

Problem1

$$\begin{aligned}
 1. (a) \quad & P(y=1) = \frac{4}{10} = \frac{2}{5} \\
 & P(y=-1) = \frac{6}{10} = \frac{3}{5} \\
 \checkmark \quad & P(x_1=1|y=1) = \frac{2}{4} \quad P(x_1=1|y=-1) = \frac{2}{5} \\
 & P(x_1=0|y=1) = \frac{1}{4} \quad P(x_1=0|y=-1) = \frac{2}{5} \\
 \checkmark \quad & P(x_2=1|y=1) = 0 \quad P(x_2=1|y=-1) = \frac{5}{6} \\
 & P(x_2=0|y=1) = 1 \quad P(x_2=0|y=-1) = \frac{1}{6} \\
 \checkmark \quad & P(x_3=1|y=1) = \frac{3}{4} \quad P(x_3=1|y=-1) = \frac{2}{3} \\
 & P(x_3=0|y=1) = \frac{1}{4} \quad P(x_3=0|y=-1) = \frac{1}{3} \\
 \checkmark \quad & P(x_4=1|y=1) = \frac{1}{2} \quad P(x_4=1|y=-1) = \frac{5}{6} \\
 & P(x_4=0|y=1) = \frac{1}{2} \quad P(x_4=0|y=-1) = \frac{1}{6} \\
 \checkmark \quad & P(x_5=1|y=1) = \frac{1}{4} \quad P(x_5=1|y=-1) = \frac{1}{3} \\
 \checkmark \quad & P(x_5=0|y=1) = \frac{3}{4} \quad P(x_5=0|y=-1) = \frac{2}{3}
 \end{aligned}$$

$$\begin{aligned}
 (b) \quad & P(y=1|x=(0,0,0,0,0)) = \frac{P(x|y=1) \cdot P(y=1)}{P(x)} \\
 & = \frac{P(x|y=1) \cdot P(y=1)}{P(y=1) \cdot P(x|y=1) + P(y=-1) \cdot P(x|y=-1)} \\
 & = \frac{(\frac{1}{4} \cdot 1 \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{3}{4}) \cdot \frac{2}{5}}{\frac{\frac{3}{320} + \frac{1}{675}}{\frac{2}{5} + \frac{1}{3}}} = \frac{\frac{3}{320}}{\frac{2}{5} + \frac{1}{3}} \\
 & P(y=-1|x=(0,0,0,0,0)) = \frac{P(x|y=-1) \cdot P(y=-1)}{P(y=1) \cdot P(x|y=1) + P(y=-1) \cdot P(x|y=-1)} \\
 & = \frac{(\frac{2}{5} \cdot \frac{1}{6} \cdot \frac{1}{3} \cdot \frac{1}{6} \cdot \frac{2}{3}) \cdot \frac{3}{5}}{\frac{\frac{3}{320} + \frac{1}{675}}{\frac{2}{5} + \frac{1}{3}}} = \frac{\frac{1}{675}}{\frac{2}{5} + \frac{1}{3}}
 \end{aligned}$$

$$\therefore P(y=1|x=(0,0,0,0,0)) > P(y=-1|x=(0,0,0,0,0))$$

\therefore predict $y=+1$

$$P(y=1 | x=(11010)) = \frac{P(x|y=1) \cdot P(y=1)}{P(x|y=1) \cdot P(y=1) + P(x|y=-1) \cdot P(y=-1)}$$

$$= \frac{\frac{3}{4} \cdot 0 \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{2}{4}}{\text{const.}} = 0$$

$$P(y=-1 | x=(11010)) = \frac{P(x|y=-1) \cdot P(y=-1)}{P(x|y=-1) \cdot P(y=-1) + P(x|y=1) \cdot P(y=1)}$$

$$= \frac{\frac{2}{5} \cdot \frac{5}{6} \cdot \frac{1}{3} \cdot \frac{5}{6} \cdot \frac{2}{3}}{0 + \frac{2}{5} \cdot \frac{5}{6} \cdot \frac{1}{3} \cdot \frac{5}{6} \cdot \frac{2}{3}} = 1$$

∴ predict $y=-1$

c. $P(y=1 | x=(11010)) = 0$. see above

d. A naive Bayes classifier is based on strong independence assumption. It assumes that the presence of a particular feature of a class is unrelated to the presence of any other feature, given the class variable.

But Bayes classifier is used in dependence variables.

For this problem, the features are independence.

Additionally, you need more parameters for Bayes classifier than naive Bayes.

For example, there are three variables.

Bayes		Naive Bayes		
	$P(x y=1)$	$P(x_1 y=1)$	$P(x_2 y=1)$	$P(x_3 y=1)$
000	α_1	0	β_1	β_2
001	α_2	1	$1-\beta_1$	$1-\beta_2$
010	α_3	It just needs 3 parameters		
1	!			
!	!			
111	$1-\alpha_1-\alpha_2-\alpha_3$			

It needs 2^3-1 parameters.

So, we should use naive Bayes for this problem.

Problem2

```
a. function [z,c,sumd] = kmeans(X,k,clnit)
% [assign, clusters, sumd] = kmeans(X, k [,initialClusters])
% Perform k-means clustering on data X (# data x # features), initializing randomly or to a set of k
clusters
% Returns:
% assign (# data x 1) index of cluster
% clusters (k x # features) cluster centers
% sumd (scalar) sum of squared euclidean distances
% Returns after 100 iterations or when converged

maxIter = 100;

[N,D] = size(X);
if (nargin < 3)
    pi = randperm(N);
    c = X(pi(1:k),:);
else
    c = clnit;
    if (isempty(k)) k=size(clnit,1); end;
end;

z = zeros(N,1);
done = false;
iter = 1;
sumdOld = inf;

while (~done)
    sumd = 0;
    for i=1:N,
        dists = sum((c - repmat(X(i,:),[k,1])).^2 , 2); % compute distances from each cluster center
        [tmp,z(i)] = min(dists); % and assign datum i to the nearest cluster
        sumd = sumd + tmp;
    end
    for j=1:k, % now update each cluster center j
        c(j,:) = mean(X(z==j,:),1); % to be the mean of the assigned data
    end
    done = (iter >= maxIter) || (sumd == sumdOld);
    sumdOld = sumd;
    iter = iter+1;
    fprintf('.');
end;

if (iter >= maxIter) warning('kmeans: stopped after reaching maximum number of iterations'); end;
```

```
[vocab] = textread('vocab.txt', '%s');  
[did, wid, cnt] = textread('docword.txt', '%d%d%d', 'headerlines', 3);
```

```
X = sparse(did, wid, cnt);  
D = max(did);  
W = max(wid);  
N = sum(cnt);
```

```
Xn = X./repmat(sum(X,2), [1, W]);
```

```
[z, c, sumd] = kmeans(Xn, 20)
```

```
sumd =
```

```
2.4669
```

b.

1. EDU>> [z1,c1,sumd1] = kmeans(Xn,20)

sumd1 =

2.1840

2. EDU>> [z2,c2,sumd2] = kmeans(Xn,20)

sumd2 =

2.1622

3. EDU>> [z3,c3,sumd3] = kmeans(Xn,20)

sumd3 =

2.2410

4. EDU>> [z4,c4,sumd4] = kmeans(Xn,20)

sumd4 =

2.1910

5. EDU>> [z5,c5,sumd5] = kmeans(Xn,20)

sumd5 =

2.5209

6. EDU>> [z6,c6,sumd6] = kmeans(Xn,20)

sumd6 =

2.5434

7. EDU>> [z7,c7,sumd7] = kmeans(Xn,20)

sumd7 =

2.2300

The cost function of 2nd run is smallest and so we pick up [z2,c2,sumd2] as the best one and use for the rest of the problem.

