In [1]:
```
import numpy as np
import pandas as pd
from word2vec import *
from nltk.tokenize import sent_tokenize, word_tokenize
import warnings
import nltk
import gensim
import torch
```

In [16]:
```
# !mkdir GloVe
# !curl -Lo GloVe/glove.840B.300d.zip http://nlp.stanford.edu/data/glove.840B.300d.zip
# !unzip GloVe/glove.840B.300d.zip -d GloVe/
!mkdir fastText
!curl -Lo fastText/crawl-300d-2M.vec.zip https://dl.fbaipublicfiles.com/fasttext/vectors-english/cra
!unzip fastText/crawl-300d-2M.vec.zip -d fastText/
```

```
mkdir: fastText: File exists
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100 1453M  100 1453M    0     0  3851k      0  0:06:26  0:06:26 --:--:-- 3852kM    0     0  3829k
0  0:06:28  0:00:17  0:06:11 3836k    0  3854k      0  0:06:26  0:00:24  0:06:02 3940k  0     0  3873k
0  0:06:24  0:00:54  0:05:30 3800k 0     0  3881k      0  0:06:23  0:01:06  0:05:17 3811k 0  3891k
0  0:06:22  0:01:16  0:05:06 3946k2  0:01:26  0:04:56 3989k     0  0:06:21  0:01:53  0:04:28 3942kk
0  0:06:21  0:03:17  0:03:04 4009k 0  3900k      0  0:06:21  0:03:32  0:02:49 3827k 0     0  3860k
0  0:06:25  0:04:18  0:02:07 1942k0  0:06:25  0:04:20  0:02:05 2781k   0     0  3863k      0  0:06:25
0:04:22  0:02:03 3815k    0  3861k      0  0:06:25  0:04:26  0:01:59 3798k   0     0  3862k      0  0:
06:25  0:04:30  0:01:55 3808k  0:06:24  0:05:41  0:00:43 3858k     0  0:06:25  0:05:54  0:00:31 3730k
0     0  3855k      0  0:06:25  0:06:08  0:00:17 3417k   0  0:06:26  0:06:10  0:00:16 3455k
Archive:  fastText/crawl-300d-2M.vec.zip
  inflating: fastText/crawl-300d-2M.vec
```

In [31]:
```
!mkdir encoder
!curl -Lo encoder/infersent1.pkl https://dl.fbaipublicfiles.com/infersent/infersent1.pkl
!curl -Lo encoder/infersent2.pkl https://dl.fbaipublicfiles.com/infersent/infersent2.pkl
```

```
mkdir: encoder: File exists
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100  146M  100  146M    0     0  3827k      0  0:00:39  0:00:39 --:--:-- 3822k    0  0:00:40  0:00:15
0:00:25 3715k0  3823k      0  0:00:39  0:00:38  0:00:01 3820k
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100  146M  100  146M    0     0  3809k      0  0:00:39  0:00:39 --:--:-- 3866k 0  0:00:39  0:00:13  0:
00:26 3771k 0  3796k      0  0:00:39  0:00:31  0:00:08 3737k
```

In [33]:
```
from models import InferSent
V = 2
MODEL_PATH = 'encoder/infersent%s.pkl' % V
params_model = {'bsize': 64, 'word_emb_dim': 300, 'enc_lstm_dim': 2048,
                'pool_type': 'max', 'dpout_model': 0.0, 'version': V}
infersent = InferSent(params_model)
infersent.load_state_dict(torch.load(MODEL_PATH))

W2V_PATH = 'fastText/crawl-300d-2M.vec'
infersent.set_w2v_path(W2V_PATH)
```

```
---------------------------------------------------------------------------
ModuleNotFoundError                       Traceback (most recent call last)
<ipython-input-33-884bcccc8c62> in <module>
----> 1 from model import InferSent
      2 V = 2
      3 MODEL_PATH = 'encoder/infersent%s.pkl' % V
      4 params_model = {'bsize': 64, 'word_emb_dim': 300, 'enc_lstm_dim': 2048,
      5                 'pool_type': 'max', 'dpout_model': 0.0, 'version': V}

ModuleNotFoundError: No module named 'model'
```

In [19]:
```
df = pd.read_csv('/Users/nehakardam/Documents/UWclasses /EE517 NLP/Project/FacultySupport_SA1_June2_
sentences = df[df.SA1.notnull()].SA1.tolist()
sentences
```

