

▼ Sentiment Classification

Generate Word Embeddings and retrieve outputs of each layer with Keras based on Cl

Word embeddings are a type of word representation that allows words with similar meaning to have a similar

It is a distributed representation for text that is perhaps one of the key breakthroughs for the impressive performance on challenging natural language processing problems.

We will use the imdb dataset to learn word embeddings as we train our dataset. This dataset contains 25,000 reviews labeled with sentiment (positive or negative).

Dataset

```
from keras.datasets import imdb
```

Dataset of 25,000 movie reviews from IMDB, labeled by sentiment (positive/negative). Reviews have been preprocessed and encoded as a sequence of word indexes (integers). For convenience, the words are indexed by their frequency, so that the word with index 1 is the most frequent word. Use the first 20 words from each review to speed up training, using `imdb.get_word_index()`.

As a convention, "0" does not stand for a specific word, but instead is used to encode any unknown word.

Aim

1. Import test and train data
2. Import the labels (train and test)
3. Get the word index and then Create key value pair for word and word_id. (12.5 points)
4. Build a Sequential Model using Keras for Sentiment Classification task. (10 points)
5. Report the Accuracy of the model. (5 points)
6. Retrieve the output of each layer in keras for a given single test sample from the trained model you built. (2.5 points)

▼ Usage:

```
from google.colab import drive
drive.mount('/content/drive/')
```

➞ Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=947318989803-6

```
Enter your authorization code:
.....
Mounted at /content/drive/
```


```
#Importing Libraries
```

```
import matplotlib.pyplot as plt
import tensorflow as tf
import numpy as np
from scipy.spatial.distance import edist
```

```

from scipy.spatial.distance import euclidean
from keras.models import Sequential
from keras.layers import Dense, GRU, Embedding, LSTM
from keras.optimizers import Adam
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences

```

 The default version of TensorFlow in Colab will soon switch to TensorFlow 2.x.
 We recommend you [upgrade](#) now or ensure your notebook will continue to use TensorFlow 1.x via the %tensorflow_version command.
 Using TensorFlow backend.


```

from keras.datasets import imdb

vocab_size = 10000 #vocab size

# vocab_size is no.of words to consider from the dataset, ordering based on frequency.
(x_train, y_train), (x_test, y_test) = imdb.load_data(num_words=vocab_size)

```

 Downloading data from <https://s3.amazonaws.com/text-datasets/imdb.npz>
 17465344/17464789 [=====] - 3s 0us/step

```

from keras.preprocessing.sequence import pad_sequences
vocab_size = 10000 #vocab size
maxlen = 300 #number of word used from each review

```

```

#load dataset as a list of ints
(x_train, y_train), (x_test, y_test) = imdb.load_data(num_words=vocab_size)
#make all sequences of the same length
x_train = pad_sequences(x_train, maxlen=maxlen)
x_test = pad_sequences(x_test, maxlen=maxlen)

```

```

# Make Word to ID dictionary
INDEX_FROM=3 # word index offset
word_to_id = imdb.get_word_index() #Get the word index
word_to_id = {k:(v+INDEX_FROM) for k,v in word_to_id.items()}


word_to_id["<PAD>"] = 0

word_to_id["<START>"] = 1

word_to_id["<UNK>"] = 2

# Make ID to Word dictionary
id_to_word = {value:key for key,value in word_to_id.items()}

```

 Downloading data from https://s3.amazonaws.com/text-datasets/imdb_word_index.json
 1646592/1641221 [=====] - 1s 1us/step

```
id_to_word
```



{34704: 'fawn',
52009: 'tsukino',
52010: 'nunnery',
16819: 'sonja',
63954: 'vani',
1411: 'woods',
16118: 'spiders',
2348: 'hanging',
2292: 'woody',
52011: 'trawling',
52012: "hold's",
11310: 'comically',
40833: 'localized',
30571: 'disobeying',
52013: "'royale",
40834: "harpo's",
52014: 'canet',
19316: 'aileen',
52015: 'acurately',
52016: "diplomat's",
25245: 'rickman',
6749: 'arranged',
52017: 'rumbustious',
52018: 'familiarness',
52019: "spider'",
68807: 'hahahah',
52020: "wood'",
40836: 'transvestism',
34705: "hangin'",
2341: 'bringing',
40837: 'seamier',
34706: 'wooded',
52021: 'bravora',
16820: 'grueling',
1639: 'wooden',
16821: 'wednesday',
52022: "'prix",
34707: 'altagracia',
52023: 'circuitry',
11588: 'crotch',
57769: 'busybody',
52024: "tart'n'tangy",
14132: 'burgade',
52026: 'thrace',
11041: "tom's",
52028: 'snuggles',
29117: 'francesco',
52030: 'complainers',
52128: 'templarios',
40838: '272',
52031: '273',
52133: 'zaniacs',
34709: '275',
27634: 'consenting',
40839: 'snuggled',
15495: 'inanimate',
52033: 'uality',
11929: 'bronte',
4013: 'errors',
3233: 'dialogs',
52034: "yomada's",
34710: "madman's",
30588: 'dialoge',