c.

```
z = zeros(20,1);
```

```
[r c] = size(z2);
```

```
for i=1:r
```

```
    n = z2(i);
```

```
    z(n) = z(n)+1;
```

```
end
```

```
EDU>> z
```

```
z =
```

```
2    (2 articles are associated with 1st cluster)
1    (1 article is associated with 2nd cluster)
2    (2 articles are associated with 3rd cluster)
2    (2 articles are associated with 4th cluster)
4    (4 articles are associated with 5th cluster)
2    (2 articles are associated with 6th cluster)
2    (2 articles are associated with 7th cluster)
4    (4 articles are associated with 8th cluster)
1    (1 article is associated with 9st cluster)
50   (50 articles are associated with 10th cluster)
11   (11 articles are associated with 11th cluster)
1    (1 article is associated with 12st cluster)
12   (12 articles are associated with 12th cluster)
2    (2 articles are associated with 14rd cluster)
32   (32 articles are associated with 15th cluster)
5    (5 articles are associated with 16th cluster)
18   (18 articles are associated with 17th cluster)
4    (4 articles are associated with 18th cluster)
16   (16 articles are associated with 19th cluster)
57   (57 articles are associated with 20th cluster)
```



```

[r c] = size(c2);
for i=1:r
    [sorted,order] = sort(c2(i,:),2,'descend');
    fprintf('Cluster %d\n', i);
    fprintf('%s ',vocab{order(1:10)});
    fprintf('\n');
end

```

Cluster 1 (Sport)
season rusie league giants seasons games game 1900 times won

Cluster 2 (Women)
plummer dictionary savannah school lady book community daughter 000 began

Cluster 3
nyt news cox times options bos stock service y2k undated

Cluster 4 (Criminal news)
drug marijuana drugs nadelmann policy criminal director foundation group americans

Cluster 5 (military)
algeria islamic pakistan government hijackers india army country front war

Cluster 6 (Criminal news)
hijackers hostages burger told passengers indian killed mrs plane finally

Cluster 7 (Family)
home stay fathers children group child kids working called convention

Cluster 8 (European news)
europe europeans european french immigration language american countries labor model

Cluster 9 (Politics - election)
spending budget candidates cuts tax health programs congress cut domestic

Cluster 10 (Sport)
game team players games season coach play league sports bowl

Cluster 11 (Sport)
season team coach million players patriots think free giants going

Cluster 12 (Test statistics)
test end houston 000 0101 0102 100 1900 1900s 1950s

Cluster 13 (Politics - election)
bradley candidates campaign mccain hampshire republican bush political party candidate

Cluster 14 (European news)
france lee called company coup africa coast damage europe giuliani

Cluster 15 (Politics - Africa)
city century war rwanda tutsi group book hutu week million

Cluster 16 (Politics - Celebration)
millennium economy island fireworks century cuba global trade political midnight

Cluster 17 (Politics - Russia)
putin yeltsin russia president russian power government country additional political

Cluster 18 (Young people)
boot camps juvenile camp tough programs justice townsend crime maryland

Cluster 19 (Technology)
y2k 2000 problem saturday system computer problems koskinen computers systems

Cluster 20 (Celebration)
times square city millennium 2000 000 night midnight york fireworks

d.

1. For the cluster associated with document 1

```
m = 1;
```

```
[r c] = size(z2);
```

```
cluster = z2(m);
```

```
n = 1;
```

```
for i=1:r
```

```
    if(cluster==z2(i))
```

```
        fprintf('%d\n',n);
```

```
        fname = sprintf('example1/20000101.%04d.txt',i);
```

```
        txt = textread(fname,'%s',10,'whitespace','\r\n');
```

```
        fprintf('%s\n',txt{:});
```

```
        n = n+1;
```

```
    if(n>12)
```

```
        break;
```

```
    end
```

```
end
```

```
end
```

cluster: 10

1 (Sport)

School has been out at Cal State Northridge since

the week before Christmas, but since you can learn something every

day, Mississippi State's women's basketball team gave a lesson.

Northridge has talked about taking its game to the next level.

The 21st-ranked Bulldogs _ the first nationally ranked team to play

here in Northridge's Division I era _ gave a glimpse of that level

in a 98-64 nonconference victory before a crowd of 165 Friday

night.

It ended CSUN's 16-game home win streak.

Whether the Matadors learned from the Bulldogs (10-0) remains to

2 (Business - industry)

A cottage industry of plain-language consultants has sprung up

in recent years. They aim to wage war on the business buzzwords,

governmental verbosity and turgid legalese that make many public documents impossible to understand.

These experts also seek to instill some of the basics of good writing, like the importance of short, direct sentences. Much of their work has come in the mutual fund industry, which for the past two years has been under an order by the Securities and Exchange Commission to simplify documents.

Samples of the consultants' work follow.

3 (Business - industry)

A cottage industry of plain-language consultants has sprung up in recent years. They aim to wage war on the business buzzwords, governmental verbosity and turgid legalese that make many public documents impossible to understand.

These experts also seek to instill some of the basics of good writing, like the importance of short, direct sentences. Much of their work has come in the mutual fund industry, which for the past two years has been under an order by the Securities and Exchange Commission to simplify documents.

Samples of the consultants' work follow.

4 (Sport)

Eddie Jones has rejoined the Charlotte Hornets and reaffirmed his plans for a quick comeback from an elbow injury.

Jones met with Hornets team physician Glenn Perry and was given a brace to wear until the torn ligament he suffered after being fouled by the New York Knicks' Kurt Thomas Dec. 20 heals.

“To be honest with you, I don't want to wear it in the games,”

Jones said. “But I might have to.”

Jones said as soon as the injury was diagnosed that he planned on returning in three weeks, not the longest-range prospect of six weeks. “Like I told people, it wasn't going to be any longer than

5 (Sport)

Just four months ago, only those living in the small farming town tucked between the Blue Ridge Mountains and Allegheny Mountains in Southwestern Virginia had any idea about what the nation was about to witness.

Now, as the national championship game approaches, quarterback Michael Vick, the steady, sturdy, leader of the second-ranked Virginia Tech Hokies, is a household name in more locales than Blacksburg.

Vick needed one season to lead the Hokies to an unbeaten record and a date with Florida State in Tuesday's Sugar Bowl. He did so by

6 (Sport)

Nobody is quite ready to do celebratory swan dives into the Detroit River just yet. Regular-season victories in Hockeytown do little except to affirm to fans that forward lines are set for the playoffs.

But as the NHL season nears the midway point, the Detroit Red Wings look closer to being a Stanley Cup team now than they did last year.

The Wings have the best record in hockey, and they continued to

string together wins despite the absences of goalie Chris Osgood and forward Slava Kozlov. After a narrow 3-2 overtime win over the

7 (Sport)

FACEOFF: BRETT HULL

Right winger, Dallas Stars

Update: Entering the weekend looking for career goal No. 600, a mark reached by only 11 other players. Second on the Stars this year with 13 goals and 28 points in 37 games.

On approaching the 610 goals scored by his father, Bobby: ``To be able to have followed in his footsteps has been a thrill for me more than anything, and to have been able to reach the level that he played, being a Hall of Famer and voted the eighth-best player in the Top 50 list that came out last year, I mean, it is a huge

8 (Sport - entertainment)

Ben Wallace's hairstyle has become one of the big sports stories in Orlando.

Wallace likes corn rows because they are nice and neat, but a poll of Magic fans show that 62 percent want him to keep the retro big Afro he has worn lately. Here's why:

The Magic won seven of nine games after Wallace let his hair free in early December. That they dropped two games last week by a combined five points might indicate that the charm has gone.

But does Wallace have the courage to change and perhaps ``doom" his team to mediocrity? Orlando was 8-9 in corn-row games.

9 (Sport)

Grant Long, whose loss to Vancouver left the Hawks without a rudder, finally is playing after suffering a concussion and a cartilage injury. In a 100-93 loss to Philadelphia, Long had eight rebounds in 22 minutes. He talked about his excitement at playing again.

“I don't know if you guys watch major league baseball a lot,” Long said, “but Sammy Sosa is one of my favorite players, and he takes off from the dugout and he tears out to center field. He's just happy to be out on the field playing major league baseball, and that's my appreciation. When I'm out there on the floor, I'm

10 (Sport)

The NFL opens the 21st century today by taking a big step backward.

The final Sunday of the season finds four 7-8 teams trying to get into the playoffs. If Detroit loses at Minnesota, it will be the first season in which two 8-8 teams advanced to the postseason.

In the NFC, Dallas, Carolina, Green Bay and the New York Giants will be doing some scoreboard watching.

In early games, Green Bay hosts Arizona and Carolina hosts New Orleans. If the Packers and Panthers win, they'll be rooting for the Giants.

11 (Sport)

Here we are, in the year 2000, and Don Casey is the coach of a red-hot team named the New Jersey Nets. Corporate synergy triumphs again.

Did you know that Casey's team is 10-3 since his new organizational ally Joe Torre motored over the George Washington Bridge to spread Yankees stardust over the then-bedraggled, 2-15 Nets? Suddenly, tantalizingly, the Nets are no longer stringing together defeats, like an order of Spaghetti Calipari.