Out[19]:
```
['Restructure quizzes and stuff. In 235 we had a weekly quiz in lieu of midterms and a final, and that helped keep people engaged and paying attention.',
 'Flexible late turn in policies, especially in my area my wifi is very inconsistent so it would be very nice if professors could be aware and understanding of that.',
 ' Leniency on deadlines. It can be hard to stay motivated during these times without dedicated study spaces. Some have inconsistent internet connections.',
 'be flexible to possible changes and take student feedback into account to change/adj0t assignments/pacing',
 'have some strict action to make sure every student will go to class. For example, count every student during each class and find if anyone is absent, give them punishment if they do not go to class j0t beca0e being lazy',
 'Post some examples on the powerpoint. Offer some chances to make up quiz point.',
 'Bi-weekly or weekly quizzes would allow for some flexibility and still ensure that students stay on top of the material. ',
 'Even though classes are being held online, I feel professors should promote active participation during class, by requiring webcams and attendance.',
 "Provide options to earn back exam points; the new system of test taking has proven incredibly challenging for me, and I've seen my test scores plummet while other students seem to have higher averages than normal. I feel like a good way to balance students who may have not been following the test rules
```

In [20]:
```
infersent.build_vocab(sentences, tokenize=True)
embeddings = infersent.encode(sentences, tokenize=True)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-20-9aa3b287e154> in <module>
----> 1 infersent.build_vocab(sentences, tokenize=True)
      2 embeddings = infersent.encode(sentences, tokenize=True)

NameError: name 'infersent' is not defined
```

In [21]:
```
1 infersent.visualize('Post some examples on the powerpoint. Offer some chances to make up quiz point.
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-21-629ff1f7d51e> in <module>
----> 1 infersent.visualize('Post some examples on the powerpoint. Offer some chances to make up quiz
 point.', tokenize=True)

NameError: name 'infersent' is not defined
```

In [22]:
```
1 !pip install nmslib
```

```
Requirement already satisfied: nmslib in /opt/anaconda3/lib/python3.8/site-packages (2.1.1)
Requirement already satisfied: pybind11<2.6.2 in /opt/anaconda3/lib/python3.8/site-packages (from nmsl
ib) (2.6.1)
Requirement already satisfied: psutil in /opt/anaconda3/lib/python3.8/site-packages (from nmslib) (5.
7.2)
Requirement already satisfied: numpy>=1.10.0 in /opt/anaconda3/lib/python3.8/site-packages (from nmsli
b) (1.19.5)
```

In [23]:
```
import nmslib

NTHREADS = 8
def create_index(a):
    index = nmslib.init(space='angulardist')
    index.addDataPointBatch(a)
    index.createIndex()
    return index
def get_knns(index, vecs, k):
    return zip(*index.knnQueryBatch(vecs, k=k,num_threads=NTHREADS))

nn_wvs = create_index(embeddings)
to_frame = lambda x: pd.DataFrame(np.array(x)[:,1:])
idxs, dists = map(to_frame, get_knns(nn_wvs, embeddings, k=10))
df = pd.concat([idxs.stack().to_frame('idx'), dists.stack().to_frame('dist')], axis=1).reset_index()
ndf = catted[["v1", "v2"]].to_numpy()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-23-ff01c268c1cb> in <module>
     10     return zip(*index.knnQueryBatch(vecs, k=k,num_threads=NTHREADS))
     11
---> 12 nn_wvs = create_index(embeddings)
     13 to_frame = lambda x: pd.DataFrame(np.array(x)[:,1:])
     14 idxs, dists = map(to_frame, get_knns(nn_wvs, embeddings, k=10))

NameError: name 'embeddings' is not defined
```

In [24]:
```
ndf
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-24-7fc6f30b0073> in <module>
----> 1 ndf

NameError: name 'ndf' is not defined
```

In [25]:
```
from sklearn.cluster import AgglomerativeClustering
clustering = AgglomerativeClustering(n_clusters=6).fit(ndf)
clustering.labels_
plot_scatter(ndf, clustering.labels_)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-25-58fb1ae7ce61> in <module>
      1 from sklearn.cluster import AgglomerativeClustering
----> 2 clustering = AgglomerativeClustering(n_clusters=6).fit(ndf)
      3 clustering.labels_
      4 plot_scatter(ndf, clustering.labels_)

NameError: name 'ndf' is not defined
```

In [26]:
```
df['labels'] = clustering.labels_
df
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-26-7681de5098ce> in <module>
----> 1 df['labels'] = clustering.labels_
      2 df

NameError: name 'clustering' is not defined
```

In [102]:
```
def plot_scatter(X,  color, alpha=0.5):
    return plt.scatter(X[:, 0],
                       X[:, 1],
                       c=color,
                       alpha=alpha,
                       edgecolor='k')
```

In [ ]:
```

```