52036: 'usenet',
40840: 'videodrome',
26341: "kid'",
52037: 'pawed',
30572: "'girlfriend'",
52038: "'pleasure",
52039: "'reloaded'",
40842: "kazakos'",
52040: 'rocque',
52041: 'mailings',
11930: 'brainwashed',
16822: 'mcanally',
52042: "tom'",
25246: 'kurupt',
21908: 'affiliated',
52043: 'babaganoosh',
40843: "noe's",
40844: 'quart',
362: 'kids',
5037: 'uplifting',
7096: 'controversy',
21909: 'kida',
23382: 'kidd',
52044: "error'",
52045: 'neurologist',
18513: 'spotty',
30573: 'cobblers',
9881: 'projection',
40845: 'fastforwarding',
52046: 'sters',
52047: "eggar's",
52048: 'etherything',
40846: 'gateshead',
34711: 'airball',
25247: 'unsinkable',
7183: 'stern',
52049: "cervi's",
40847: 'dnd',
11589: 'dna',
20601: 'insecurity',
52050: "'reboot'",
11040: 'treikovsky',
52051: 'jaekel',
52052: 'sidebars',
52053: "sforza's",
17636: 'distortions',
52054: 'mutinies',
30605: 'sermons',
40849: '7ft',
52055: 'boobage',
52056: "o'bannon's",
23383: 'populations',
52057: 'chulak',
27636: 'mesmerize',
52058: 'quinnell',
10310: 'yahoo',
52060: 'meteorologist',
42580: 'beswick',
15496: 'boorman',
40850: 'voicework',
52061: "ster'",
22925: 'blustering',
52062: 'hj',
27637: 'intake',

5624: 'morally',
40852: 'jumbling',
52063: 'browsersock',
52064: "'porky's'",
16824: 'gershon',
40853: 'ludicrously',
52065: 'coprophilia',
40854: 'expressively',
19503: "india's",
34713: "post's",
52066: 'wana',
5286: 'wang',
30574: 'wand',
25248: 'wane',
52324: 'edgeways',
34714: 'titanium',
40855: 'pinta',
181: 'want',
30575: 'pinto',
52068: 'whoopeddoodles',
21911: 'tchaikovsky',
2106: 'travel',
52069: "'victory'",
11931: 'copious',
22436: 'gouge',
52070: "chapters'",
6705: 'barbra',
30576: 'uselessness',
52071: "wan'",
27638: 'assimilated',
16119: 'petiot',
52072: 'most\x85and',
3933: 'dinosaurs',
355: 'wrong',
52073: 'seda',
52074: 'stollen',
34715: 'sentencing',
40856: 'ouroboros',
40857: 'assimilates',
40858: 'colorfully',
27639: 'glenne',
52075: 'dongen',
4763: 'subplots',
52076: 'kiloton',
23384: 'chandon',
34716: "effect'",
27640: 'snugly',
40859: 'kuei',
9095: 'welcomed',
30074: 'dishonor',
52078: 'concurrence',
23385: 'stoicism',
14899: "guys'",
52080: "beroemd'",
6706: 'butcher',
40860: "melfi's",
30626: 'aargh',
20602: 'playhouse',
11311: 'wickedly',
1183: 'fit',
52081: 'labratory',
40862: 'lifeline',
1930: 'screaming',
4290: 'fix'.

52082: 'cineliterate',
52083: 'fic',
52084: 'fia',
34717: 'fig',
52085: 'fmvs',
52086: 'fie',
52087: 'reentered',
30577: 'fin',
52088: 'doctresses',
52089: 'fil',
12609: 'zucker',
31934: 'ached',
52091: 'counsel',
52092: 'paterfamilias',
13888: 'songwriter',
34718: 'shivam',
9657: 'hurting',
302: 'effects',
52093: 'slaughter',
52094: "'flame'",
52095: 'sommerset',
52096: 'interwhined',
27641: 'whacking',
52097: 'bartok',
8778: 'barton',
21912: 'frewer',
52098: "fi'",
6195: 'ingrid',
30578: 'sribor',
52099: 'appropriately',
52100: 'wobblyhand',
52101: 'tantalisingly',
52102: 'ankylosaurus',
17637: 'parasites',
52103: 'children',
52104: "jenkins'",
52105: 'metafiction',
17638: 'golem',
40863: 'indiscretion',
23386: "reeves'",
57784: "inamorata's",
52107: 'brittannica',
7919: 'adapt',
30579: "russo's",
48249: 'guitarists',
10556: 'abbott',
40864: 'abbots',
17652: 'lanisha',
40866: 'magickal',
52108: 'mattter',
52109: "'willy",
34719: 'pumpkins',
52110: 'stuntpeople',
30580: 'estimate',
40867: 'ugghhh',
11312: 'gameplay',
52111: "wern't",
40868: "n'sync",
16120: 'sickeningly',
40869: 'chiara',
4014: 'disturbed',
40870: 'portmanteau',
52112: 'ineffectively',
82146: "dubious"