Stephon Marbury turned into Derek Jeter. Keith Van Horn morphed into Bernie Williams. The Nets have been sharing, bonding, winning.

12 (Sport)

It was not until after the Players' Riot of 2019, in which hundreds of fans were injured and two died, that America began to take seriously the possibility that sports was no national campfire but rather a dangerously out-of-control brush fire that could consume the entire country. The president's hastily assembled Rocker Commission summed up the situation in this plaintive statement: ``Twenty years ago, a grown man in a New York Rangers jersey would be an object of mild amusement. Ten years ago, he would be questioned by mental health authorities. Today, he would be detained by the police as a potential terrorist. How has this

For document1, we predict cluster10, which belongs to Sport Catalog predicted by problem c. Among these 12 documents, there are roughly 10 documents belong to Sport Catalog. Also, it is the same label with the cluster label we got from problem c. So it is a very good cluster for those documents. But they may also belong to other clusters that are labeled by Sport Catalog.

By far, we can get the conclusion that this is a good cluster, but may not the best one. In order to check if it is a best one, we need to read other documents in other Catalog label.

However, anyway, the result for this one, it is already very good.

2. For the cluster associated with document 15

```
m = 15;
[r c] = size(z2);
cluster = z2(m);
fprintf('cluster: %d\n',cluster);
n = 1;
for i=1:r
    if(cluster==z2(i))
        fprintf('%d\n',n);
        fname = sprintf('example1/20000101.%04d.txt',i);
        txt = textread(fname,'%s',10,'whitespace','\r\n');
        fprintf('%s\n',txt{:});
        n = n+1;
        if(n>12)
            break;
        end
    end
end
end
```

cluster: 19

1 (Technology)

Despite a few sputters and glitches, the world's computers appear to have survived the Year 2000 rollover without major problems _ and with humanity's faith in technology intact, at least for another day.

As clocks passed midnight around the world, there were a flurry of reports of apparently minor problems: a timing device at a electric plant in Wisconsin jumped ahead 35 days, but it was quickly reset; a monitoring system at a Japanese nuclear plant malfunctioned, but it did not affect the operation of the reactor; in Australia, ticketing machines on some buses jammed.

2 (Story)

Don't call Jim and Susan Smith survivalists.

In fact, you don't even have to call them Jim and Susan _ those

aren't their real names.

Thing is, they don't want their real names used. They have a good thing _ a secure thing _ going, and they'd just as soon not publicize it.

They're not digging bomb shelters and running around in camouflage. The whole idea of locked-and-loaded Uzis and concertina wire festooned around the perimeter and hundreds of cases of pork and beans in the bunker gives them the creeps.

3 (Story)

Don't call Jim and Susan Smith survivalists.

In fact, you don't even have to call them Jim and Susan _ those aren't their real names.

Thing is, they don't want their real names used. They have a good thing _ a secure thing _ going, and they'd just as soon not publicize it.

They're not digging bomb shelters and running around in camouflage. The whole idea of locked-and-loaded Uzis and concertina wire festooned around the perimeter and hundreds of cases of pork and beans in the bunker gives them the creeps.

4 (Technology – Y2K)

The eastern half of the United States smoothly followed the rest of the world into the new century early Saturday with only a smattering of minor Y2K glitches reported.

Slot machines in Delaware went bonkers and clocks at a Wisconsin electricity plant went haywire, as if to highlight the lack of

serious problems in the nation.

“We have seen no indications of any significant Y2K problems,” said John Koskinen, chief of the White House’s Y2K command center, based on reports from the Eastern and Central time zones.

He reported some congestion on telephone networks on the East

5 (Technology)

Now that the Year 2000 has arrived around the world without significant disruptions of power, transportation, commerce and other activities dependent on computers, the question arises: Was the threat of technology failure overstated, or was it all those hundreds of billions of dollars that were spent to fix things that prevented a catastrophe?

Most computer experts insist that the problem was a widespread and major threat before work began. “There was puffery by vendors and some money was wasted, but these were real problems,” said Leon Kappelman, a University of North Texas associate professor of

6 (Technology – Y2K)

As the world glided smoothly into the new century, the United States reported Saturday only a smattering of minor Y2K glitches, but officials cautioned that problems could still crop up, especially starting Monday.

John Koskinen, chief of the White House’s Y2K command center, said Monday will be a crucial milestone because it will be the first day of normal business operations following the long holiday weekend.

Otherwise, he offered generally upbeat news about the nation's progress. ``We have not been able to find anything of great

7 (Technology – Y2K)

As the world glided smoothly into the new century, the United States reported Saturday only a smattering of minor Y2K glitches, but officials cautioned that problems could still crop up, especially starting Monday.

John Koskinen, chief of the White House's Y2K command center, said Monday will be a crucial milestone because it will be the first day of normal business operations following the long holiday weekend.

Otherwise, he offered generally upbeat news about the nation's progress. ``We have not been able to find anything of great

8 (Technology – Y2K)

A stockpile of \$750,000 worth of spare parts set aside for Year 2000 fixes remained in a storeroom Saturday in New York City's computer center in Brooklyn.

The only power failure associated with the transition to 2000 came when a foil balloon in Queens rose and struck some wires, blacking out 100 customers, utility officials said.

City officials said it was too early to say that no computer repairs would be needed, with systems of lesser importance scheduled to kick on in the days ahead. But they emphasized that globally, it was increasingly evident that something remarkable had

9 (Technology – Y2K)

As the world glided smoothly into the new century,
the United States reported Saturday only a smattering of minor Y2K
glitches, but officials cautioned that problems could still crop
up, especially starting Monday.

John Koskinen, chief of the White House's Y2K command center,
said Monday will be a crucial milestone because it will be the
first day of normal business operations following the long holiday
weekend.

Otherwise, he offered generally upbeat news about the nation's
progress. ``We have not been able to find anything of great

10 (Technology – Y2K)

Moving in (w) Washington and (f) financial categories.

By MARILYN GEEWAX

c. 2000 Cox News Service

WASHINGTON _ While country after country continued to report
success this weekend in smothering the Y2K computer bug, U.S.
leaders delayed declarations of victory Saturday. They say the true
test of Y2K will come Monday.

``It isn't over yet," said Rep. Steve Horn, R-Calif., chairman
of the House subcommittee that investigated Y2K readiness. ``We
have to see what, if any, damage will be done" once computers face

11 (Technology – Y2K)

Moving in (w) Washington and (f) financial categories.

By MARILYN GEEWAX

c. 2000 Cox News Service

WASHINGTON _ While country after country continued to report success this weekend in smothering the Y2K computer bug, U.S. leaders delayed declarations of victory Saturday. They say the true test of Y2K will come Monday.

“It isn't over yet,” said Rep. Steve Horn, R-Calif., chairman of the House subcommittee that investigated Y2K readiness. “We have to see what, if any, damage will be done” once computers face

12 (Technology – Y2K)

As the world glided smoothly into the new century, the United States reported Saturday only a smattering of minor Y2K glitches, but officials cautioned that problems could still crop up, especially starting Monday.

John Koskinen, chief of the White House's Y2K command center, said Monday will be a crucial milestone because it will be the first day of normal business operations following the long holiday weekend.

Otherwise, he offered generally upbeat news about the nation's progress. “We have not been able to find anything of great

For document15, we predict cluster19, which belongs to Technology Catalog predicted by problem c. Among these 12 documents, there are roughly 10 documents belong to Technology Catalog. Also, it is the same label with the cluster label we got from problem c. The most important thing we can see is that in this cluster, we can see most documents are not only belong to technology area but also belong to more specific catalog – Y2K of technology area for those documents. So it is very high chance for them to stay in the best cluster.

3. For the cluster associated with document 30

```
m = 30;
[r c] = size(z2);
cluster = z2(m);
fprintf('cluster: %d\n',cluster);
n = 1;
for i=1:r
    if(cluster==z2(i))
        fprintf('%d\n',n);
        fname = sprintf('example1/20000101.%04d.txt',i);
        txt = textread(fname,'%s',10,'whitespace','\r\n');
        fprintf('%s\n',txt{:});
        n = n+1;
        if(n>12)
            break;
        end
    end
end
end
```

cluster: 17

1 (Politics - Russia)

When Boris Yeltsin unveiled his millennium surprise on New Year's Eve, he used the word "power" five times in his brief retirement speech.

