

▼ Sentiment Classification

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

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

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

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

Dataset

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

Dataset of 25,000 movies reviews from IMDB, labeled by sentiment (positive/negative). Reviews have
encoded as a sequence of word indexes (integers). For convenience, the words are indexed by their fr
that has index 1 is the most frequent word. Use the first 20 words from each review to speed up traini

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

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. Retrive the output of each layer in keras for a given single test sample from the trained model you built. (2.5 p

▼ Usage:

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

```
vocab_size = 10000 #vocab size
```

```
(x_train, y_train), (x_test, y_test) = imdb.load_data(num_words=vocab_size) # vocab_size is n
```

☞ Using TensorFlow backend.

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 %tensc
1.x magic: [more info](#).

Downloading data from <https://s3.amazonaws.com/text-datasets/imdb.npz>

17165311/17161789 [-----] - 25.0MB/s

```
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)
```

```
x_train[0]
```

```

↳ 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,  1, 14, 22, 16, 43, 530,
        973, 1622, 1385,  65, 458, 4468,  66, 3941,  4, 173,  36,
        256,   5,  25, 100,  43, 838, 112,  50, 670,  2,   9,
          35, 480, 284,   5, 150,   4, 172, 112, 167,  2, 336,
        385,  39,   4, 172, 4536, 1111,  17, 546,  38, 13, 447,
          4, 192,  50,  16,   6, 147, 2025,  19, 14, 22,   4,
       1920, 4613, 469,   4,  22,  71,  87, 12, 16, 43, 530,
          38,  76,  15, 13, 1247,   4,  22, 17, 515, 17, 12,
          16, 626,  18,   2,   5,  62, 386, 12,   8, 316,   8,
         106,   5,   4, 2223, 5244,  16, 480,  66, 3785,  33,   4,
        130,  12,  16,  38, 619,   5,  25, 124,  51,  36, 135,
          48,  25, 1415,  33,   6,  22, 12, 215,  28,  77,  52,
           5,  14,  407,  16,  82,   2,   8,   4, 107, 117, 5952,
          15, 256,   4,   2,   7, 3766,   5, 723,  36,  71,  43,
         530, 476,  26, 400, 317,  46,   7,   4,   2, 1029, 13,
        104,  88,   4, 381,  15, 297,  98,  32, 2071,  56,  26,
        141,   6, 194, 7486,  18,   4, 226,  22,  21, 134, 476,
          26, 480,   5, 144,  30, 5535,  18,  51,  36,  28, 224,
          92,  25, 104,   4, 226,  65,  16,  38, 1334,  88,  12,
          16, 283,   5,  16, 4472, 113, 103,  32,  15,  16, 5345,
          19, 178,  32], dtype=int32)

```

```
x_test[0]
```



```
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,  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, 591, 202, 14, 31, 6, 717, 10, 10, 2,
        2, 5, 4, 360, 7, 4, 177, 5760, 394, 354, 4,
       123, 9, 1035, 1035, 1035, 10, 10, 13, 92, 124, 89,
       488, 7944, 100, 28, 1668, 14, 31, 23, 27, 7479, 29,
       220, 468, 8, 124, 14, 286, 170, 8, 157, 46, 5,
        27, 239, 16, 179, 2, 38, 32, 25, 7944, 451, 202,
       14, 6, 717], dtype=int32)
```

```
import tensorflow as tf
word_to_id = tf.keras.datasets.imdb.get_word_index()
word_to_id = {k:(v) for k,v in word_to_id.items()}
```

```
↳ Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-datasets/imdb\_w  
1646592/1641221 [=====] - 0s 0us/step
```

