Homework 4
CS 259 Numerical Methods for Data Science
Prof. **David Bindel**
TA. **Yurong You, Xinran Zhu**

Hongyu Yan (516030910595)
ACM Class, Zhiyuan College, SJTU
Due Date: June 26th, 2018
Submit Date: July 19, 2018

Problem 1

```
>> main
idc =  1729
  chip (1.43e-01)  neuro (9.12e-02)  canceling (6.97e-02)  analog (4.22e-02)  analogue (4.14e-02)  synapse (3.26e-02)  board (3.14e-02)  loud (2.96e-02)  noi
se (2.30e-02)  active (2.14e-02)
  wta (9.18e-02)  voltage (6.27e-02)  chip (5.86e-02)  offset (4.47e-02)  cmos (4.16e-02)  adaptable (3.78e-02)  winner (3.23e-02)  analog (3.16e-02)  vin (3
.15e-02)  dynamically (2.63e-02)
  circuit (1.19e-01)  winner (1.13e-01)  transistor (4.96e-02)  lazzaro (4.19e-02)  neuron (3.98e-02)  mead (3.03e-02)  mahowald (2.95e-02)  ryckebusch (2.78
e-02)  resistor (2.60e-02)  losing (2.50e-02)
  pulse (1.05e-01)  murray (6.69e-02)  stream (6.34e-02)  synapse (6.31e-02)  vos (5.27e-02)  transistor (4.47e-02)  transconductance (3.63e-02)  mosfet (3.5
4e-02)  tarassenko (3.49e-02)  chip (3.32e-02)
  circuit (1.21e-01)  adaptation (4.78e-02)  intensity (4.48e-02)  delbrfick (4.41e-02)  laminar (4.41e-02)  background (4.08e-02)  photocurrent (3.60e-02)
delbriick (3.56e-02)  transistor (3.47e-02)  conductance (3.11e-02)
>>
```

Figure 1: Result

The result shows that the most relevant articles may not have ”circuit” as top words. It seems that ”circuit” may refer to biological neural network, which means the documents’ top words may contain ”synapse”, ”neuro” etc. It may also refer to electronics related themes, which means the documents’ top words may be ”chip”, ”pulse” etc.

Here is my code.

```
1
2  [W, vocab] = load_docword('.', 'nips');
3  [WW, Dtf, Didf] = tf_idf(W);
4
5
6  [U, S, V] = svds(WW, 20);
7  new_WW = U * S * V';
8
9  idc = 0;
10 for i = 1:length(vocab)
11   if strcmp(vocab{i}, 'circuit')
12     idc = i
13   end
14 end
15
16 score = new_WW(:, idc);
17 [v, idx] = sort(score, 'descend');
18
19 num_doc = 5;
20 num_word = 10;
21 for i = 1:num_doc
22   show_top_words(WW(idx(i), :), vocab, num_word);
23 end
```