charges', 0.6659150123596191),

('prosecutors', 0.6433315873146057),  
('prosecution', 0.6348966360092163)],  
'intellectual': [('literary', 0.5628653168678284),  
('cultural', 0.561545729637146),  
('academic', 0.5525931119918823),  
('scholarly', 0.5495522022247314),  
('creativity', 0.5457181930541992)],  
'motion': [('motions', 0.7237242460250854),  
('action', 0.5294706225395203),  
('reconsideration', 0.4987879991531372),  
('no-confidence', 0.49767208099365234),  
('censure', 0.49187707901000977)],  
'privacy': [('protections', 0.6106308698654175),  
('confidentiality', 0.5989964008331299),  
('liberties', 0.5912916660308838),  
('protection', 0.5544588565826416),  
('protect', 0.5485154390335083)],  
'punishment': [('punishments', 0.7635425925254822),  
('sentence', 0.6709460020065308),  
('imprisonment', 0.6657165288925171),  
('meted', 0.649448573589325),  
('harsher', 0.645781397819519)],  
'security': [('military', 0.6203919649124146),  
('officials', 0.6168738603591919),  
('officers', 0.6108697652816772),  
('intelligence', 0.6077421307563782),  
('police', 0.6074000000953674)],  
'terrorism': [('terror', 0.8303008079528809),  
('terrorist', 0.8002285361289978),  
('terrorists', 0.7479064464569092),  
('extremism', 0.7211341857910156),

('threat', 0.652489960193634)]}]}

Top 5 most similar words obtained from law2vec model

{'abuse': [('violence', 0.6047216057777405),

('sexual', 0.5924835205078125),

('self-destructive', 0.5540093183517456),

('assault', 0.5521867871284485),

('mistreatment', 0.5408604145050049)]],

'act': [('statute', 0.6347061395645142),

('§', 0.6320472955703735),

('provision', 0.6195109486579895),

('mccarranferguson', 0.6141680479049683),

('irca', 0.613204836845398)]],

'action': [('suit', 0.7625703811645508),

('actions', 0.7550638914108276),

('lawsuit', 0.7105170488357544),

('claim', 0.6544851064682007),

('proceeding', 0.6417317986488342)]],

'alcohol': [('liquor', 0.6247892379760742),

('spirits', 0.620631217956543),

('intoxicating', 0.6028286218643188),

('beer', 0.5810186862945557),

('vinous', 0.5808978080749512)]],

'appeal': [('appeals', 0.6785864233970642),

('review', 0.6456307172775269),

('hearing', 0.6360105872154236),

('appellate', 0.625510573387146),

('appealed', 0.6197022199630737)]],

'article': [('convention', 0.6230964660644531),

('section', 0.588966965675354),

('articles', 0.5592589974403381),

('clause', 0.5505915880203247),  
('provisions', 0.5461112260818481)],  
'complaint': [('grievance', 0.6954891085624695),  
('allegations', 0.6646998524665833),  
('allegation', 0.663662314414978),  
('complaints', 0.6476339101791382),  
('counterclaim', 0.6411800384521484)],  
'crime': [('offense', 0.7254292964935303),  
('murder', 0.6850289106369019),  
('crimes', 0.6729871034622192),  
('felony', 0.6589514017105103),  
('violent', 0.645916223526001)],  
'drugs': [('drug', 0.7857030034065247),  
('narcotic', 0.7241787910461426),  
('addicts', 0.6322356462478638),  
('psychotropic', 0.6315125226974487),  
('medicines', 0.6211462020874023)],  
'felony': [('offense', 0.8303117752075195),  
('misdemeanor', 0.799248993396759),  
('felonies', 0.7530264258384705),  
('offenses', 0.7107501029968262),  
('convicted', 0.7094066143035889)],  
'fraud': [('fraudulent', 0.644085168838501),  
('theft', 0.6300138235092163),  
('deceit', 0.6072374582290649),  
('misrepresentation', 0.6061434745788574),  
('bribery', 0.604240894317627)],  
'illegal': [('unlawful', 0.7168081402778625),  
('corrupt', 0.6443915367126465),  
('improper', 0.589661180973053),  
('illicit', 0.5682709813117981),