word_to_id


```
{'fawn': 34701,
'tsukino': 52006,
'nunnery': 52007,
'sonja': 16816,
'vani': 63951,
'woods': 1408,
'spiders': 16115,
'hanging': 2345,
'woody': 2289,
'trawling': 52008,
"hold's": 52009,
'comically': 11307,
'localized': 40830,
'disobeying': 30568,
"'royale": 52010,
"harpo's": 40831,
'canet': 52011,
'aileen': 19313,
'acurately': 52012,
'diplomat's": 52013,
'rickman': 25242,
'arranged': 6746,
'rumbustious': 52014,
'familiarness': 52015,
"spider'": 52016,
'hahahah': 68804,
"wood'": 52017,
'transvestism': 40833,
"hangin'": 34702,
'bringing': 2338,
'seamier': 40834,
'wooded': 34703,
'bravora': 52018,
'grueling': 16817,
'wooden': 1636,
'wednesday': 16818,
"'prix": 52019,
'altagracia': 34704,
'circuitry': 52020,
'crotch': 11585,
'busybody': 57766,
"tart'n'tangy": 52021,
'burgade': 14129,
'thrace': 52023,
"tom's": 11038,
'snuggles': 52025,
'francesco': 29114,
'complainers': 52027,
'templarios': 52125,
'272': 40835,
'273': 52028,
'zaniacs': 52130,
'275': 34706,
'consenting': 27631,
'snuggled': 40836,
'inanimate': 15492,
'uality': 52030,
```

```
'bronte': 11926,  
'errors': 4010,  
'dialogs': 3230,  
"yomada's": 52031,  
"madman's": 34707,  
'dialoge': 30585,  
'usenet': 52033,  
'videodrome': 40837,  
"kid'": 26338,  
'pawed': 52034,  
"'girlfriend'": 30569,  
"'pleasure": 52035,  
"'reloaded'": 52036,  
"kazakos'": 40839,  
'rocque': 52037,  
'mailings': 52038,  
'brainwashed': 11927,  
'mcanally': 16819,  
"tom'": 52039,  
'kurupt': 25243,  
'affiliated': 21905,  
'babaganoosh': 52040,  
"noe's": 40840,  
'quart': 40841,  
'kids': 359,  
'uplifting': 5034,  
'controversy': 7093,  
'kida': 21906,  
'kidd': 23379,  
"error'": 52041,  
'neurologist': 52042,  
'spotty': 18510,  
'cobblers': 30570,  
'projection': 9878,  
'fastforwarding': 40842,  
'sters': 52043,  
"eggar's": 52044,  
'etherything': 52045,  
'gateshead': 40843,  
'airball': 34708,  
'unsinkable': 25244,  
'stern': 7180,  
"cervi's": 52046,  
'dnd': 40844,  
'dna': 11586,  
'insecurity': 20598,  
"'reboot'": 52047,  
'treikovsky': 11037,  
'jaekel': 52048,  
'sidebars': 52049,  
"sforza's": 52050,  
'distortions': 17633,  
'mutinies': 52051,  
'sermons': 30602,  
'7ft': 40846,  
'boobage': 52052,  
"o'bannon's": 52053,  
'populations': 23380,
```

'chulak': 52054,
'mesmerize': 27633,
'quinnell': 52055,
'yahoo': 10307,
'meteorologist': 52057,
'beswick': 42577,
'boorman': 15493,
'voicework': 40847,
'ster': 52058,
'blustering': 22922,
'hj': 52059,
'intake': 27634,
'morally': 5621,
'jumbling': 40849,
'bowersock': 52060,
'porky's': 52061,
'gershon': 16821,
'ludicrously': 40850,
'coprophilia': 52062,
'expressively': 40851,
'india's': 19500,
'post's': 34710,
'wana': 52063,
'wang': 5283,
'wand': 30571,
'wane': 25245,
'edgeways': 52321,
'titanium': 34711,
'pinta': 40852,
'want': 178,
'pinto': 30572,
'whoopdedoodles': 52065,
'tchaikovsky': 21908,
'travel': 2103,
'victory': 52066,
'copious': 11928,
'gouge': 22433,
'chapters': 52067,
'barbra': 6702,
'uselessness': 30573,
'wan': 52068,
'assimilated': 27635,
'petiot': 16116,
'most\x85and': 52069,
'dinosaurs': 3930,
'wrong': 352,
'seda': 52070,
'stollen': 52071,
'sentencing': 34712,
'ouroboros': 40853,
'assimilates': 40854,
'colorfully': 40855,
'glenne': 27636,
'dongen': 52072,
'subplots': 4760,
'kiloton': 52073,
'chandon': 23381,
'affair': 21712

```
effect': 34713,  
'snugly': 27637,  
'kuei': 40856,  
'welcomed': 9092,  
'dishonor': 30071,  
'concurrence': 52075,  
'stoicism': 23382,  
"guys'": 14896,  
"beroemd'": 52077,  
'butcher': 6703,  
"melfi's'": 40857,  
'aargh': 30623,  
'playhouse': 20599,  
'wickedly': 11308,  
'fit': 1180,  
'labratory': 52078,  
'lifeline': 40859,  
'screaming': 1927,  
'fix': 4287,  
'cineliterate': 52079,  
'fic': 52080,  
'fia': 52081,  
'fig': 34714,  
'fmvs': 52082,  
'fie': 52083,  
'reentered': 52084,  
'fin': 30574,  
'doctresses': 52085,  
'fil': 52086,  
'zucker': 12606,  
'ached': 31931,  
'counsil': 52088,  
'paterfamilias': 52089,  
'songwriter': 13885,  
'shivam': 34715,  
'hurting': 9654,  
'effects': 299,  
'slauther': 52090,  
"'flame'": 52091,  
'sommerset': 52092,  
'interwhined': 52093,  
'whacking': 27638,  
'bartok': 52094,  
'barton': 8775,  
'frewer': 21909,  
"fi'": 52095,  
'ingrid': 6192,  
'stribor': 30575,  
'appropriately': 52096,  
'wobblyhand': 52097,  
'tantalisingly': 52098,  
'ankylosaurus': 52099,  
'parasites': 17634,  
'children': 52100,  
"jenkins'": 52101,  
'metafiction': 52102,  
'golem': 17635,  
'indiscretion': 40860,
```