When Lenin issued his famous call to revolution on Oct. 24, 1917, he also used the word "power" five times. Lenin was getting ready to seize it. Yeltsin, who will go down in history as the man who toppled the totalitarian system Lenin wrought, said he was relinquishing it.

Yet power, and the thirst for it, is the connection between two Russian leaders who only a few years ago symbolized opposite

2 (Politics - Russia)

When Boris Yeltsin unveiled his millennium surprise on New Year's Eve, he used the word "power" five times in his brief retirement speech.

When Lenin issued his famous call to revolution on Oct. 24, 1917, he also used the word "power" five times. Lenin was getting ready to seize it. Yeltsin, who will go down in history as the man who toppled the totalitarian system Lenin wrought, said he was relinquishing it.

Yet power, and the thirst for it, is the connection between two Russian leaders who only a few years ago symbolized opposite

3 (Business and Computer)

At times more words, not fewer, are needed to decode particularly turgid passages. Here's one example cited by Lutz, with a suggested revision.

Before: When business processes are automated, employees are careful not to fall into the trap of applying new technology to old, inefficient work procedures. Instead, a needs assessment is completed to identify system requirements, and then automated systems are designed to accomplish these goals.

After: Computers don't improve the way you do business, if you simply do business the same old way using the computers. So, before

4 (Business and Computer)

At times more words, not fewer, are needed to decode particularly turgid passages. Here's one example cited by Lutz, with a suggested revision.

Before: When business processes are automated, employees are careful not to fall into the trap of applying new technology to old, inefficient work procedures. Instead, a needs assessment is completed to identify system requirements, and then automated systems are designed to accomplish these goals.

After: Computers don't improve the way you do business, if you simply do business the same old way using the computers. So, before

5 (Politics)

After independence came to both nations in 1962, the polarization in both only increased with round after round of ethnic slaughters. People were killed merely for belonging to one ethnic group or the other, something that had never happened in either country.

In Rwanda, that cycle came to a crisis under the presidency of Juvenal Habyarimana, a Hutu whose administration demonized all Tutsi as supporters of rebels fighting to oust him. His death in a plane crash in April 1994, in which the Burundian president also died, set off the 100 days of slaughter of Tutsis.

6 (Politics)

After independence came to both nations in 1962, the polarization in both only increased with round after round of ethnic slaughters. People were killed merely for belonging to one ethnic group or the other, something that had never happened in either country.

In Rwanda, that cycle came to a crisis under the presidency of Juvenal Habyarimana, a Hutu whose administration demonized all Tutsi as supporters of rebels fighting to oust him. His death in a plane crash in April 1994, in which the Burundian president also died, set off the 100 days of slaughter of Tutsis.

7 (Politics)

After independence came to both nations in 1962, the polarization in both only increased with round after round of ethnic slaughters. People were killed merely for belonging to one ethnic group or the other, something that had never happened in either country.

In Rwanda, that cycle came to a crisis under the presidency of Juvenal Habyarimana, a Hutu whose administration demonized all

Tutsi as supporters of rebels fighting to oust him. His death in a plane crash in April 1994, in which the Burundian president also died, set off the 100 days of slaughter of Tutsis.

8 (Politics - Russia)

On his first full day as Russia's acting president, Vladimir Putin left before dawn to visit with Russian troops fighting in Chechnya.

While his predecessor, Boris Yeltsin, spent his first day of retirement at a government country house outside Moscow, Putin and his wife, Lyudmila _ he in a turtleneck sweater, she in pants and a sheepskin coat _ spent what was left of the momentous New Year's holiday greeting Russian soldiers and handing out hunting knives as presidential souvenirs.

Besides offering the nation a new image of an active, concerned

9 (Politics - Russia)

Only Vladimir Putin knows for certain where he wants to take Russia as the nation's acting president. But if his visit last week to the Railways Ministry is any indication, Russia may be in for an interesting trip indeed.

As the Moscow daily Kommersant reported, Putin performed the rote ceremonial duties at a packed meeting of ministry employees eight days ago, doling out awards and praising the railroads' contribution to the economy.

But then his tone turned icy. Why, he asked, did the railroads beseech the Kremlin for money when certain customers got special

10 (Politics)

Such shortcomings can be overcome, he said, but only by devising and swiftly carrying out a plan for the nation's restructuring and recovery. "The paramount word is fast," he wrote, "and we have no time for a slow start."

Sobchak, who recruited Putin to a team of economic experts when he was St. Petersburg's reformist mayor in the early 1990s, said the new president would benefit from his years as an intelligence agent in East Germany, where he was in regular contact with Western businesses and governments.

"He is a convinced supporter of the market economy," Sobchak

11 (Politics)

Barely six months ago, Brazil's government achieved a long-cherished objective with a carefully negotiated merger of all three of its military forces into a single ministry that is theoretically under civilian control.

But now, to the government's embarrassment, the air force, the most nationalistic of Brazil's armed services, is very noisily challenging that concept.

The dispute has already led to the dismissal of the air force chief, Gen. Walter Brauer. But it has also raised questions about the proper role of the Brazilian military in politics and the

12 (Politics - Russia)

Acting president Vladimir Putin moved quickly Saturday to establish himself as Russia's new leader by flying to the embattled republic of Chechnya to praise the army. Russia's most popular politician took over the reins of power after Boris Yeltsin's surprise New Year's Eve announcement he was resigning as president after eight tumultuous years. Putin held the post as prime minister and, under the Russian Constitution, takes over Yeltsin's job until new elections are held. The slight, judo-loving former KGB officer is favored to win the elections, tentatively set for March 26. However, analysts agree

For document30, we predict cluster17, which belongs to Politics Catalog predicted by problem c. Among these 12 documents, there are roughly 10 documents belong to Politics Catalog. Also, it is the same label with the cluster label we got from problem c. The most important thing we can see is that in this cluster, we can see some(5) documents are not only belong to Politics area but also belong to more specific catalog – Russia Politics for those documents. So it is very high chance for them to stay in the best cluster.

e.

```
[z10,c10,sumd10] = kmeans(Xn,40)
```

```
sumd10 =
```

```
1.7644
```

The result cost with 40 clusters is less than the one with 20 clusters. More clusters mean more specific class for each document.

(1). For the cluster associated with document1

```
m = 1; %1,15,30
[r c] = size(z10);
cluster = z10(m);
fprintf('cluster: %d\n',cluster);
n = 1;
for i=1:r
    if(cluster==z10(i))
        fprintf('%d\n',n);
        fname = sprintf('example1/20000101.%04d.txt',i);
        txt = textread(fname,'%s',10,'whitespace','\r\n');
        fprintf('%s\n',txt{:});
        n = n+1;
        if(n>12)
            break;
        end
    end
end
end
```

cluster: 27

1 (Sport)

School has been out at Cal State Northridge since the week before Christmas, but since you can learn something every day, Mississippi State's women's basketball team gave a lesson. Northridge has talked about taking its game to the next level. The 21st-ranked Bulldogs _ the first nationally ranked team to play here in Northridge's Division I era _ gave a glimpse of that level in a 98-64 nonconference victory before a crowd of 165 Friday night.

It ended CSUN's 16-game home win streak.

Whether the Matadors learned from the Bulldogs (10-0) remains to

2 (Intro a character)

Ray Hubbard, a television producer and broadcasting executive and a pioneer in the medium, died on Dec. 27 in Kenwood, Calif. He

was 75 and had been suffering from Parkinson's disease.

Hubbard was vice president for programming and production at Post-Newsweek Stations in Washington, the broadcast group of The Washington Post, from 1969 to 1976. Before that he was the national programming manager and executive producer of public affairs for Group W, the Westinghouse Broadcasting Co. in New York, where he helped to develop the Steve Allen, Mike Douglas, Merv Griffin and Regis Philbin shows and ``P.M. East'' with Mike Wallace.

3 (Art)

One of the happiest aspects of the Alvin Ailey American Dance Theater winter season, which ends Sunday, has been the performing.

Burnished to a glow by Judith Jamison and Masazumi Chaya, longtime former Ailey dancers who are now, respectively, the troupe's artistic director and associate artistic director, the performers' high technical level and individuality were seen to their best advantage in Hans van Manen's ``Polish Pieces'' on Wednesday afternoon at City Center.

The piece, choreographed in 1995 and set to a driving score by

4 (Sport)

How do you mark the passage of 1,000 years of human history? Simple, in this city of chronic gamblers and cold-eyed oddsmakers: You head for the track.