('fraudulent', 0.5605493783950806)],  
'immigrant': [('immigrants', 0.6497452855110168),  
('nonquota', 0.5691838264465332),  
('alien', 0.5544954538345337),  
('asylum', 0.550568699836731),  
('citizenship', 0.5427181720733643)],  
'indictment': [('conviction', 0.7802726030349731),  
('summary', 0.6821379661560059),  
('imprisonment', 0.6107341051101685),  
('indictable', 0.571308970451355),  
('triable', 0.5636763572692871)],  
'intellectual': [('copyrights', 0.53022301197052),  
('patents', 0.5186152458190918),  
('copyright', 0.5155873894691467),  
('trademark', 0.497243732213974),  
('wipo', 0.4911758601665497)],  
'motion': [('motions', 0.7267795205116272),  
('petition', 0.6432109475135803),  
('dismiss', 0.6236007213592529),  
('leave', 0.6216575503349304),  
('cross-motion', 0.6065188646316528)],  
'privacy': [('confidentiality', 0.6086761355400085),  
('communications', 0.5904741287231445),  
('liberty', 0.5871350765228271),  
('freedom', 0.5565742254257202),  
('freedoms', 0.5516698956489563)],  
'punishment': [('penalty', 0.6692326068878174),  
('punishments', 0.6650903820991516),  
('sentencing', 0.629499614238739),  
('sentence', 0.6238275170326233),  
('imprisonment', 0.6178720593452454)],

```
'security': [('social', 0.6683281064033508),
             ('health', 0.4961587190628052),
             ('administration', 0.4936230182647705),
             ('retirement', 0.49084150791168213),
             ('curity', 0.48942241072654724)],
'terrorism': [('terrorist', 0.7091106176376343),
              ('trafficking', 0.6308254599571228),
              ('counter-terrorism', 0.6254228353500366),
              ('violent', 0.5942960977554321),
              ('laundering', 0.5858644247055054)]}
```

Top 5 similar words are more or less synonyms in GloVe while in Law2Vec, the similar words are related to the semantic concept of the word.

**B) Compute the cosine similarity between the following word pairs:**

➤ (“sentence”, “writing”), (“sentence”, “reading”), (“sentence”, “death”),  
 (“punishment”, “death”)

Solution:-

Cosine similarity between (“sentence”, “writing”) by GloVe and Law2vec

By GloVe: 0.332891

By Law2vec: 0.180392

Cosine similarity between (“sentence”, “reading”) by GloVe and Law2vec

By GloVe: 0.286445

By Law2vec: 0.325889

Cosine similarity between (“sentence”, “death”) by GloVe and Law2vec

By GloVe: 0.549144

By Law2vec: 0.665044

Cosine similarity between (“punishment”, “death”) by GloVe and Law2vec

By GloVe: 0.510231

By Law2vec: 0.507011

Sentence has more a “judgement” or “punishment” connotation in Law2Vec embedding.



Similarly Death and Sentence are more similar in the case of Law2Vec

**C) Compute the word mover's distance between the following document pairs:**

- ("I got capital punishment", "I got sentenced to death")
- ("let me finish my sentence", "I got sentenced to death")
- ("let me finish my sentence", "I am still writing")
- ("let me finish my sentence", "I am still reading")

Word Movers distance between I got capital punishment, I got sentenced to death

By GloVe: 4.882263

By Law2Vec: 2.623782

Word Movers distance between let me finish my sentence, I got sentenced to death

By GloVe: 6.076215

By Law2Vec: 4.705064

Word Movers distance between let me finish my sentence, I am still writing

By GloVe: 6.644206

By Law2Vec: 4.682960

Word Movers distance between let me finish my sentence, I am still reading

By GloVe: 6.784415

By Law2vec: 4.550605

Similar to the cosine similarity the punishment connotation of Sentence is clearly evident in the case of Law2vec.