"reeves": 23383,
"inamorata's": 57781,
'brittannica': 52104,
'adapt': 7916,
"russo's": 30576,
'guitarists': 48246,
'abbott': 10553,
'abbots': 40861,
'lanisha': 17649,
'magickal': 40863,
'mattter': 52105,
"'willy": 52106,
'pumpkins': 34716,
'stuntpeople': 52107,
'estimate': 30577,
'ugghhh': 40864,
'gameplay': 11309,
"wern't": 52108,
"n'sync": 40865,
'sickeningly': 16117,
'chiara': 40866,
'disturbed': 4011,
'portmanteau': 40867,
'ineffectively': 52109,
"duchonvey's": 82143,
"nasty": 37519,
'purpose': 1285,
'lazars': 52112,
'lightened': 28105,
'kaliganj': 52113,
'popularism': 52114,
"damme's": 18511,
'stylistics': 30578,
'mindgaming': 52115,
'spoilerish': 46449,
"'corny": 52117,
'boerner': 34718,
'olds': 6792,
'bakelite': 52118,
'renovated': 27639,
'forrester': 27640,
"lumiere's": 52119,
'gaskets': 52024,
'needed': 884,
'smight': 34719,
'master': 1297,
"edie's": 25905,
'seeber': 40868,
'hiya': 52120,
'fuzziness': 52121,
'genesis': 14897,
'rewards': 12607,
'enthrall': 30579,
"'about": 40869,
"recollection's": 52122,
'mutilated': 11039,
'fatherlands': 52123,
"fischer's": 52124,


```

'positively': 5399,
'270': 34705,
'ahmed': 34720,
'zatoichi': 9836,
'bannister': 13886,
'anniversaries': 52127,
'helm's": 30580,
"'work'": 52128,
'exclaimed': 34721,
"'unfunny'": 52129,
'274': 52029,
'feeling': 544,
"wanda's": 52131,
'dolan': 33266,
'278': 52133,
'peacoat': 52134,
'brawny': 40870,
'mishra': 40871,
'worlders': 40872,
'protags': 52135,
'skullcap': 52136,
'dastagir': 57596,
'affairs': 5622,
'wholesome': 7799,
'hymen': 52137,
'paramedics': 25246,
'unpersons': 52138,
'heavyarms': 52139,
'affaire': 52140,
'couliesses': 52141,
'hymen': 40873,
'kremlin': 52142,
'shipments': 30581,
'pixilated': 52143,
"'00s": 30582,
'diminishing': 18512,
'cinematic': 1357,
'resonates': 14898,
'simplify': 40874,
"nature'": 40875,
'temptresses': 40876,
'reverence': 16822,
'resonated': 19502,
'dailey': 34722,
'2\x85': 52144,
'treize': 27641,
'majo': 52145,
'kiya': 21910,
'woolnough': 52146,
'thanatos': 39797,
'sandoval': 35731,
'dorama': 40879,
"o'shaughnessy": 52147,
'tech': 4988,
'fugitives': 32018,
'teck': 30583,
"'e'": 76125,
'doan'": 10001

```

```
uoesn l . 40001,  
'purged': 52149,  
'saying': 657,  
"martians": 41095,  
'norliss': 23418,  
'dickey': 27642,  
'dicker': 52152,  
"'sependipity": 52153,  
'padded': 8422,  
'ordell': 57792,  
"sturges": 40882,  
'independentcritics': 52154,  
'tempted': 5745,  
"atkinson's": 34724,  
'hounded': 25247,  
'apace': 52155,  
'clicked': 15494,  
"'humor'": 30584,  
"martino's": 17177,  
"'supporting": 52156,  
'warmongering': 52032,  
"zemeckis's": 34725,  
'lube': 21911,  
'shocky': 52157,  
'plate': 7476,  
'plata': 40883,  
'sturgess': 40884,  
"nerds": 40885,  
'plato': 20600,  
'plath': 34726,  
'platt': 40886,  
'mcnab': 52159,  
'clumsiness': 27643,  
'altogether': 3899,  
'massacring': 42584,  
'bicenntinial': 52160,  
'skaal': 40887,  
'droning': 14360,  
'lds': 8776,  
'jaguar': 21912,  
"cale's": 34727,  
'nicely': 1777,  
'mummy': 4588,  
"lot's": 18513,  
'patch': 10086,  
'kerkhof': 50202,  
"leader's": 52161,  
"'movie": 27644,  
'unconfirmed': 52162,  
'heirloom': 40888,  
'wrangle': 47360,  
'emotion\x85': 52163,  
"'stargate'": 52164,  
'pinoy': 40889,  
'conchatta': 40890,  
'broeke': 41128,  
'advisedly': 40891,  
"barker's": 17636,
```