At 12:45 a.m., as the last wisps of smoke from New Year's Eve fireworks drifted away from the Happy Valley Racecourse, the crack

of a starter's gun set off the Millennium Cup. The Hong Kong Jockey Club billed it as the first horse race of the 2000s, though rumors swirled that it may have been edged out for first-in-the-world status by post-midnight races in New Zealand and Australia. Floral Joy won \$110,000 from a total purse worth \$200,000. It

5 (Celebration)

Stars rained over the nation's capital on New Year's eve as President Clinton ushered in the new year by lighting a fuse that sent a thundering shower of brilliant white stars cascading down the sides of the Washington Monument. ``Tonight, the torch passed to a new century of young Americans," Clinton told the more than 200,000 people gathered on the National Mall to ring in the Year 2000. ``Tonight, we celebrate the change of centuries and the dawning of a new millennium." Amid tight security, the crowd enjoyed relatively balmy weather _ the temperature was in the higher 30s _ and seemed unconcerned

6 (Celebration)

Los Angeles' city parties and concerts on New Year's Eve were a dud of millennial proportions Friday, drawing far smaller crowds of revelers than expected. A rare rainstorm coupled with Y2K boredom and uncertainty kept much of Los Angeles home from the five city parties, and even The Eagles concert failed to fill the Staples Center and left scalpers eating high-priced tickets. At the city's icon, the Hollywood sign, the laser light show and

fireworks displays were nixed. Officials gave contradictory explanations for why they were canceled and why officials continued

7 (Celebration – new year)

Thousands ventured to the Staples Center to wind down the 1990s with The Eagles, Linda Ronstadt and Jackson Browne, in a concert featuring pillars of the L.A. music scene.

Outside the Staples Center, ticketholders for the Eagles seemed to be in good spirits even as they waited in the chilly drizzle.

The clothing ran the gamut, from jeans to business suits to ...

Star Wars outfits, as they came for one reason: to indulge in a bit of nostalgia.

"I decided to do something I've never done before to ring out the New Year," said Andrew Milinkevich, a Thousand Oaks attorney

8 (Algiers)

High above the sparkling blue sweep of the Bay of Algiers, with panoramic views over whitewashed colonial palaces and palm-treed gardens that still make the city one of the jewels of the Mediterranean, there is a magnificent hostelry.

Once a palace, it became, at the height of France's 130 years of colonial rule in Algeria, the St. George's Hotel. On an upper floor, with a balcony scented in the springtime with mimosa blossoms, there is a suite with a brass plaque at the door recording the fact that General Dwight D. Eisenhower occupied the suite for several months in 1942 and 1943, when U.S. and allied

9 (Algiers)

High above the sparkling blue sweep of the Bay of Algiers, with panoramic views over whitewashed colonial palaces and palm-treed gardens that still make the city one of the jewels of the Mediterranean, there is a magnificent hostelry.

Once a palace, it became, at the height of France's 130 years of colonial rule in Algeria, the St. George's Hotel. On an upper floor, with a balcony scented in the springtime with mimosa blossoms, there is a suite with a brass plaque at the door recording the fact that General Dwight D. Eisenhower occupied the suite for several months in 1942 and 1943, when U.S. and allied

10 (Algiers)

As significant as its resources, Algeria fills a long stretch of the southern Mediterranean, and thus is of compelling military and political significance. And what happens in Algeria is of profound significance to many of America's key allies in Western Europe. Algiers is only an overnight ferry journey from Marseilles and even closer to Spain, allowing Algeria's enfeebled economy to send at least 3 million Algerians to Europe as immigrant workers in recent years.

And then there is the war, and how it may have led to Ressaam driving off that ferry in Washington state. In this, perhaps more

11 (Algiers)

As significant as its resources, Algeria fills a long stretch of the southern Mediterranean, and thus is of compelling military and political significance. And what happens in Algeria is of profound

significance to many of America's key allies in Western Europe.

Algiers is only an overnight ferry journey from Marseilles and even closer to Spain, allowing Algeria's enfeebled economy to send at least 3 million Algerians to Europe as immigrant workers in recent years.

And then there is the war, and how it may have led to Ressam driving off that ferry in Washington state. In this, perhaps more

12 (Hard to predict)

"We choose freedom because it gives us the power to live, learn, and love," wrote students from Tanapag Elementary School in the Northern Mariana Islands. Their essay, along with Ray Charles' sunglasses, photos of Earth from space, a Twinkie, Corningware, a computer chip, the Bill of Rights, a World War II helmet, a cell phone and a piece of the Berlin Wall will be among the items sealed this summer in the National Millennium Time Capsule until 2100.

The time capsule prototype was unveiled Friday by President Clinton and Hillary Rodham Clinton in Washington.

It is on display at the Smithsonian Institution's National

There are just 2 documents belong to Sport Catalog. Also, the differences among documents among this cluster are very large. There are 4 documents may belong to Algiers Catalog and there are 3 documents belong to Celebration Catalog. So this cluster may not be Sport Label/Catalog. The result is much worse than kmeans.

(2). For the cluster associated with document15

```
m = 15; %1,15,30
[r c] = size(z10);
cluster = z10(m);
fprintf('cluster: %d\n',cluster);
n = 1;
for i=1:r
    if(cluster==z10(i))
        fprintf('%d\n',n);
        fname = sprintf('example1/20000101.%04d.txt',i);
        txt = textread(fname,'%s',10,'whitespace','\r\n');
        fprintf('%s\n',txt{:});
        n = n+1;
        if(n>12)
            break;
        end
    end
end
end
```

cluster: 40

1 (commercial aviation)

Airports around the world, ordinarily quiet on New Year's Eve, were unusually so Friday as thousands of potential travelers stayed put while the world's air traffic system rolled over to the year 2000.

Yet, despite such fears and recent concerns about terrorism, commercial aviation entered the new year without incident as midnight arrived around the globe.

The major test for commercial aviation came at midnight in London when the global air traffic control system, which observes Greenwich Mean Time, crossed over into the new year without

2 (Technology – Y2K)

Despite a few sputters and glitches, the world's computers appear to have survived the Year 2000 rollover without major

problems _ and with humanity's faith in technology intact, at least for another day.

As clocks passed midnight around the world, there were a flurry of reports of apparently minor problems: a timing device at a electric plant in Wisconsin jumped ahead 35 days, but it was quickly reset; a monitoring system at a Japanese nuclear plant malfunctioned, but it did not affect the operation of the reactor; in Australia, ticketing machines on some buses jammed.

3 (Story)

Don't call Jim and Susan Smith survivalists.

In fact, you don't even have to call them Jim and Susan _ those aren't their real names.

Thing is, they don't want their real names used. They have a good thing _ a secure thing _ going, and they'd just as soon not publicize it.

They're not digging bomb shelters and running around in camouflage. The whole idea of locked-and-loaded Uzis and concertina wire festooned around the perimeter and hundreds of cases of pork and beans in the bunker gives them the creeps.

4 (Story)

Don't call Jim and Susan Smith survivalists.

In fact, you don't even have to call them Jim and Susan _ those aren't their real names.

Thing is, they don't want their real names used. They have a good thing _ a secure thing _ going, and they'd just as soon not

publicize it.

They're not digging bomb shelters and running around in camouflage. The whole idea of locked-and-loaded Uzis and concertina wire festooned around the perimeter and hundreds of cases of pork and beans in the bunker gives them the creeps.

5 (Technology – Y2K)

The eastern half of the United States smoothly followed the rest of the world into the new century early Saturday with only a smattering of minor Y2K glitches reported.

Slot machines in Delaware went bonkers and clocks at a Wisconsin electricity plant went haywire, as if to highlight the lack of serious problems in the nation.

“We have seen no indications of any significant Y2K problems,” said John Koskinen, chief of the White House's Y2K command center, based on reports from the Eastern and Central time zones.

He reported some congestion on telephone networks on the East

6 (air plane)

It was three hours before midnight on Friday and

John A. Koskinen did not appear worried a bit. His head was back.

His eyes were closed. And a flight attendant aboard the last Delta Shuttle flight of 1999 from New York to Washington was urging him, to no avail, to return his seat to the upright position.

“Excuse me, sir,” the attendant said, pulling at the seat.

“Excuse me.” It was three hours before midnight and the federal government's year 2000 coordinator, President Clinton's point man

on the potentially debilitating computer problems, was fast asleep.