82146: 'ducnonvey s',
37522: 'nasty',
1288: 'purpose',
52115: 'lazars',
28108: 'lightened',
52116: 'kaliganj',
52117: 'popularism',
18514: 'damme's',
30581: 'stylistics',
52118: 'mindgaming',
46452: 'spoilerish',
52120: "'corny'",
34721: 'boerner',
6795: 'olds',
52121: 'bakelite',
27642: 'renovated',
27643: 'forrester',
52122: 'lumiere's',
52027: 'gaskets',
887: 'needed',
34722: 'smight',
1300: 'master',
25908: 'edie's',
40871: 'seeber',
52123: 'hiya',
52124: 'fuzziness',
14900: 'genesis',
12610: 'rewards',
30582: 'enthrall',
40872: "'about",
52125: "recollection's",
11042: 'mutilated',
52126: 'fatherlands',
52127: "fischer's",
5402: 'positively',
34708: '270',
34723: 'ahmed',
9839: 'zatoichi',
13889: 'bannister',
52130: 'anniversaries',
30583: "helm's",
52131: "'work'",
34724: 'exclaimed',
52132: "'unfunny'",
52032: '274',
547: 'feeling',
52134: "wanda's",
33269: 'dolan',
52136: '278',
52137: 'peacoat',
40873: 'brawny',
40874: 'mishra',
40875: 'worlders',
52138: 'protags',
52139: 'skullcap',
57599: 'dastagir',
5625: 'affairs',
7802: 'wholesome',
52140: 'hymen',
25249: 'paramedics',
52141: 'unpersons',
52142: 'heavyarms',
52143: 'affaire',
52144: 'coulisses',

40876: 'hymer',
52145: 'kremlin',
30584: 'shipments',
52146: 'pixilated',
30585: "'00s",
18515: 'diminishing',
1360: 'cinematic',
14901: 'resonates',
40877: 'simplify',
40878: "nature'",
40879: 'temptresses',
16825: 'reverence',
19505: 'resonated',
34725: 'dailey',
52147: '2\x85',
27644: 'treize',
52148: 'majo',
21913: 'kiya',
52149: 'woolnough',
39800: 'thanatos',
35734: 'sandoval',
40882: 'dorama',
52150: "o'shaughnessy",
4991: 'tech',
32021: 'fugitives',
30586: 'teck',
76128: "'e'",
40884: 'doesn't',
52152: 'purged',
660: 'saying',
41098: "martians'",
23421: 'norliss',
27645: 'dickey',
52155: 'dicker',
52156: "'sependipity",
8425: 'padded',
57795: 'ordell',
40885: "sturges'",
52157: 'independentcritics',
5748: 'tempted',
34727: "atkinson's",
25250: 'hounded',
52158: 'apace',
15497: 'clicked',
30587: "'humor'",
17180: "martino's",
52159: "'supporting",
52035: 'warmongering',
34728: "zemeckis's",
21914: 'lube',
52160: 'shocky',
7479: 'plate',
40886: 'plata',
40887: 'sturgess',
40888: "nerds'",
20603: 'plato',
34729: 'plath',
40889: 'platt',
52162: 'mcnab',
27646: 'clumsiness',
3902: 'altogether',
42587: 'massacring',
52163: 'bicenntinial',
40890: 'skaal',

14363: 'droning',
8779: 'lds',
21915: 'jaguar',
34730: "cale's",
1780: 'nicely',
4591: 'mummy',
18516: "lot's",
10089: 'patch',
50205: 'kerkhof',
52164: "leader's",
27647: "'movie",
52165: 'unconfirmed',
40891: 'heirloom',
47363: 'wrangle',
52166: 'emotion\x85',
52167: "'stargate",
40892: 'pinoy',
40893: 'conchatta',
41131: 'broeke',
40894: 'advisedly',
17639: "barker's",
52169: 'descours',
775: 'lots',
9262: 'lotr',
9882: 'irs',
52170: 'lott',
40895: 'xvi',
34731: 'irk',
52171: 'irl',
6890: 'ira',
21916: 'belzer',
52172: 'irc',
27648: 'ire',
40896: 'requisites',
7696: 'discipline',
52964: 'lyoko',
11313: 'extend',
876: 'nature',
52173: "'dickie",
40897: 'optimist',
30589: 'lapping',
3903: 'superficial',
52174: 'vestment',
2826: 'extent',
52175: 'tendons',
52176: "heller's",
52177: 'quagmires',
52178: 'miyako',
20604: 'moocow',
52179: "coles",
40898: 'lookit',
52180: 'ravenously',
40899: 'levitating',
52181: 'perfunctorily',
30590: 'lookin',
40901: "lot'",
52182: 'lookie',
34873: 'fearlessly',
52184: 'libyan',
40902: 'fondles',
35717: 'gopher',
40904: 'wearying',
52185: "nz's",
27640: 'minusce'