'descours': 52166,
'lots': 772,
'lotr': 9259,
'irs': 9879,
'lott': 52167,
'xvi': 40892,
'irk': 34728,
'irl': 52168,
'ira': 6887,
'belzer': 21913,
'irc': 52169,
'ire': 27645,
'requisites': 40893,
'discipline': 7693,
'lyoko': 52961,
'extend': 11310,
'nature': 873,
"'dickie'": 52170,
'optimist': 40894,
'lapping': 30586,
'superficial': 3900,
'vestment': 52171,
'extent': 2823,
'tendons': 52172,
"heller's": 52173,
'quagmires': 52174,
'miyako': 52175,
'moocow': 20601,
"coles'": 52176,
'lookit': 40895,
'ravenously': 52177,
'levitating': 40896,
'perfunctorily': 52178,
'lookin': 30587,
"lot'": 40898,
'lookie': 52179,
'fearlessly': 34870,
'libyan': 52181,
'fondles': 40899,
'gopher': 35714,
'wearying': 40901,
"nz's": 52182,
'minuses': 27646,
'puposelessly': 52183,
'shandling': 52184,
'decapitates': 31268,
'humming': 11929,
"'nother": 40902,
'smackdown': 21914,
'underdone': 30588,
'frf': 40903,
'triviality': 52185,
'fro': 25248,
'bothers': 8777,
"'kensington": 52186,
'much': 73,
'muco': 34730,
'wiseguv': 22615,

```

"richie's": 27648,
'tonino': 40904,
'unleavened': 52187,
'fry': 11587,
"'tv'": 40905,
'toning': 40906,
'obese': 14361,
'sensationalized': 30589,
'spiv': 40907,
'spit': 6259,
'arkin': 7364,
'charleton': 21915,
'jeon': 16823,
'boardroom': 21916,
'doubts': 4989,
'spin': 3084,
'hepo': 53083,
'wildcat': 27649,
'venoms': 10584,
'misconstrues': 52191,
'mesmerising': 18514,
'misconstrued': 40908,
'rescinds': 52192,
'prostrate': 52193,
'majid': 40909,
'climbed': 16479,
'canoeing': 34731,
'majin': 52195,
'animie': 57804,
'sylke': 40910,
'conditioned': 14899,
'waddell': 40911,
'3\x85': 52196,
'hyperdrive': 41188,
'conditioner': 34732,
'bricklayer': 53153,
'hong': 2576,
'memoriam': 52198,
'inventively': 30592,
"levant's": 25249,
'portobello': 20638,
'remand': 52200,
'mummified': 19504,
'honk': 27650,
'spews': 19505,
'visitations': 40912,
'mummifies': 52201,
'cavanaugh': 25250,
'zeon': 23385,
"jungle's": 40913,
'viertel': 34733,
'frenchmen': 27651,
'torpedoes': 52202,
'schlessinger': 52203,
'torpedoed': 34734,
'blister': 69876,
'cinefest': 52204,
'frenchmen': 27651