That is not to say that Koskinen was not taking the arrival of

7 (Technology – Y2K)

As the world glided smoothly into the new century,

the United States reported Saturday only a smattering of minor Y2K

glitches, but officials cautioned that problems could still crop

up, especially starting Monday.

John Koskinen, chief of the White House's Y2K command center,

said Monday will be a crucial milestone because it will be the

first day of normal business operations following the long holiday

weekend.

Otherwise, he offered generally upbeat news about the nation's

progress. ``We have not been able to find anything of great

8 (Technology – Y2K)

As the world glided smoothly into the new century,

the United States reported Saturday only a smattering of minor Y2K

glitches, but officials cautioned that problems could still crop

up, especially starting Monday.

John Koskinen, chief of the White House's Y2K command center,

said Monday will be a crucial milestone because it will be the

first day of normal business operations following the long holiday

weekend.

Otherwise, he offered generally upbeat news about the nation's

progress. ``We have not been able to find anything of great

9 (Technology – Y2K)

As the world glided smoothly into the new century,
the United States reported Saturday only a smattering of minor Y2K
glitches, but officials cautioned that problems could still crop
up, especially starting Monday.

John Koskinen, chief of the White House's Y2K command center,
said Monday will be a crucial milestone because it will be the
first day of normal business operations following the long holiday
weekend.

Otherwise, he offered generally upbeat news about the nation's
progress. ``We have not been able to find anything of great

10 (New Year)

New Year's Day 2000 held good news and bad news for
the Holy Land. The evening before had passed with few technological
glitches or security threats, but there was no apparent arrival of
the Messiah, either.

Neither Israel nor the Palestinian-controlled areas reported any
problems with basic services on Saturday. While air traffic had
been called off overnight, the first flight left Israel's Ben
Gurion International Airport for Zurich without incident Saturday
morning.

Most businesses were closed. Banks, as a precaution, were not

11 (Technology – Y2K)

As the world glided smoothly into the new century,
the United States reported Saturday only a smattering of minor Y2K
glitches, but officials cautioned that problems could still crop

up, especially starting Monday.

John Koskinen, chief of the White House's Y2K command center, said Monday will be a crucial milestone because it will be the first day of normal business operations following the long holiday weekend.

Otherwise, he offered generally upbeat news about the nation's progress. ``We have not been able to find anything of great

12 (Technology – Y2K)

As the world glided smoothly into the new century, the United States reported Saturday only a smattering of minor Y2K glitches, but officials cautioned that problems could still crop up, especially starting Monday.

John Koskinen, chief of the White House's Y2K command center, said Monday will be a crucial milestone because it will be the first day of normal business operations following the long holiday weekend.

Otherwise, he offered generally upbeat news about the nation's progress. ``We have not been able to find anything of great

For document15, we predict cluster40. Among these 12 documents, there are roughly 7 documents belong to the same Catalog - Technology. Compared with problem d, this result is a little bit worse.

For other assigned documents to this cluster, the clusters are significantly different. In order to make predict better, we need to reduce the number of cluster.

(3) For the cluster associated with document30

```
m = 30; %1,15,30
[r c] = size(z10);
cluster = z10(m);
fprintf('cluster: %d\n',cluster);
n = 1;
for i=1:r
    if(cluster==z10(i))
        fprintf('%d\n',n);
        fname = sprintf('example1/20000101.%04d.txt',i);
        txt = textread(fname,'%s',10,'whitespace','\r\n');
        fprintf('%s\n',txt{:});
        n = n+1;
        if(n>12)
            break;
        end
    end
end
end
```

cluster: 2

1 (Politics - Russia)

When Boris Yeltsin unveiled his millennium surprise on New Year's Eve, he used the word "power" five times in his brief retirement speech.

When Lenin issued his famous call to revolution on Oct. 24, 1917, he also used the word "power" five times. Lenin was getting ready to seize it. Yeltsin, who will go down in history as the man who toppled the totalitarian system Lenin wrought, said he was relinquishing it.

Yet power, and the thirst for it, is the connection between two Russian leaders who only a few years ago symbolized opposite

2 (Politics - Russia)

When Boris Yeltsin unveiled his millennium surprise on New

Year's Eve, he used the word "power" five times in his brief retirement speech.

When Lenin issued his famous call to revolution on Oct. 24, 1917, he also used the word "power" five times. Lenin was getting ready to seize it. Yeltsin, who will go down in history as the man who toppled the totalitarian system Lenin wrought, said he was relinquishing it.

Yet power, and the thirst for it, is the connection between two Russian leaders who only a few years ago symbolized opposite

3 (Politics - Russia)

As always, Boris Yeltsin picked his moment.

By choosing to resign on the last day of the old year, Yeltsin guaranteed that almost all Russians would be deeply engaged in celebrating the arrival of the new one, traditionally the day of the biggest family feast here. Better than that, he acted at a time of enormous political flux: the start of a protracted break in the political year, and just 12 days after the election of a new Parliament, where rival blocs have yet to solidify.

The advantage of acting in this political void is that Yeltsin avoided the usual flurry of public discussion and political

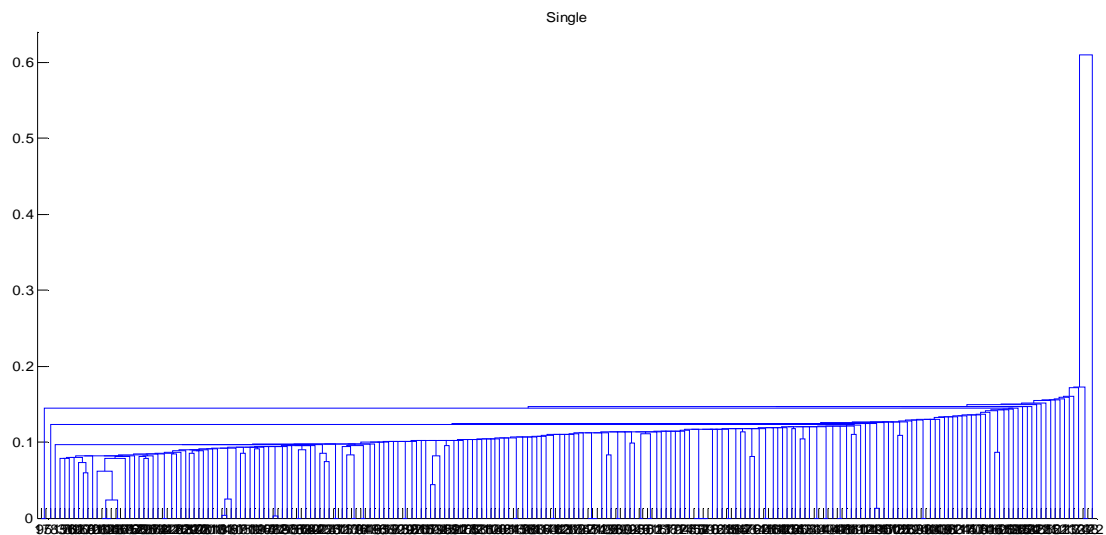
For document30, we predict cluster2. Among these 3 documents, all of them are not only belong to the same Catalog – Politics but also belong to a more specific catalog - Russia. Compared with problem d, this result is much better.

The result is already very nice and so we don't need to do any modify.

Problem3

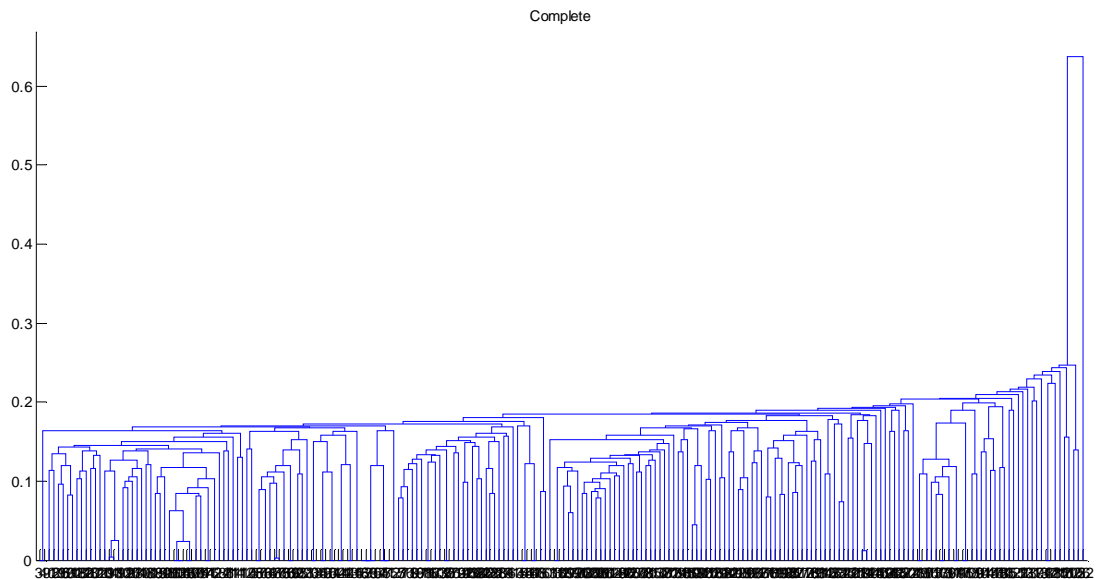
a.

```
X = full(Xn)  
Z = linkage(X,'single','euclidean');  
dendrogram(Z,0)
```



b.

```
X = full(Xn)
Z = linkage(X,'complete','euclidean');
dendrogram(Z,0)
```



There is difference between “single” linkage and “complete” linkage. For this problem, in average the “complete” linkage combines two clusters with more dissimilarities than “single” linkage.

