Table of Contents

Question 1

```
format

A=[ 3 -5 -5 ; 8 6 0 ; -7 2 2 ; 6 0 3 ; 8 1 7 ];
b=[ 10 1 -6 10 9 ]';

fprintf('Question 1a\n')
x = inv(A'*A)*A'*b

fprintf('Question 1b\n')
a = A(:,1);b = A(:,2);c = A(:,3);
q1 = a/norm(a);
B = b - b'*q1 * q1; q2 = B/norm(B);
C = c - c'*q1*q1 - c'*q2*q2;  q3 = C/norm(C);
Q = [q1,q2,q3];
R = Q'*A
```

```
Question 1a

x =

    1.0427
   -1.3417
    0.3722

Question 1b

R =

   14.8997    1.8121    3.0202
    0.0000    7.9194    3.8547
    0.0000   -0.0000    7.9385
```

Question 2

```
fprintf('Question 2a\n')
X=[ 5.6 4 6 4.4 5 5.6 4 6 4.4 5 5.6 4;
6.2 4.7 6.5 5 5.6 6.4 4.7 6.8 5 5.8 6.4 4.7;
4.9 5.1 4.9 5.1 5 4.9 5.1 4.9 5.1 5 4.9 5.1;
5.9 5.2 6 5.3 5.6 6.1 5.1 6.3 5.3 5.8 6.2 5.2 ];

Y=[ 19.9 20.1 20 19.9 20.1 20 19.9 20.1 20 19.9 20.1 20;
20.6 20.1 20.1 20.6 20.1 20.7 21.1 20.5 20.8 20.7 20.3 21;
```

```
20.9 19.1 20 20.9 19.1 20 20.9 19.1 20 20.9 19.1 20 ];

[M,N] = size(X);
[L,~] = size(Y);

means_X = mean(X,2)
means_Y = mean(Y,2)

for m=1:M; X(m,:)=X(m,:)-means_X(m); end
for l=1:L; Y(l,:)=Y(l,:)-means_Y(l); end

sd_X = sqrt(sumsqr(X)/(N*M))
sd_Y = sqrt(sumsqr(Y)/(N*L))

X = X/sd_X
Y = Y/sd_Y

fprintf('Question 2b\n')
F=[X;Y];
[U,S,V] = svd(F);

sigmas = diag(S).^2 / N;


fprintf('percent variance explained by each PC:\n');disp(sigmas/
sum(sigmas))

fprintf('PC #1 (the first column of U:)\n');disp(U(:,1))

fprintf('Question 2c\n')

C = X*Y'/N

fprintf('Question 2d\n')

[U,S,V] = svd(C);
disp('The singular values are:')
disp(diag(S)');
disp('based on the singular values, only the first SVD mode')
disp('is important.  Consider therefore U(:,1),V(:,1):')
disp('U(:,1):')
disp(U(:,1)')
disp('V(:,1):')
disp(V(:,1)')

fprintf('Question 2e\n')
disp('Total covariance:'); disp(sumsqr(C))
disp('Sum over the singular values squared:'); disp(sum(diag(S).^2))

fprintf('percent variance explained by each SVD mode:\n');
disp(diag(S).^2/sum(diag(S).^2))
```

Question 2a

```
means_X =

    4.9667
    5.6500
    5.0000
    5.6667


means_Y =

   20.0000
   20.5500
   20.0000


sd_X =

    0.5778


sd_Y =

    0.4670


X =

  Columns 1 through 7

    1.0962   -1.6731    1.7885   -0.9808    0.0577    1.0962   -1.6731
    0.9519   -1.6442    1.4712   -1.1250   -0.0865    1.2981   -1.6442
   -0.1731    0.1731   -0.1731    0.1731         0   -0.1731    0.1731
    0.4039   -0.8077    0.5769   -0.6346   -0.1154    0.7500   -0.9808

  Columns 8 through 12

    1.7885   -0.9808    0.0577    1.0962   -1.6731
    1.9904   -1.1250    0.2596    1.2981   -1.6442
   -0.1731    0.1731         0   -0.1731    0.1731
    1.0962   -0.6346    0.2308    0.9231   -0.8077


Y =

  Columns 1 through 7

   -0.2141    0.2141         0   -0.2141    0.2141         0   -0.2141
    0.1071   -0.9637   -0.9637    0.1071   -0.9637    0.3212    1.1778
    1.9273   -1.9273         0    1.9273   -1.9273         0    1.9273

  Columns 8 through 12

    0.2141         0   -0.2141    0.2141         0
   -0.1071    0.5354    0.3212   -0.5354    0.9637
```

```
   -1.9273         0    1.9273   -1.9273         0

Question 2b
percent variance explained by each PC:
    0.6347
    0.3251
    0.0370
    0.0030
    0.0002
    0.0001
    0.0000

PC #1 (the first column of U:)
   -0.5749
   -0.5904
    0.0686
   -0.3222
   -0.0459
    0.1992
    0.4130

Question 2c

C =

    0.0494   -0.3892   -0.4448
    0.0556   -0.3583   -0.5004
   -0.0062    0.0432    0.0556
    0.0371   -0.1884   -0.3336

Question 2d
The singular values are:
    0.9392    0.0680    0.0000

based on the singular values, only the first SVD mode
is important.  Consider therefore U(:,1),V(:,1):
U(:,1):
   -0.6296   -0.6578    0.0753   -0.4065

V(:,1):
   -0.0886    0.5968    0.7974