```

```
turlough : 34735,  
'mainsequence': 52205,  
'mentors': 40914,  
'academic': 9094,  
'stillness': 20602,  
'academia': 40915,  
'lonelier': 52206,  
'nibby': 52207,  
"losers'": 52208,  
'cineastes': 40916,  
'corporate': 4449,  
'massaging': 40917,  
'bellow': 30593,  
'absurdities': 19506,  
'expetations': 53241,  
'nyfiken': 40918,  
'mehras': 75638,  
'lasse': 52209,  
'visability': 52210,  
'militarily': 33946,  
"elder'": 52211,  
'gainsbourg': 19023,  
'hah': 20603,  
'hai': 13420,  
'haj': 34736,  
'hak': 25251,  
'hal': 4311,  
'ham': 4892,  
'duffer': 53259,  
'haa': 52213,  
'had': 66,  
'advancement': 11930,  
'hag': 16825,  
"hand'": 25252,  
'hay': 13421,  
'mcnamara': 20604,  
"mozart's": 52214,  
'duffel': 30731,  
'haq': 30594,  
'har': 13887,  
'has': 44,  
'hat': 2401,  
'hav': 40919,  
'haw': 30595,  
'figtings': 52215,  
'elders': 15495,  
'underpanted': 52216,  
'pninson': 52217,  
'unequivocally': 27652,  
"barbara's": 23673,  
"bello'": 52219,  
'indicative': 12997,  
'yawnfest': 40920,  
'hexploitation': 52220,  
"loder's": 52221,  
'sleuthing': 27653,  
"justin's": 32622,  
"'ball'": 52222,
```

```
"'summer": 52223,  
"'demons": 34935,  
"mormon's": 52225,  
"laughton's": 34737,  
'debell': 52226,  
'shipyard': 39724,  
'unabashedly': 30597,  
'disks': 40401,  
'crowd': 2290,  
'crowe': 10087,  
"vancouver's": 56434,  
'mosques': 34738,  
'crown': 6627,  
'culpas': 52227,  
'crows': 27654,  
'surrell': 53344,  
'flowless': 52229,  
'sheirk': 52230,  
"'three": 40923,  
"peterson": 52231,  
'ooverall': 52232,  
'perchance': 40924,  
'bottom': 1321,  
'chabert': 53363,  
'sneha': 52233,  
'inhuman': 13888,  
'ichii': 52234,  
'ursla': 52235,  
'completly': 30598,  
'moviedom': 40925,  
'raddick': 52236,  
'brundage': 51995,  
'brigades': 40926,  
'starring': 1181,  
"'goal": 52237,  
'caskets': 52238,  
'willcock': 52239,  
"threesome's": 52240,  
"mosque": 52241,  
"cover's": 52242,  
'spaceships': 17637,  
'anomalous': 40927,  
'ptsd': 27655,  
'shirdan': 52243,  
'obscenity': 21962,  
'lemmings': 30599,  
'duccio': 30600,  
"levene's": 52244,  
"'gorby": 52245,  
"teenager's": 25255,  
'marshall': 5340,  
'honeymoon': 9095,  
'shoots': 3231,  
'despised': 12258,  
'okabasho': 52246,  
'fabric': 8289,  
'cannavale': 18515,  
'raped': 3537.
```

```

    'tutt's': 52247,
    'grasping': 17638,
    'despises': 18516,
    'thief's': 40928,
    'rapes': 8926,
    'raper': 52248,
    'eyre': 27656,
    'walchek': 52249,
    'elmo's': 23386,
    'perfumes': 40929,
    'spurting': 21918,
    'exposition\x85': 52250,
    'denoting': 52251,
    'thesaurus': 34740,
    'shoot': 40930,
    'bonejack': 49759,
    'simpsonian': 52253,
    'hebetude': 30601,
    'hallow's': 34741,
    'desperation\x85': 52254,
    'incinerator': 34742,
    'congratulations': 10308,
    'humbled': 52255,
    'else's': 5924,
    'trelovski': 40845,
    'rape': 52256,
    'chapters': 59386,
    '1600s': 52257,
    'martian': 7253,
    'nicest': 25256,
    'eyred': 52259,
    'passenger': 9457,
    'disgrace': 6041,
    'moderne': 52260,
    'barrymore': 5120,
    'yankovich': 52261,
    'moderns': 40931,
    'studlied': 52262,
    'bedsheet': 52263,
    'decapitation': 14900,
    'slurring': 52264,
    'nunsplotation': 52265,
    'character': 34743,
    'cambodia': 9880,
    'rebelious': 52266,
    'pasadena': 27657,
    'crowne': 40932,
    'bedchamber': 52267,
    'conjectural': 52268,
    'appologize': 52269,
    'halfassing': 52270,
    'paycheque': 57816,
    'palms': 20606,
    'islands': 52271,
    'hawked': 40933,
    'palme': 21919,
    'conservatively': 40934,
    '': 40935

```

'larp': 64007,
'palma': 5558,
'smelling': 21920,
'aragorn': 12998,
'hawker': 52272,
'hawkes': 52273,
'explosions': 3975,
'loren': 8059,
"pyle's": 52274,
'shootout': 6704,
"mike's": 18517,
"driscoll's": 52275,
'cogsworth': 40935,
"britian's": 52276,
'childs': 34744,
"portrait's": 52277,
'chain': 3626,
'whoever': 2497,
'puttered': 52278,
'childe': 52279,
'maywether': 52280,
'chair': 3036,
"rance's": 52281,
'machu': 34745,
'ballet': 4517,
'grapples': 34746,
'summerize': 76152,
'freelance': 30603,
"andrea's": 52283,
'\x91very': 52284,
'coolidge': 45879,
'mache': 18518,
'balled': 52285,
'grappled': 40937,
'macha': 18519,
'underlining': 21921,
'macho': 5623,
'oversight': 19507,
'machi': 25257,
'verbally': 11311,
'tenacious': 21922,
'windshields': 40938,
'paychecks': 18557,
'jerk': 3396,
"good'": 11931,
'prancer': 34748,
'prances': 21923,
'olympus': 52286,
'lark': 21924,
'embark': 10785,
'gloomy': 7365,
'jehaan': 52287,
'turaqui': 52288,
"child'": 20607,
'locked': 2894,
'pranced': 52289,
'exact': 2588,
'unattuned': 52290,