27049: 'minuses',
52186: 'purposelessly',
52187: 'shandling',
31271: 'decapitates',
11932: 'humming',
40905: "'nother",
21917: 'smackdown',
30591: 'underdone',
40906: 'frf',
52188: 'triviality',
25251: 'fro',
8780: 'bothers',
52189: "'kensington",
76: 'much',
34733: 'muco',
22618: 'wiseguy',
27651: "richie's",
40907: 'tonino',
52190: 'unleavened',
11590: 'fry',
40908: "'tv'",
40909: 'toning',
14364: 'obese',
30592: 'sensationalized',
40910: 'spiv',
6262: 'spit',
7367: 'arkin',
21918: 'charleton',
16826: 'jeon',
21919: 'boardroom',
4992: 'doubts',
3087: 'spin',
53086: 'hepo',
27652: 'wildcat',
10587: 'venoms',
52194: 'misconstrues',
18517: 'mesmerising',
40911: 'misconstrued',
52195: 'rescinds',
52196: 'prostrate',
40912: 'majid',
16482: 'climbed',
34734: 'canoeing',
52198: 'majin',
57807: 'animie',
40913: 'sylke',
14902: 'conditioned',
40914: 'waddell',
52199: '3\x85',
41191: 'hyperdrive',
34735: 'conditioner',
53156: 'bricklayer',
2579: 'hong',
52201: 'memoriam',
30595: 'inventively',
25252: "levant's",
20641: 'portobello',
52203: 'remand',
19507: 'mummified',
27653: 'honk',
19508: 'spews',
40915: 'visitations',
52204: 'mummifies',
25253: 'cavanaugh',
...

23388: 'zeon',
40916: "jungle's",
34736: 'viertel',
27654: 'frenchmen',
52205: 'torpedoes',
52206: 'schlessinger',
34737: 'torpedoed',
69879: 'blister',
52207: 'cinefest',
34738: 'furlough',
52208: 'mainsequence',
40917: 'mentors',
9097: 'academic',
20605: 'stillness',
40918: 'academia',
52209: 'lonelier',
52210: 'nibby',
52211: "losers'",
40919: 'cineastes',
4452: 'corporate',
40920: 'massaging',
30596: 'bellow',
19509: 'absurdities',
53244: 'expetations',
40921: 'nyfiken',
75641: 'mehras',
52212: 'lasse',
52213: 'visability',
33949: 'militarily',
52214: "elder'",
19026: 'gainsbourg',
20606: 'hah',
13423: 'hai',
34739: 'haj',
25254: 'hak',
4314: 'hal',
4895: 'ham',
53262: 'duffer',
52216: 'haa',
69: 'had',
11933: 'advancement',
16828: 'hag',
25255: "hand'",
13424: 'hay',
20607: 'mcnamara',
52217: "mozart's",
30734: 'duffel',
30597: 'haq',
13890: 'har',
47: 'has',
2404: 'hat',
40922: 'hav',
30598: 'haw',
52218: 'figtings',
15498: 'elders',
52219: 'underpanted',
52220: 'pninson',
27655: 'unequivocally',
23676: "barbara's",
52222: "bello'",
13000: 'indicative',
40923: 'yawnfest',
52223: 'hexploitation',
52224: "loder's",

27656: 'sleuthing',
32625: "justin's",
52225: "'ball",
52226: "'summer",
34938: "'demons'",
52228: "mormon's",
34740: "laughton's",
52229: 'debell',
39727: 'shipyard',
30600: 'unabashedly',
40404: 'disks',
2293: 'crowd',
10090: 'crowe',
56437: "vancouver's",
34741: 'mosques',
6630: 'crown',
52230: 'culpas',
27657: 'crows',
53347: 'surrell',
52232: 'flowless',
52233: 'sheirk',
40926: "'three",
52234: "peterson'",
52235: 'ooverall',
40927: 'perchance',
1324: 'bottom',
53366: 'chabert',
52236: 'sneha',
13891: 'inhuman',
52237: 'ichii',
52238: 'ursla',
30601: 'completly',
40928: 'moviedom',
52239: 'raddick',
51998: 'brundage',
40929: 'brigades',
1184: 'starring',
52240: "'goal'",
52241: 'caskets',
52242: 'willcock',
52243: "threesome's",
52244: "mosque'",
52245: "cover's",
17640: 'spaceships',
40930: 'anomalous',
27658: 'ptsd',
52246: 'shirdan',
21965: 'obscenity',
30602: 'lemmings',
30603: 'duccio',
52247: "levene's",
52248: "'gorby'",
25258: "teenager's",
5343: 'marshall',
9098: 'honeymoon',
3234: 'shoots',
12261: 'despised',
52249: 'okabasho',
8292: 'fabric',
18518: 'cannavale',
3540: 'raped',
52250: "tutt's",
17641: 'grasping',
18519: 'desnises'