Question 2e
Total covariance:
    0.8867

Sum over the singular values squared:
    0.8867

percent variance explained by each SVD mode:
    0.9948
    0.0052
    0.0000
```

Question 3

```
fprintf('Question 3a\n')
fprintf('i.) logical variables\n')
a=[ 0 1 1 0 0 0 1 1 0 0 ]; b=[ 1 1 0 0 0 0 0 1 0 1 ];
fprintf('union of a and b:\n'); disp(union(a,b))
fprintf('intersect of a and b:\n'); disp(intersect(a,b))
J_index = sum(a & b)/sum(a | b); fprintf(1,'Jaccard Similarity=%g
\n',J_index)

fprintf('\nii.) integers\n')
a=[ 10 1 2 4 6 7 8 1 6 8 ]; b=[ 6 8 2 6 6 6 1 7 6 6 ];
fprintf('union of a and b:\n'); disp(union(a,b))
fprintf('intersect of a and b:\n'); disp(intersect(a,b))
J_index=length(intersect(a,b))/length(union(a,b)); fprintf(1,'Jaccard
 Similarity=%g\n',J_index)

fprintf('\niii.) words\n')
a={'april','is','the','cruellest','month','breeding','lilacs','out','of','the','de
b={'a','april','earth','lilacs','memory','winter','and','breeding','covering','cru
fprintf('union of a and b:\n'); disp(union(a,b))
fprintf('intersect of a and b:\n'); disp(intersect(a,b))
J_index=length(intersect(a,b))/length(union(a,b)); fprintf(1,'Jaccard
 Similarity=%g\n',J_index)

fprintf('\niv.) text files\n')
fid1 = fopen('words1.txt');
fid2 = fopen('words2.txt');
delimiters={' ','\r','\n','\t','â##','â##','â##','.',',','-'};
C1 = textscan(fid1,'%s','delimiter',delimiters);
carray1 = char(C1{:});
C2 = textscan(fid2,'%s','delimiter',delimiters);
carray2 = char(C2{:});
fclose(fid1); fclose(fid2);
a={C1{1}{:}};
b={C2{1}{:}};
fprintf('union of a and b:\n'); disp(union(a,b))
fprintf('intersect of a and b:\n'); disp(intersect(a,b))
J_index=length(intersect(a,b))/length(union(a,b)); fprintf(1,'Jaccard
 Similarity=%g\n',J_index)

fprintf('v.) Jaccard Similarity\n')
a=[ 4 1 1 2 4 4
5 3 4 1 3 4
2 3 5 1 2 1
2 3 4 4 1 2
1 2 4 1 5 2
4 1 1 3 1 5
4 1 3 1 3 4
1 1 3 1 4 1
1 2 1 4 3 1];

[~,N]=size(a);
```

```
Jaccards = zeros(N);

for n = 1:N
   for m = 1:N
       Jaccards(n,m) = length(intersect(a(:,n),a(:,m)))/
length(union(a(:,n),a(:,m)));
   end
end

Jaccards

fprintf('Question 3b\n')

shingles = {'of this El Nino expected';'this El Nino expected in';
    'in coming months By Michael';'By Michael Casey Published
 December';
    'are expecting this winter''s El';'this winter''s El Nino could';
    'could be the strongest since';'be the strongest since 1997-98'};

crcs = zeros(1,8);

for i=1:8
    crcs(i) = crc32(shingles{i});
end

format long
fprintf('crc values for each shingle:\n')
disp(crcs')

hashes = zeros(1,8);

for j = 1:8
    n=prod(double(shingles{j}));
    hash_value=mod(n,100003);
    hashes(j) = hash_value;
end

fprintf('hashvalues for each shingle:\n')
disp(hashes')
```

```
Question 3a
i.) logical variables
union of a and b:
     0     1

intersect of a and b:
     0     1

Jaccard Similarity=0.333333

ii.) integers
union of a and b:
     1     2     4     6     7     8    10
```

```
intersect of a and b:
     1     2     6     7     8

Jaccard Similarity=0.714286

iii.) words
union of a and b:
  Columns 1 through 6

    'a'    'and'    'april'    'breeding'    'covering'    'cruellest'

  Columns 7 through 12

    'dead'    'desire'    'dried'    'dull'    'earth'    'feeding'

  Columns 13 through 19

    'forgetful'    'in'    'is'    'land'    'lilacs'    'memory'
 'mixing'

  Columns 20 through 26

    'month'    'of'    'out'    'rain'    'roots'    'spring'
 'stirring'

  Columns 27 through 29

    'the'    'winter'    'with'

intersect of a and b:
  Columns 1 through 6

    'and'    'april'    'breeding'    'cruellest'    'dead'
 'desire'

  Columns 7 through 10

    'dull'    'is'    'lilacs'    'memory'

Jaccard Similarity=0.344828

