```
%load data
load 'dataset'/wordVecV.mat
```

```
%compute M M is a matrix of 1s and 0s indicating if such terms exist
%indepent of frequency
M = double(V~=0);
M_t = M;

%compute M~
for i=1:size(M,2)
    M_t(:,i) = M(:,i)/norm(M(:,i));
end
```

c) Compute svd of M~

```
[U,S,V] = svd(M_t);
s = svd(M_t)
```

```
s = 10×1
    1.5366
    1.0192
    0.9587
    0.9539
    0.9413
    0.9289
    0.8977
    0.8919
    0.8687
    0.8161
```

d) Assume k = 9

```
k = 9;
distances = zeros(10);

for i = 1:size(M_t, 2)-1
    vec_1 = M_t(:,i);
    for e = i+1:size(M_t,2)
        vec_2 = M_t(:,e);
        distances(i,e) = acos((vec_1(1:k)'*vec_2(1:k))/(norm(vec_1(1:k))*norm(vec_2(1:
    end
end

min_distance = min(distances(distances>0));
[doc1, doc2] = find(distances == min_distance);
doc1
```

```
doc1 = 2
```

```
doc2
```

```
doc2 = 8
```

e) Repeat for lower k. Find lowest k that doesn't change answer. Find documents for k - 1.

```
k = 8;
distances = zeros(10);

for i = 1:size(M_t, 2)-1
    vec_1 = M_t(:,i);
    for e = i+1:size(M_t,2)
        vec_2 = M_t(:,e);
        distances(i,e) = acos((vec_1(1:k)'*vec_2(1:k))/(norm(vec_1(1:k))*norm(vec_2(1:
    end
end

min_distance = min(distances(distances>0));
[doc1, doc2] = find(distances == min_distance);
doc1
```

```
doc1 = 2
```

```
doc2
```

```
doc2 = 8
```

```

k = 7;
distances = zeros(10);

for i = 1:size(M_t, 2)-1
    vec_1 = M_t(:,i);
    for e = i+1:size(M_t,2)
        vec_2 = M_t(:,e);
        distances(i,e) = acos((vec_1(1:k)'*vec_2(1:k))/(norm(vec_1(1:k))*norm(vec_2(1:
    end
end

min_distance = min(distances(distances>0));
[doc1, doc2] = find(distances == min_distance);
doc1
```

```
doc1 = 2
```

```
doc2
```

```
doc2 = 8
```

```

k = 6;
distances = zeros(10);

for i = 1:size(M_t, 2)-1
    vec_1 = M_t(:,i);
    for e = i+1:size(M_t,2)
        vec_2 = M_t(:,e);
        distances(i,e) = acos((vec_1(1:k)'*vec_2(1:k))/(norm(vec_1(1:k))*norm(vec_2(1:
    end
```

```
end

min_distance = min(distances(distances>0));
[doc1, doc2] = find(distances == min_distance);
doc1
```

```
doc1 = 2
```

```
doc2
```

```
doc2 = 8
```

```

k = 5;
distances = zeros(10);

for i = 1:size(M_t, 2)-1
    vec_1 = M_t(:,i);
    for e = i+1:size(M_t,2)
        vec_2 = M_t(:,e);
        distances(i,e) = acos((vec_1(1:k)'*vec_2(1:k))/(norm(vec_1(1:k))*norm(vec_2(1:
    end
end

min_distance = min(distances(distances>0));
[doc1, doc2] = find(distances == min_distance);
doc1
```

```
doc1 = 2
```

```
doc2
```

```
doc2 = 8
```

```

k = 4;
distances = zeros(10);

for i = 1:size(M_t, 2)-1
    vec_1 = M_t(:,i);
    for e = i+1:size(M_t,2)
        vec_2 = M_t(:,e);
        distances(i,e) = acos((vec_1(1:k)'*vec_2(1:k))/(norm(vec_1(1:k))*norm(vec_2(1:
    end
end

min_distance = min(distances(distances>0));
[doc1, doc2] = find(distances == min_distance);
doc1
```

```
doc1 = 2
```

```
doc2
```

```
doc2 = 5
```

```
disp('k=5 is the lowest that keeps the same answer. k=4 gives documents 2 and 5.')
```