40931: 'thief's',
8929: 'rapes',
52251: 'raper',
27659: 'eyre',
52252: 'walchek',
23389: 'elmo's',
40932: 'perfumes',
21921: 'spurting',
52253: 'exposition'\x85',
52254: 'denoting',
34743: 'thesaurus',
40933: 'shoot',
49762: 'bonejack',
52256: 'simpsonian',
30604: 'hebetude',
34744: 'hallow's',
52257: 'desperation\x85',
34745: 'incinerator',
10311: 'congratulations',
52258: 'humbled',
5927: 'else's',
40848: 'treikovski',
52259: 'rape',
59389: "'chapters'",
52260: '1600s',
7256: 'martian',
25259: 'nicest',
52262: 'eyred',
9460: 'passenger',
6044: 'disgrace',
52263: 'moderne',
5123: 'barrymore',
52264: 'yankovich',
40934: 'moderns',
52265: 'studlied',
52266: 'bedsheet',
14903: 'decapitation',
52267: 'slurring',
52268: "'nunsplotation'",
34746: "'character'",
9883: 'cambodia',
52269: 'rebellious',
27660: 'pasadena',
40935: 'crowne',
52270: "'bedchamber",
52271: 'conjectural',
52272: 'appologize',
52273: 'halfassing',
57819: 'paycheque',
20609: 'palms',
52274: "'islands",
40936: 'hawked',
21922: 'palme',
40937: 'conservatively',
64010: 'larp',
5561: 'palma',
21923: 'smelling',
13001: 'aragorn',
52275: 'hawker',
52276: 'hawkes',
3978: 'explosions',
8062: 'loren',
52277: 'pyle's',
52278: 'pyle's',

6707: 'snootout',
18520: 'mike's',
52278: 'driscoll's',
40938: 'cogsworth',
52279: 'britian's',
34747: 'childs',
52280: 'portrait's',
3629: 'chain',
2500: 'whoever',
52281: 'puttered',
52282: 'childe',
52283: 'maywether',
3039: 'chair',
52284: 'rance's',
34748: 'machu',
4520: 'ballet',
34749: 'grapples',
76155: 'summerize',
30606: 'freelance',
52286: 'andrea's',
52287: '\x91very',
45882: 'coolidge',
18521: 'mache',
52288: 'balled',
40940: 'grappled',
18522: 'macha',
21924: 'underlining',
5626: 'macho',
19510: 'oversight',
25260: 'machi',
11314: 'verbally',
21925: 'tenacious',
40941: 'windshields',
18560: 'paychecks',
3399: 'jerk',
11934: "good'",
34751: 'prancer',
21926: 'prances',
52289: 'olympus',
21927: 'lark',
10788: 'embark',
7368: 'gloomy',
52290: 'jehaan',
52291: 'turaqui',
20610: "child'",
2897: 'locked',
52292: 'pranced',
2591: 'exact',
52293: 'unattuned',
786: 'minute',
16121: 'skewed',
40943: 'hodgins',
34752: 'skewer',
52294: 'think\x85',
38768: 'rosenstein',
52295: 'helmit',
34753: 'wrestlemanias',
16829: 'hindered',
30607: "martha's",
52296: 'cheree',
52297: "pluckin'",
40944: 'ogles',
11935: 'heavyweight',
82193: 'aada',