iv.) text files
union of a and b:
  Columns 1 through 8

    ''    '"the'    '$4'    '$8m'    '(19)82'    '(19)97'    '(AO)'
 '(NOAA)'

  Columns 9 through 17

    '(The'    '00'    '1'    '10lb'    '11'    '116'    '13'    '17'
  '175in'
```

```
Columns 18 through 25

  '189'    '19(82)'    '1950'    '1970'    '1997'    '1st'    '2'
'20'

Columns 26 through 33

  '200'    '2010'    '2015'    '30%'    '31'    '40lb'    '5%'
'50%'

Columns 34 through 40

  '500'    '60'    '832in'    '89in'    '98'    'AO'
'Administration'

Columns 41 through 46

  'Alaska'    'Alpine'    'Andy'    'Angeles'    'April'    'Arctic'

Columns 47 through 51

  'Associated'    'Atmospheric'    'Australia'    'Below'
'Boomtime'

Columns 52 through 57

  'Buffalo'    'But'    'By'    'California'    'Californians'
'Casey'

Columns 58 through 63

  'Center'    'Christ'    'Christmas'    'City'    'Climate'
'Coast'

Columns 64 through 69

  'Conservation'    'Council'    'December'    'Del'    'Don'
'Donald'

Columns 70 through 75

  'East'    'El'    'England'    'Even'    'February'    'Florida'

Columns 76 through 81

  'Forecasters'    'FoxNews'    'GMT'    'Greg'    'Guardian'
'Halpert'

Columns 82 through 88

  'Hawaii'    'Heavenly'    'Heureux'    'However'    'I'    'If'
'In'

Columns 89 through 96
```

```
  'It'    'Its'    'January'    'L'    'Lake'    'Los'    'Many'
'March'

 Columns 97 through 102

  'Marine'    'Meadows'    'Michael'    'Michelle'    'Mike'
'Milman'

 Columns 103 through 108

  'Mountain'    'My'    'NOAA'    'National'    'Nevada'
'Nevertheless'

 Columns 109 through 114

  'New'    'Nino'    'NiÃ±o'    'NiÃ±os'    'Northeast'
'November'

 Columns 115 through 120

  'Ocean'    'Oceanic'    'Oefinger'    'Oliver'    'Oregon'
'Oscillation'

 Columns 121 through 126

  'Pacific'    'Paraguay'    'People'    'Pete'    'Plains'
'Prediction'

 Columns 127 through 132

  'Press'    'Press)'    'Published'    'Rather'    'Related:'
'Resort'

 Columns 133 through 139

  'Rey'    'Rick'    'S'    'Sierra'    'Sonntag'    'South'
'Spanish'

 Columns 140 through 145

  'Sportfishing'    'Squaw'    'States'    'Tahoe'    'Texas'
'That'

 Columns 146 through 152

  'The'    'There'    'They'    'This'    'Thursday'    'Trumps'
'U'

 Columns 153 through 159

  'US'    'United'    'Urban'    'Valley'    'Washington'    'Water'
  'We'
```

```
 Columns 160 through 166

   'Webber'    'West'    'While'    'Wirth'    'Worst'    'York'
'a'

 Columns 167 through 172

   'about'    'above'    'absolute'    'accepts'    'according'
'across'

 Columns 173 through 179

   'actually'    'after'    'air'    'all'    'along'    'already'
'also'

 Columns 180 through 185

   'amounts'    'an'    'and'    'animals'    'anticipated'
'aquifers'

 Columns 186 through 192

   'arctic'    'are'    'area'    'around'    'as'    'associated'
 'at'

 Columns 193 through 198

   'attention'    'attracted'    'attribute'    'average'    'back'
 'based'

 Columns 199 through 205

   'basin'    'be'    'because'    'been'    'being'    'believe'
'benefit'

 Columns 206 through 211

   'between'    'bigger'    'billion'    'black'    'blamed'
'blanketing'

 Columns 212 through 217

   'blue'    'boom'    'boy'    'breaker'    'breaking'    'bring'

 Columns 218 through 222

   'bringing'    'brings'    'brought'    'buffoon'    'bushfires'

 Columns 223 through 228

   'business'    'businesses'    'but'    'can'    'cause'
'caused'

 Columns 229 through 233
```

```
  'caution'    'cautioned'    'center'    'central'    'certainly'

 Columns 234 through 238

  'certainty'    'challenging'    'change'    'chaos'
'characteristic'

 Columns 239 through 245