Definitely, we prefer to pick the clusters generated by “single” linkage. Because, during the process procedure, the number of documents in each cluster is very closed. Additionally, we prefer to combine two clusters with lots of similarities, but not dissimilarities.

c. I use "complete" linkage.

```
X = full(Xn);  
Z = linkage(X,'complete','euclidean');  
c = cluster(Z,'maxclust',20);  
z = zeros(20,1);  
[r c2] = size(c);  
for i=1:r  
    n = c(i);  
    z(n) = z(n)+1;  
end
```

```
Z =  
3    (3 articles are associated with 1st cluster)  
185  (185 articles are associated with 2nd cluster)  
2    (2 articles are associated with 3th cluster)  
4    (4 articles are associated with 4th cluster)  
2    (2 articles are associated with 5th cluster)  
1    (1 article is associated with 6th cluster)  
14   (14 articles are associated with 7th cluster)  
1    (1 article is associated with 8th cluster)  
1    (1 article is associated with 9th cluster)  
2    (2 articles are associated with 10th cluster)  
1    (1 article is associated with 11th cluster)  
1    (1 article is associated with 12th cluster)  
1    (1 article is associated with 13th cluster)  
1    (1 article is associated with 14th cluster)  
2    (2 articles are associated with 15th cluster)  
1    (1 article is associated with 16th cluster)  
1    (1 article is associated with 17th cluster)  
2    (3 articles are associated with 18th cluster)  
2    (3 articles are associated with 19th cluster)  
1    (1 article is associated with 20th cluster)
```

d. For the cluster associated with document1: (it is the same cluster with document15, because document1 and document15 belong to the same cluster)

```
X = full(Xn);
Z = linkage(X,'complete','euclidean');
c = cluster(Z,'maxclust',20);
m = 1; %1,15,30
[r c2] = size(c);
clust = c(m);
fprintf('cluster: %d\n',clust);
n = 1;
for i=1:r
    if(clust==c(i))
        fprintf('%d\n',n);
        fname = sprintf('example1/20000101.%04d.txt',i);
        txt = textread(fname,'%s',10,'whitespace','\r\n');
        fprintf('%s\n',txt{:});
        n = n+1;
    if(n>12)
        break;
    end
end
end
```

cluster: 2

1 (Sport)

School has been out at Cal State Northridge since the week before Christmas, but since you can learn something every day, Mississippi State's women's basketball team gave a lesson. Northridge has talked about taking its game to the next level. The 21st-ranked Bulldogs _ the first nationally ranked team to play here in Northridge's Division I era _ gave a glimpse of that level in a 98-64 nonconference victory before a crowd of 165 Friday night.

It ended CSUN's 16-game home win streak.

Whether the Matadors learned from the Bulldogs (10-0) remains to

2 (Celebration)

EDS: New top and updates throughout.

By SCOTT MONTGOMERY

c.2000 Cox News Service

WASHINGTON _ Clear skies and a string of flawless midnight New Year's rollovers around the globe enticed a massive crowd to the National Mall for a fireworks spectacular ignited by President Clinton to greet the 21st century.

On a night when Clinton called on America to make its ``dreams stronger than its memories," 2000 arrived in the nation's capital in a shower of light over the Washington Monument that left

3 (Aviation)

Airports around the world, ordinarily quiet on New Year's Eve, were unusually so Friday as thousands of potential travelers stayed put while the world's air traffic system rolled over to the year 2000.

Yet, despite such fears and recent concerns about terrorism, commercial aviation entered the new year without incident as midnight arrived around the globe.

The major test for commercial aviation came at midnight in London when the global air traffic control system, which observes Greenwich Mean Time, crossed over into the new year without

4 (intro a character)

Ray Hubbard, a television producer and broadcasting executive and a pioneer in the medium, died on Dec. 27 in Kenwood, Calif. He was 75 and had been suffering from Parkinson's disease.

Hubbard was vice president for programming and production at Post-Newsweek Stations in Washington, the broadcast group of The Washington Post, from 1969 to 1976. Before that he was the national programming manager and executive producer of public affairs for Group W, the Westinghouse Broadcasting Co. in New York, where he helped to develop the Steve Allen, Mike Douglas, Merv Griffin and Regis Philbin shows and "P.M. East" with Mike Wallace.

5 (Celebration)

THIRD MILLENNIUM CELEBRATIONS CAST LIGHT ON TWO MEXICOS
By FRANC CONTRERAS

c.1999 Special to the Cox News Service

OAXACA CITY, Mexico _ While most of Mexico welcomed the Third Millennium with fireworks, concerts and long-standing traditions, the celebration was muted in rural areas by fear of rebel attacks and the so-called Y2K bug's negative impact on tourism.

All across Mexico and here, in this well-known colonial city, families joined together for traditional New Year's Eve supper and counted down to mid-night wearing red underwear, which they believe

6 (Celebration)

Two thousand years after Christ's obscure birth in a dusty town in Judea, the world's 6 billion people _ most of them non-Christian and many of them preoccupied with terrorism, computers, diets, bank accounts, politics and the perils of the future _ rode their turning blue planet across time's invisible line Friday and, by common consent, looked into the dawn of a new millennium.

What they saw first was a party. It was garish, glittering and global, and millions, setting religious considerations and personal concerns aside, joined in the festivities to celebrate the conjunction of a new year, a new century and a new thousand-year

7 (Celebration)

The millennium, an idea with overtones ranging from Biblical to commercial, had swelled recently into a coercive miniculture as the countdown ticked away and a flood of books, articles, television specials and studied commentaries by academic, political and

religious leaders reflected upon the last thousand years of human achievements and missteps, and speculated on the next thousand. Purists still insist that the millennium will not start until Jan. 1, 2001, and they are no doubt right. The calendar, at best, is arbitrary. There were problems from ancient times based on inaccuracies in measuring the year's duration and its uneven

8 (Celebration)

The festivities for most of India's 1 billion people were muted by comparison with those in wealthier nations, but hundreds of thousands danced, drank and ate at open-air stalls in New Delhi and partygoers were out in force in other major cities, celebrating the peaceful end of an eight-day hostage crisis aboard an Indian Airlines hours before midnight.

Iran and its Persian Gulf neighbors largely ignored the new millennium, which fell during the Islamic holy month of Ramadan, a time for prayer and reflection.

With religious bans on celebrations that might desecrate holy

9 (Celebration)

Although it had seen many huge crowds over the years, for inaugurations, Fourth of July fireworks, protests and parades, never in memory had the Mall in the nation's capital drawn multitudes for New Year's. But on this mild winter night with a lovely sunset, hundreds of thousands of people here thronged to the Mall for a long-planned celebration on the last New Year's Eve of the 1990s.

"It's a very friendly crowd," Sgt. Frank R. Onolfi of the U.S. Park Police said as afternoon turned to evening. But by shortly after 10 p.m., there were so many people trying to get close to the

10 (Celebration)

How do you mark the passage of 1,000 years of human history? Simple, in this city of chronic gamblers and cold-eyed oddsmakers: You head for the track.

At 12:45 a.m., as the last wisps of smoke from New Year's Eve fireworks drifted away from the Happy Valley Racecourse, the crack of a starter's gun set off the Millennium Cup. The Hong Kong Jockey Club billed it as the first horse race of the 2000s, though rumors swirled that it may have been edged out for first-in-the-world status by post-midnight races in New Zealand and Australia.