11315: 'chopping',
61537: 'strongboy',
41345: 'hegemonic',
40945: 'adorns',
41349: 'xxth',
34754: 'nobuhiro',
52301: 'capitães',
52302: 'kavogianni',
13425: 'antwerp',
6541: 'celebrated',
52303: 'roarke',
40946: 'baggins',
31273: 'cheeseburgers',
52304: 'matras',
52305: "nineties",
52306: "'craig'",
13002: 'celebrates',
3386: 'unintentionally',
14365: 'drafted',
52307: 'climby',
52308: '303',
18523: 'oldies',
9099: 'climbs',
9658: 'honour',
34755: 'plucking',
30077: '305',
5517: 'address',
40947: 'menjou',
42595: "'freak'",
19511: 'dwindling',
9461: 'benson',
52310: 'white's',
40948: 'shamelessness',
21928: 'impacted',
52311: 'upatz',
3843: 'cusack',
37570: "flavia's",
52312: 'effette',
34756: 'influx',
52313: 'booooooooo',
52314: 'dimitrova',
13426: 'houseman',
25262: 'bigas',
52315: 'boyle',
52316: 'phillipenes',
40949: 'fakery',
27661: "grandpa's",
27662: 'darnell',
19512: 'undergone',
52318: 'handbags',
21929: 'perished',
37781: 'pooped',
27663: 'vigour',
3630: 'opposed',
52319: 'etude',
11802: "caine's",
52320: 'doozers',
34757: 'photojournals',
52321: 'perishes',
34758: 'constrains',
40951: 'migenes',
30608: 'consoled',
16830: 'alastair',
52322: 'wvs'.

52323: 'ooooooh',
34759: 'approving',
40952: 'consoles',
52067: 'disparagement',
52325: 'futureistic',
52326: 'rebounding',
52327: "'date",
52328: 'gregoire',
21930: 'rutherford',
34760: 'americanised',
82199: 'novikov',
1045: 'following',
34761: 'munroe',
52329: "morita'",
52330: 'christenssen',
23109: 'oatmeal',
25263: 'fossey',
40953: 'livered',
13003: 'listens',
76167: "'marci",
52333: "otis's",
23390: 'thanking',
16022: 'maude',
34762: 'extensions',
52335: 'ameteurish',
52336: "commender's",
27664: 'agricultural',
4521: 'convincingly',
17642: 'fueled',
54017: 'mahattan',
40955: "paris's",
52339: 'vulkan',
52340: 'stapes',
52341: 'odysessy',
12262: 'harmon',
4255: 'surfing',
23497: 'halloran',
49583: 'unbelieveably',
52342: "'offed'",
30610: 'quadrant',
19513: 'inhabiting',
34763: 'nebbish',
40956: 'forebears',
34764: 'skirmish',
52343: 'ocassionally',
52344: "'resist",
21931: 'impactful',
52345: 'spicier',
40957: 'touristy',
52346: "'football'",
40958: 'webpage',
52348: 'exurbia',
52349: 'jucier',
14904: 'professors',
34765: 'structuring',
30611: 'jig',
40959: 'overlord',
25264: 'disconnect',
82204: 'sniffle',
40960: 'slimeball',
40961: 'jia',
16831: 'milked',
40962: 'banjoes',
12222: 'sim'

1240: 'jim',
52351: 'workforces',
52352: 'jip',
52353: 'rotweiller',
34766: 'mundaneness',
52354: "'ninja'",
11043: "dead'",
40963: "cipriani's",
20611: 'modestly',
52355: "professor'",
40964: 'shacked',
34767: 'bashful',
23391: 'sorter',
16123: 'overpowering',
18524: 'workmanlike',
27665: 'henpecked',
18525: 'sorted',
52357: "jōb's",
52358: "'always",
34768: "'baptists",
52359: 'dreamcatchers',
52360: "'silence'",
21932: 'hickory',
52361: 'fun\x97yet',
52362: 'breakumentary',
15499: 'didn',
52363: 'didi',
52364: 'pealing',
40965: 'dispite',
25265: "italy's",
21933: 'instability',
6542: 'quarter',
12611: 'quartet',
52365: 'padmé',
52366: "'bleedmedry",
52367: 'pahalniuk',
52368: 'honduras',
10789: 'bursting',
41468: "pablo's",
52370: 'irremediably',
40966: 'presages',
57835: 'bowlegged',
65186: 'dalip',
6263: 'entering',
76175: 'newsradio',
54153: 'presaged',
27666: "giallo's",
40967: 'bouyant',
52371: 'amerterish',
18526: 'rajni',
30613: 'leeves',
34770: 'macauley',
615: 'seriously',
52372: 'sugercoma',
52373: 'grimstead',
52374: "'fairy'",
30614: 'zenda',
52375: "'twins'",
17643: 'realisation',
27667: 'highsmith',
7820: 'raunchy',
40968: 'incentives',
52377: 'flatson',
35100: 'snooker',

```

16832: 'crazies',
14905: 'crazier',
7097: 'grandma',
52378: 'napunsaktha',
30615: 'workmanship',
52379: 'reisner',
61309: "sanford's",
52380: '\x91doña',
6111: 'modest',
19156: "everything's",
40969: 'hamer',
52382: "couldn't",
13004: 'quibble',
52383: 'socking',
21934: 'tingler',
52384: 'gutman',
40970: 'lachlan',
52385: 'tableaus',
52386: 'headbanger',
2850: 'spoken',
34771: 'cerebrally',
23493: "'road",
21935: 'tableaux',
40971: "proust's",
40972: 'periodical',
52388: "shoveller's",
25266: 'tamara',
17644: 'affords',
3252: 'concert',
87958: "yara's",
52389: 'someone',
8427: 'lingering',
41514: "abraham's",
34772: 'beesley',
34773: 'cherbourg',
28627: 'kagan',
9100: 'snatch',
9263: "miyazaki's",
25267: 'absorbs',
40973: "koltai's",
64030: 'tingled',
19514: 'crossroads',
16124: 'rehab',
52392: 'falworth',
52393: 'sequals',
...}