  'chief'    'child"'    'climate'    'coast'    'cold'    'com'
'comes'

 Columns 246 through 251

  'coming'    'conditions'    'contiguous'    'cooler'    'could'
'couldn'

 Columns 252 through 257

  'country'    'couple'    'cowboy'    'crabs'    'created'
'current'

 Columns 258 through 263

  'd'    'damage'    'days'    'declared'    'definitely'
'degrees'

 Columns 264 through 268

  'demand'    'depleted'    'deputy'    'desperate'    'difference'

 Columns 269 through 275

  'director'    'disasters'    'do'    'don'    'dose'    'drier'
'driven'

 Columns 276 through 281

  'drop'    'drought'    'during'    'eagerly'    'earlier'
'early'

 Columns 282 through 287

  'ease'    'east'    'eastern'    'easters'    'endured'
'entire'

 Columns 288 through 293

  'equatorial'    'even'    'event'    'events'    'every'
'evidence'

 Columns 294 through 298
```

```
  'excitement'    'executive'    'expect'    'expected'
'expecting'

 Columns 299 through 305

  'experts'    'fact'    'far'    'feet'    'fish'    'fishing'
'five'

 Columns 306 through 311

  'flooding'    'for'    'forecasting'    'four'    'from'
'fruitful'

 Columns 312 through 317

  'fuelling'    'future'    'generated'    'get'    'getting'
'good'

 Columns 318 through 324

  'got'    'grip'    'grounded'    'had'    'half'    'hammerhead'
 'has'

 Columns 325 through 332

  'hat'    'have'    'haven'    'he'    'heat'    'heavy'    'help'
  'here'

 Columns 333 through 338

  'historical'    'historically'    'history'    'hit'    'holiday'
  'hope'

 Columns 339 through 343

  'however'    'in'    'including'    'increased'    'increasingly'

 Columns 344 through 349

  'industry'    'influences'    'influx'    'inspired'    'into'
'is'

 Columns 350 through 357

  'isn'    'it'    'its'    'jet'    'just'    'keen'    'key'
'killed'

 Columns 358 through 364

  'know'    'lack'    'land'    'landslides'    'large'    'last'
'late'

 Columns 365 through 371
```

```
    'lean'    'least'    'less'    'like'    'likely'    'little'
'll'

 Columns 372 through 378

    'long'    'look'    'looking'    'lot'    'low'    'm'
'magnitude'

 Columns 379 through 384

    'making'    'man'    'many'    'marine'    'marlin'    'masses'

 Columns 385 through 389

    'massive'    'meaning'    'meanwhile'    'measurable'
'measured'

 Columns 390 through 395

    'measures'    'mild'    'miles'    'modelling'    'moisture'
'momentum'

 Columns 396 through 401

    'months'    'more'    'most'    'mountain'    'much'
'mudslides'

 Columns 402 through 408

    'nationwide'    'natural'    'near'    'next'    'no'    'nor'
'north'

 Columns 409 through 415

    'northern'    'not'    'now'    'number'    'numbers'    'ocean'
  'of'

 Columns 416 through 422

    'off'    'officer'    'on'    'only'    'open'    'opened'
'operate'

 Columns 423 through 428

    'operating'    'operators'    'optimism'    'or'    'other'
'our'

 Columns 429 through 435

    'out'    'over'    'overcome'    'owner'    'pack'    'panic'
'parched'

 Columns 436 through 441
```

```
  'part'    'past'    'patterns'    'penetrate'    'pent'
'people'

 Columns 442 through 446

  'peril'    'period'    'periodical'    'phenomenal'
'phenomenon'

 Columns 447 through 452

  'places'    'planned'    'playing'    'plenty'    'point'
'pointing'

 Columns 453 through 457

  'polar'    'positive'    'possibly'    'precipitation'
'predicting'

 Columns 458 through 462

  'prediction'    'previous'    'prior'    'probability'
'probably'

 Columns 463 through 468

  'proved'    'provides'    'punch'    'pushing'    'quite'
'rain'

 Columns 469 through 475

  'rainfall'    'rainiest'    'ran'    're'    'reach'    'real'