'minute': 783,
'skewed': 16118,
'hodgins': 40940,
'skewer': 34749,
'think\x85': 52291,
'rosenstein': 38765,
'helmit': 52292,
'wrestlemanias': 34750,
'hindered': 16826,
"martha's": 30604,
'cheree': 52293,
"pluckin'": 52294,
'ogles': 40941,
'heavyweight': 11932,
'aada': 82190,
'chopping': 11312,
'strongboy': 61534,
'hegemonic': 41342,
'adorns': 40942,
'xxth': 41346,
'nobuhiro': 34751,
'capitães': 52298,
'kavogianni': 52299,
'antwerp': 13422,
'celebrated': 6538,
'roarke': 52300,
'baggins': 40943,
'cheeseburgers': 31270,
'matras': 52301,
"nineties'": 52302,
"'craig'": 52303,
'celebrates': 12999,
'unintentionally': 3383,
'drafted': 14362,
'climby': 52304,
'303': 52305,
'oldies': 18520,
'climbs': 9096,
'honour': 9655,
'plucking': 34752,
'305': 30074,
'address': 5514,
'menjou': 40944,
"'freak'": 42592,
'dwindling': 19508,
'benson': 9458,
'white's': 52307,
'shamelessness': 40945,
'impacted': 21925,
'upatz': 52308,
'cusack': 3840,
"flavia's": 37567,
'effette': 52309,
'influx': 34753,
'booooooooo': 52310,
'dimitrova': 52311,
'houseman': 13423,
'higas': 25259.

```

'boylen': 52312,
'phillipenes': 52313,
'fakery': 40946,
"grandpa's": 27658,
'darnell': 27659,
'undergone': 19509,
'handbags': 52315,
'perished': 21926,
'pooped': 37778,
'vigour': 27660,
'opposed': 3627,
'etude': 52316,
"caine's": 11799,
'doozers': 52317,
'photojournals': 34754,
'perishes': 52318,
'constrains': 34755,
'migenes': 40948,
'consoled': 30605,
'alastair': 16827,
'wvs': 52319,
'ooooooh': 52320,
'approving': 34756,
'consoles': 40949,
'disparagement': 52064,
'futureistic': 52322,
'rebounding': 52323,
"'date": 52324,
'gregoire': 52325,
'rutherford': 21927,
'americanised': 34757,
'novikov': 82196,
'following': 1042,
'munroe': 34758,
"morita'": 52326,
'christenssen': 52327,
'oatmeal': 23106,
'fossey': 25260,
'livered': 40950,
'listens': 13000,
"'marci": 76164,
"otis's": 52330,
'thanking': 23387,
'maude': 16019,
'extensions': 34759,
'ameteurish': 52332,
"commender's": 52333,
'agricultural': 27661,
'convincingly': 4518,
'fueled': 17639,
'mahattan': 54014,
"paris's": 40952,
'vulkan': 52336,
'stapes': 52337,
'odysessy': 52338,
'harmon': 12259,
'surfing': 4252,

```

'halloran': 23494,
'unbelievably': 49580,
"'offed'": 52339,
'quadrant': 30607,
'inhabiting': 19510,
'nebbish': 34760,
'forebears': 40953,
'skirmish': 34761,
'ocassionally': 52340,
"'resist'": 52341,
'impactful': 21928,
'spicier': 52342,
'touristy': 40954,
"'football'": 52343,
'webpage': 40955,
'exurbia': 52345,
'jucier': 52346,
'professors': 14901,
'structuring': 34762,
'jig': 30608,
'overlord': 40956,
'disconnect': 25261,
'sniffle': 82201,
'slimeball': 40957,
'jia': 40958,
'milked': 16828,
'banjoes': 40959,
'jim': 1237,
'workforces': 52348,
'jip': 52349,
'rotweiller': 52350,
'mundaneness': 34763,
"'ninja'": 52351,
"'dead'": 11040,
"'cipriani's'": 40960,
'modestly': 20608,
"'professor'": 52352,
'shacked': 40961,
'bashful': 34764,
'sorter': 23388,
'overpowering': 16120,
'workmanlike': 18521,
'henpecked': 27662,
'sorted': 18522,
"'jõb's'": 52354,
"'always'": 52355,
"'baptists'": 34765,
'dreamcatchers': 52356,
"'silence'": 52357,
'hickory': 21929,
'fun\x97yet': 52358,
'breakumentary': 52359,
'didn': 15496,
'didi': 52360,
'pealing': 52361,
'dispite': 40962,
"'italy's'": 25262,
'instability': 21930,