```

```

print("Train-set size: ", len(x_train))
print("Test-set size: ", len(x_test))

```

```

☞ Train-set size: 25000
   Test-set size: 25000

```

```
np.array(x_train[1])
```

```
☞
```

```
array([[ 0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  1, 194, 1153, 194, 8255, 78, 228, 5, 6, 1463,
4369, 5012, 134, 26, 4, 715, 8, 118, 1634, 14, 394,
20, 13, 119, 954, 189, 102, 5, 207, 110, 3103, 21,
14, 69, 188, 8, 30, 23, 7, 4, 249, 126, 93,
4, 114, 9, 2300, 1523, 5, 647, 4, 116, 9, 35,
8163, 4, 229, 9, 340, 1322, 4, 118, 9, 4, 130,
4901, 19, 4, 1002, 5, 89, 29, 952, 46, 37, 4,
455, 9, 45, 43, 38, 1543, 1905, 398, 4, 1649, 26,
6853, 5, 163, 11, 3215, 2, 4, 1153, 9, 194, 775,
7, 8255, 2, 349, 2637, 148, 605, 2, 8003, 15, 123,
125, 68, 2, 6853, 15, 349, 165, 4362, 98, 5, 4,
228, 9, 43, 2, 1157, 15, 299, 120, 5, 120, 174,
11, 220, 175, 136, 50, 9, 4373, 228, 8255, 5, 2,
656, 245, 2350, 5, 4, 9837, 131, 152, 491, 18, 2,
32, 7464, 1212, 14, 9, 6, 371, 78, 22, 625, 64,
1382, 9, 8, 168, 145, 23, 4, 1690, 15, 16, 4,
1355, 5, 28, 6, 52, 154, 462, 33, 89, 78, 285,
16, 145, 95], dtype=int32)
```

▼ Build Keras Embedding Layer Model

We can think of the Embedding layer as a dictionary that maps a index assigned to a word to a word vector. This is used in a few ways:

- The embedding layer can be used at the start of a larger deep learning model.
- Also we could load pre-train word embeddings into the embedding layer when we create our model.
- Use the embedding layer to train our own word2vec models.

The keras embedding layer doesn't require us to onehot encode our words, instead we have to give each word an index. For the imdb dataset we've loaded this has already been done, but if this wasn't the case we could use sklearn

```
# Building the model
```

```
model = Sequential()
embedding_size = 8
```

```
⚠ WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:3974: tf.nn.
```

```
model.add(Embedding(input_dim=vocab_size,
                    output_dim=embedding_size,
                    name='Embedding_layer'))
```

```
⌕ WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:397: tf.nn.conv2d is deprecated and will be removed in a future version.
Instructions for updating:
Use tf.nn.conv2d_v2 instead.
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:397: tf.nn.conv2d is deprecated and will be removed in a future version.
Instructions for updating:
Use tf.nn.conv2d_v2 instead.
```

```
model.add(LSTM(units=16, return_sequences=True))

model.add(LSTM(units=8, return_sequences=True))

model.add(LSTM(units=4))

model.add(Dense(1, activation='sigmoid'))

optimizer = Adam(lr=1e-3)
```

```
model.compile(loss='binary_crossentropy',
              optimizer=optimizer,
              metrics=['accuracy'])
```

```
⌕ WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/optimizers.py:793: The name tf.nn.conv2d is deprecated. Please use tf.nn.conv2d_v2 instead.
Instructions for updating:
Use tf.nn.conv2d in 2.0, which has the same broadcast rule as np.where
```

```
model.summary()
```

⌕ Model: "sequential_1"

Layer (type)	Output Shape	Param #
=====		
Embedding_layer (Embedding)	(None, None, 8)	80000

lstm_1 (LSTM)	(None, None, 16)	1600

lstm_2 (LSTM)	(None, None, 8)	800

lstm_3 (LSTM)	(None, 4)	208

dense_1 (Dense)	(None, 1)	5
=====		
Total params: 82,613		
Trainable params: 82,613		
Non-trainable params: 0		

```
model.fit(x_train, y_train,
        validation_split=0.05, epochs=2, batch_size=64)
```