Floral Joy won \$110,000 from a total purse worth \$200,000. It

11 (technology)

Despite a few sputters and glitches, the world's computers appear to have survived the Year 2000 rollover without major problems _ and with humanity's faith in technology intact, at least for another day.

As clocks passed midnight around the world, there were a flurry of reports of apparently minor problems: a timing device at a electric plant in Wisconsin jumped ahead 35 days, but it was quickly reset; a monitoring system at a Japanese nuclear plant

malfunctioned, but it did not affect the operation of the reactor;
in Australia, ticketing machines on some buses jammed.

12 (Story)

Don't call Jim and Susan Smith survivalists.

In fact, you don't even have to call them Jim and Susan _ those
aren't their real names.

Thing is, they don't want their real names used. They have a
good thing _ a secure thing _ going, and they'd just as soon not
publicize it.

They're not digging bomb shelters and running around in
camouflage. The whole idea of locked-and-loaded Uzis and concertina
wire festooned around the perimeter and hundreds of cases of pork
and beans in the bunker gives them the creeps.

There are 7 documents belong to the same catalog - Celebration. But the differences between each
cluster are much large. Because we use the "complete" linkage, which combines two clusters with the
furtherest distance. Compared with kmeans, it should a little bit worse.

(2). For the cluster associated with document30

```
X = full(Xn);
Z = linkage(X,'complete','euclidean');
c = cluster(Z,'maxclust',20);
m = 30; %1,15,30
[r c2] = size(c);
clust = c(m);
fprintf('cluster: %d\n',clust);
n = 1;
for i=1:r
    if(clust==c(i))
        fprintf('%d\n',n);
        fname = sprintf('example1/20000101.%04d.txt',i);
        txt = textread(fname,'%s',10,'whitespace','\r\n');
        fprintf('%s\n',txt{:});
        n = n+1;
    if(n>12)
        break;
    end
end
end
end
```

cluster: 7

1 (Politics - Russia)

When Boris Yeltsin unveiled his millennium surprise on New Year's Eve, he used the word "power" five times in his brief retirement speech.

When Lenin issued his famous call to revolution on Oct. 24, 1917, he also used the word "power" five times. Lenin was getting ready to seize it. Yeltsin, who will go down in history as the man who toppled the totalitarian system Lenin wrought, said he was relinquishing it.

Yet power, and the thirst for it, is the connection between two Russian leaders who only a few years ago symbolized opposite

2 (Politics - Russia)

When Boris Yeltsin unveiled his millennium surprise on New Year's Eve, he used the word "power" five times in his brief retirement speech.

When Lenin issued his famous call to revolution on Oct. 24, 1917, he also used the word "power" five times. Lenin was getting ready to seize it. Yeltsin, who will go down in history as the man who toppled the totalitarian system Lenin wrought, said he was relinquishing it.

Yet power, and the thirst for it, is the connection between two Russian leaders who only a few years ago symbolized opposite

3 (Politics - Russia)

On his first full day as Russia's acting president, Vladimir Putin left before dawn to visit with Russian troops fighting in Chechnya.

While his predecessor, Boris Yeltsin, spent his first day of retirement at a government country house outside Moscow, Putin and his wife, Lyudmila _ he in a turtleneck sweater, she in pants and a sheepskin coat _ spent what was left of the momentous New Year's holiday greeting Russian soldiers and handing out hunting knives as presidential souvenirs.

Besides offering the nation a new image of an active, concerned

4 (Politics - Russia)

Only Vladimir Putin knows for certain where he wants to take Russia as the nation's acting president. But if his visit last week to the Railways Ministry is any indication, Russia may be in

for an interesting trip indeed.

As the Moscow daily Kommersant reported, Putin performed the rote ceremonial duties at a packed meeting of ministry employees eight days ago, doling out awards and praising the railroads' contribution to the economy.

But then his tone turned icy. Why, he asked, did the railroads beseech the Kremlin for money when certain customers got special

5 (Politics - Russia)

Such shortcomings can be overcome, he said, but only by devising and swiftly carrying out a plan for the nation's restructuring and recovery. "The paramount word is fast," he wrote, "and we have no time for a slow start."

Sobchak, who recruited Putin to a team of economic experts when he was St. Petersburg's reformist mayor in the early 1990s, said the new president would benefit from his years as an intelligence agent in East Germany, where he was in regular contact with Western businesses and governments.

"He is a convinced supporter of the market economy," Sobchak

6 (Story)

On most days, 11-year-old Azamat Malikov spends at least an hour or two doing what comes naturally to many people in Kyrgyzstan. He works at memorizing what is said to be the world's longest epic poem.

One sunny afternoon, Azamat, resplendent in a dark blue velvet robe, chanted the story of a battle so vividly that he even

frightened a visitor who could speak no Kyrgyz. Azamat rocked to and fro, waved his arms and changed his expressions wildly to illustrate the text.

“My grandfather asked me to please learn the Manas epic,”

7 (Politics)

In 1995 Akayev convened regional leaders and other dignitaries for a lavish celebration of the 1,000th anniversary of the Manas epic. The date was purely symbolic, and the expense left a bad taste in a country where most people's living standards have plummeted since the fall of Communism.

During his speech at the celebration, Akayev said that everyone in Kyrgyzstan should adhere to the seven basic principles of Manas. This raised some eyebrows, since no one had ever heard of such principles. It later turned out that the president himself had divined them.

8 (Politics - Russia)

Acting president Vladimir Putin moved quickly Saturday to establish himself as Russia's new leader by flying to the embattled republic of Chechnya to praise the army.

Russia's most popular politician took over the reins of power after Boris Yeltsin's surprise New Year's Eve announcement he was resigning as president after eight tumultuous years. Putin held the post as prime minister and, under the Russian Constitution, takes over Yeltsin's job until new elections are held.

The slight, judo-loving former KGB officer is favored to win the

elections, tentatively set for March 26. However, analysts agree

9 (Politics - Russia)

As always, Boris Yeltsin picked his moment.

By choosing to resign on the last day of the old year, Yeltsin guaranteed that almost all Russians would be deeply engaged in celebrating the arrival of the new one, traditionally the day of the biggest family feast here. Better than that, he acted at a time of enormous political flux: the start of a protracted break in the political year, and just 12 days after the election of a new Parliament, where rival blocs have yet to solidify.

The advantage of acting in this political void is that Yeltsin avoided the usual flurry of public discussion and political

10 (Politics - Russia)

The New York Times said in an editorial for Sunday, Jan. 2:

In suddenly resigning as Russian president on Friday, Boris Yeltsin showed once again why he will be remembered as a practitioner of the unexpected political stratagem. Like so many of Yeltsin's gambits over the years, the move upended expectations and scrambled Russian politics. It also came with a promise of immunity from prosecution for Yeltsin for any misdeeds of his government.

That will likely roil the coming presidential campaign, a fitting legacy for a courageous but disappointingly erratic man who guided Russia through the first years of a turbulent, still unfinished

11 (Politics - Russia)

Not content to sit pat a day after assuming power in

Russia, the acting president, Vladimir Putin, reminded voters Saturday why they like him so much, with a surprise visit to troops fighting the war in Chechnya.

Putin's overnight trip to the Russian-held city of Gudermes, to present soldiers with medals and with hunting knives engraved with his signature, and to offer a ringing vote of support for their efforts to crush separatist militants, bore the traits that make him the favorite to win in a landslide in an early presidential vote set for March 26.

12 (Politics - Russia)

Not content to sit pat a day after assuming power in Russia, the acting president, Vladimir Putin, reminded voters Saturday why they like him so much, with a surprise visit to troops fighting the war in Chechnya.

Putin's overnight trip to the Russian-held city of Gudermes, to present soldiers with medals and with hunting knives engraved with his signature, and to offer a ringing vote of support for their efforts to crush separatist militants, bore the traits that make him the favorite to win in a landslide in an early presidential vote set for March 26.

For document30, we predict cluster7, which belongs to Politics Catalog. Among these 12 documents, there are 10 documents belong to Politics Catalog. Also, it is the same label with the cluster label we got from problem c. The most important thing we can see is that in this cluster, we can see most(9) documents are not only belong to Politics area but also belong to a more specific catalog – Russia Politics for those documents. So it is very high chance for them to stay in the best cluster.

Also, it is better than kmeans.