'quarter': 6539,
'quartet': 12608,
'padmé': 52362,
"'bleedmedry'": 52363,
'pahalniuk': 52364,
'honduras': 52365,
'bursting': 10786,
"pablo's": 41465,
'irremediably': 52367,
'presages': 40963,
'bowlegged': 57832,
'dalip': 65183,
'entering': 6260,
'newsradio': 76172,
'presaged': 54150,
"giallo's": 27663,
'bouyant': 40964,
'amerterish': 52368,
'rajni': 18523,
'leeves': 30610,
'macauley': 34767,
'seriously': 612,
'sugercoma': 52369,
'grimstead': 52370,
"'fairy'": 52371,
'zenda': 30611,
"'twins'": 52372,
'realisation': 17640,
'highsmith': 27664,
'raunchy': 7817,
'incentives': 40965,
'flatson': 52374,
'snooker': 35097,
'crazies': 16829,
'crazier': 14902,
'grandma': 7094,
'napunsaktha': 52375,
'workmanship': 30612,
'reisner': 52376,
"sanford's": 61306,
'\x91doña': 52377,
'modest': 6108,
"everything's": 19153,
'hamer': 40966,
"couldn't'": 52379,
'quibble': 13001,
'socking': 52380,
'tingler': 21931,
'gutman': 52381,
'lachlan': 40967,
'tableaus': 52382,
'headbanger': 52383,
'spoken': 2847,
'cerebrally': 34768,
"'road'": 23490,
'tableaux': 21932,
"proust's": 40968,
'periodical': 40969

```

periodicall : 40300,
"shoveller's": 52385,
'tamara': 25263,
'affords': 17641,
'concert': 3249,
"yara's": 87955,
'someone': 52386,
'lingering': 8424,
"abraham's": 41511,
'beesley': 34769,
'cherbourg': 34770,

```

▼ 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. It can be 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

```

import glob

from sklearn.utils import shuffle
import numpy as np

from keras.preprocessing.text import Tokenizer
from keras.models import Sequential, load_model
from keras.layers import LSTM, Dense, Dropout, Embedding, Masking, Bidirectional
from keras.optimizers import Adam
from keras import backend

model = Sequential()

# Embedding layer
model.add(
    Embedding(
        input_dim=10000,
        output_dim=100,
        weights=None,
        trainable=True))

# Recurrent layer
model.add(
    LSTM(
        64, return_sequences=False, dropout=0.1,
        recurrent_dropout=0.1))

```

```
# Fully connected layer
model.add(Dense(64, activation='relu'))

# Dropout for regularization
model.add(Dropout(0.5))

# Output layer
model.add(Dense(1, activation='softmax'))
# Compile the model
model.compile(
    optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

model.summary()
```

```
⏏ WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_
Instructions for updating:
Please use `rate` instead of `keep_prob`. Rate should be set to `rate = 1 - keep_prob`.
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/optimizers.py:793:
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/tensorflow_core/python/op
Instructions for updating:
Use tf.where in 2.0, which has the same broadcast rule as np.where
Model: "sequential_1"
```

Layer (type)	Output Shape	Param #
=====		
embedding_1 (Embedding)	(None, None, 100)	1000000

lstm_1 (LSTM)	(None, 64)	42240

dense_1 (Dense)	(None, 64)	4160

dropout_1 (Dropout)	(None, 64)	0

dense_2 (Dense)	(None, 1)	65
=====		
Total params: 1,046,465		
Trainable params: 1,046,465		
Non-trainable params: 0		

```
model.fit(x_train, y_train, batch_size=64, epochs=3)
```

```

↳ WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_
Epoch 1/3
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_
25000/25000 [=====] - 238s 10ms/step - loss: 7.9712 - acc: 0.50
Epoch 2/3
25000/25000 [=====] - 231s 9ms/step - loss: 7.9712 - acc: 0.500
Epoch 3/3
25000/25000 [=====] - 232s 9ms/step - loss: 7.9712 - acc: 0.500
<keras.callbacks.History at 0x7fae5c960dd8>

```

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

```

from keras import backend as K
def layer_outs_for_single_instance(model, instance):
    outputs = [layer.output for layer in model.layers]          # all layer outputs
    function = K.function([model.input, K.learning_phase()], outputs) # evaluation function

    if instance.any():
        layer_outs = function([np.array([instance,]), 1.])
        print('_'*75)
        for idx, out in enumerate(layer_outs):
            print('\033[1mLayer# :', idx + 1)
            print('\033[1mName   :', model.layers[idx].name)
            print('\033[1mShape  : ' + str(out.shape) + '\033[0m')
            print(out)
            print('_'*75)

layer_outs_for_single_instance(model, x_test[np.random.choice(len(y_test))])