⌕

Train on 23750 samples, validate on 1250 samples

Epoch 1/2

23750/23750 [=====] - 421s 18ms/step - loss: 0.4538 - acc: 0.7950 - va

Epoch 2/2

23750/23750 [=====] - 402s 17ms/step - loss: 0.2547 - acc: 0.9055 - va

<keras.callbacks.History at 0x7f4d98208588>

```
result = model.evaluate(x_test, y_test)
```

```
25000/25000 [=====] - 146s 6ms/step
```

Report the Accuracy of the model. (5 points)

```
print("Accuracy: {}".format(result[1]))
```

```
Accuracy: 0.871
```

```
print("Accuracy: {:.2%}".format(result[1]))
```

```
Accuracy: 87.10%
```

Retrive the output of each layer in keras for a given single test sample from built

```
import tensorflow as tf
from keras import backend as k
```

```
#Define function to get output of all the layers in the model for specific test input
```

```
with tf.Session() as sess:
```

```
    sess.run(tf.global_variables_initializer())
```

```
    sess.run(tf.tables_initializer())
```

```
def getLayerOutput(layer):
```

```
    get_Layer_Output = k.function([model.layers[0].input], [layer.output])
```

```
    return get_Layer_Output([x_test[0:1,]])[0]
```

```
layer_output = []
```

```
for layer in model.layers:
```

```
    layer_output.append(getLayerOutput(layer))
```

```
print(layer_output)
```

```
[array([[[ 0.0041155 , -0.04370933,  0.01085538, ...,  0.02347941,
          -0.01528121, -0.04100442],
         [ 0.0041155 , -0.04370933,  0.01085538, ...,  0.02347941,
          -0.01528121, -0.04100442],
         [ 0.0041155 , -0.04370933,  0.01085538, ...,  0.02347941,
          -0.01528121, -0.04100442],
         ...,
         [ 0.04258528,  0.03693824, -0.02728037, ..., -0.03450607,
           0.02184614,  0.02130951],
         [-0.02791059, -0.01948614, -0.03185003, ..., -0.03451742,
           0.01664151, -0.03993685],
         [-0.02878357, -0.03301308, -0.01381617, ...,  0.01601266,
           0.01202636,  0.03784463]]], dtype=float32), array([[[ 0.00486718, -0.00203985, -0.001
          -0.00370962, -0.00099983],
         [ 0.00765251, -0.0043822 , -0.00185332, ...,  0.00443793,
          -0.00594004, -0.00244151],
         [ 0.00914937, -0.0067409 , -0.00180035, ...,  0.00535144,
          -0.00735939, -0.00391256],
         ...,
         [-0.00976832,  0.00027523,  0.00278522, ..., -0.00364425,
           0.00289563,  0.00205749],
         [-0.0089741 ,  0.00143643,  0.00037401, ..., -0.00136078,
           0.00743207,  0.0028485 ],
         [-0.00260326,  0.00127924, -0.00212376, ..., -0.00215453,
           0.00725914,  0.0080314 ]]], dtype=float32), array([[-0.00032545, -0.0003693 , -0.000
          -0.0005567 ,  0.00119426],
         [-0.00107241, -0.00074781, -0.00225401, ..., -0.00329013,
          -0.00126665,  0.00268792],
         [-0.00215548, -0.0009654 , -0.00354357, ..., -0.00509788,
          -0.00195908,  0.00403407],
         ...,
         [ 0.00080943,  0.00174695, -0.00021325, ...,  0.00236173,
           0.00429004, -0.00336745],
         [ 0.00163181,  0.00281414,  0.00076217, ...,  0.00363725,
           0.0036981 , -0.00439277],
         [ 0.00196767,  0.00349656,  0.00135669, ...,  0.00359932,
           0.00220501, -0.00449431]]], dtype=float32), array([[ 0.00103098, -0.00036572, -0.0028
          dtype=float32), array([[0.4999851]], dtype=float32)]
```

```
# Getting a specific output layer
layer_output[2]
```

```
↳ array([[-0.00032545, -0.0003693 , -0.00096983, ..., -0.00142934,
          -0.0005567 ,  0.00119426],
         [-0.00107241, -0.00074781, -0.00225401, ..., -0.00329013,
          -0.00126665,  0.00268792],
         [-0.00215548, -0.0009654 , -0.00354357, ..., -0.00509788,
          -0.00195908,  0.00403407],
         ...,
         [ 0.00080943,  0.00174695, -0.00021325, ...,  0.00236173,
           0.00429004, -0.00336745],
         [ 0.00163181,  0.00281414,  0.00076217, ...,  0.00363725,
           0.0036981 , -0.00439277],
         [ 0.00196767,  0.00349656,  0.00135669, ...,  0.00359932,
           0.00220501, -0.00449431]]], dtype=float32)
```