```

↳

Layer# : 1**Name : embedding_1****Shape : (1, 300, 100)**

```
[[[-0.01472874  0.03862207 -0.00789142 ...  0.0353833  -0.02726978
   -0.01185354]
 [-0.01472874  0.03862207 -0.00789142 ...  0.0353833  -0.02726978
   -0.01185354]
 [-0.01472874  0.03862207 -0.00789142 ...  0.0353833  -0.02726978
   -0.01185354]
 ...
 [ 0.02734203  0.0395744   0.00077729 ... -0.02001605  0.01918605
   0.01954628]
 [-0.03852933  0.02782103 -0.00494694 ...  0.02900696  0.04584566
   0.0492037 ]
 [ 0.02756107  0.00461642 -0.0453699   ... -0.0076103  -0.02313288
  -0.00421534]]]
```

Layer# : 2**Name : lstm_1****Shape : (1, 64)**

```
[[ 9.45082866e-04  3.36543960e-03  1.78135298e-02 -7.86926318e-03
  1.98247680e-03  3.05092847e-03  9.42264870e-03 -1.24558862e-02
 -5.21513913e-03  1.89300291e-02  7.33488286e-03  7.11964583e-03
  8.43445479e-04 -8.40545818e-03  1.68919731e-02  1.72298793e-02
 -6.04572333e-03 -1.50117436e-02  1.58114098e-02 -2.97207414e-04
  3.01633903e-04  1.52029619e-02 -1.85263238e-03 -5.72952768e-03
 -3.44699714e-03  1.46512818e-02 -1.17764238e-03  4.85222973e-03
  6.44007232e-04  3.62510444e-04  5.89778088e-03 -2.07916298e-03
 -1.39600942e-02 -1.00411510e-03 -5.26397629e-03  1.07000293e-02
  9.30486806e-03 -4.77149663e-03 -3.57207656e-03 -1.03053236e-02
 -1.75238978e-02 -1.75218924e-03  1.37255462e-02  4.74871881e-03
 -6.51200302e-03 -3.38529097e-03 -1.10049695e-02  5.61200036e-03
 -3.39799553e-05 -7.39507983e-03 -1.15060983e-02  5.37597435e-03
 -5.02444617e-03  1.80633401e-03 -1.27596650e-02  2.13552325e-04
  8.29503592e-03 -1.05528021e-02  5.51900477e-04 -6.61057932e-03
  1.37532232e-02  2.45877588e-03  8.87858495e-03 -1.44654121e-02]]]
```

Layer# : 3**Name : dense_1****Shape : (1, 64)**

```
[[8.9872978e-05 1.5477960e-03 0.0000000e+00 1.4021548e-02 0.0000000e+00
 0.0000000e+00 3.4300541e-04 1.0132790e-02 1.2591403e-02 0.0000000e+00
 6.3492097e-03 0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
 0.0000000e+00 3.9210822e-03 0.0000000e+00 2.2311523e-03 0.0000000e+00
 1.9005340e-02 1.5973752e-02 2.0752603e-02 0.0000000e+00 0.0000000e+00
 0.0000000e+00 3.8883255e-03 0.0000000e+00 0.0000000e+00 6.9625485e-03
 0.0000000e+00 4.5670718e-03 0.0000000e+00 0.0000000e+00 1.1287245e-03
 0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
 6.7499820e-03 0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
 2.5882502e-03 3.5217353e-03 1.5466179e-02 0.0000000e+00 7.8768758e-03
 0.0000000e+00 0.0000000e+00 7.2057080e-03 1.5723193e-04 3.6621983e-03
 0.0000000e+00 1.7753971e-03 1.1343910e-02 0.0000000e+00 0.0000000e+00
 0.0000000e+00 3.1806063e-03 2.5454632e-03 0.0000000e+00]]]
```

Layer# : 4**Name : dropout_1**

Shape : (1, 64)

```
[[0.00017975 0.00309559 0.          0.0280431  0.          0.
  0.00068601 0.          0.          0.          0.          0.
  0.          0.          0.          0.          0.          0.
  0.0044623  0.          0.          0.0319475  0.04150521 0.
  0.          0.          0.          0.          0.          0.
  0.          0.          0.          0.          0.          0.
  0.          0.          0.          0.          0.          0.
  0.          0.          0.          0.0051765  0.          0.
  0.          0.          0.          0.          0.          0.00031446
  0.0073244  0.          0.          0.          0.          0.
  0.          0.00636121 0.00509093 0.          0.          0.]]
```

Layer# : 5

Name : dense_2

Shape : (1, 1)

[[1.]]