'really'

 Columns 476 through 481

  'reason'    'rebound'    'received'    'recent'    'record'
'red'

 Columns 482 through 486

  'refill'    'region'    'regions'    'relief'    'remarkably'

 Columns 487 through 491

  'reporters'    'rescue'    'resort'    'resorts'    'responsible'

 Columns 492 through 497

  'restaurants'    'retailers'    'risking'    'role'    'rolling'
 'run'

 Columns 498 through 504
```

```
  'running'    's'    'said'    'saw'    'say'    'science'
'scientist'

 Columns 505 through 511

  'scientists'    'sea'    'season'    'see'    'seeing'    'seen'
 'sees'

 Columns 512 through 517

  'several'    'sharks'    'she'    'shifted'    'should'
'signal'

 Columns 518 through 522

  'significant'    'significantly'    'since'    'single'    'ski'

 Columns 523 through 528

  'skiers'    'skiing'    'slopes'    'small'    'snow'
'snowboard'

 Columns 529 through 533

  'snowboarders'    'snowfall'    'snowpack'    'snowstorms'    'so'

 Columns 534 through 539

  'some'    'someone'    'something'    'soon'    'sorts'
'southern'

 Columns 540 through 545

  'sparked'    'spate'    'species'    'spent'    'stark'    'start'

 Columns 546 through 551

  'state'    'still'    'storms'    'strange'    'streams'
'strengthened'

 Columns 552 through 556

  'strengthening'    'strong'    'strongest'    'stuff'
'substantial'

 Columns 557 through 562

  'substantially'    'such'    'suggests'    'sure'    'surpasses'
  't'

 Columns 563 through 568

  'technology'    'temperatures'    'term'    'than'    'that'
'the'
```

```
 Columns 569 through 575

   'them'    'there'    'these'    'they'    'think'    'third'
'this'

 Columns 576 through 581

   'those'    'threat'    'three'    'thrilled'    'through'
'throughout'

 Columns 582 through 588

   'time'    'to'    'told'    'tough'    'tour'    'trapped'
'trends'

 Columns 589 through 594

   'trips'    'tropical'    'truth'    'turnaround'    'turtles'
'two'

 Columns 595 through 599

   'typically'    'unbelievable'    'uncertain'    'under'
'unlikely'

 Columns 600 through 605

   'unprecedented'    'until'    'unusual'    'up'    'upside'
've'

 Columns 606 through 611

   'view'    'visitor'    'volumes'    'vortex'    'want'    'warm'

 Columns 612 through 617

   'warmer'    'warmest'    'warming'    'warn'    'warned'    'was'

 Columns 618 through 624

   'water'    'waters'    'wave'    'way'    'we'    'weather'
'week'

 Columns 625 through 631

   'welcome'    'well'    'western'    'wetter'    'what'    'when'
 'where'

 Columns 632 through 638

   'which'    'while'    'white'    'who'    'whole'    'why'
'widespread'
```

```
  Columns 639 through 644

    'will'    'winter'    'winter's'    'with'    'won'    'wondering'

  Columns 645 through 650

    'woodwork'    'world'    'worst'    'would'    'year'    'years'

  Columns 651 through 654

    'yellowtails'    'yet'    'you'    'â##'

intersect of a and b:
  Columns 1 through 7

    ''    '11'    '2015'    '60'    'Administration'    'Atmospheric'
   'But'

  Columns 8 through 13

    'California'    'Californians'    'December'    'El'    'I'
 'January'

  Columns 14 through 19

    'National'    'NiÃ±o'    'Ocean'    'Oceanic'    'Pacific'
 'The'

  Columns 20 through 27

    'There'    'This'    'We'    'a'    'above'    'after'    'all'
 'along'

  Columns 28 through 34

    'already'    'and'    'are'    'around'    'as'    'at'
 'average'

  Columns 35 through 41

    'be'    'been'    'blamed'    'but'    'comes'    'coming'
 'current'

  Columns 42 through 47

    'director'    'during'    'event'    'expect'    'expected'
 'far'

  Columns 48 through 55

    'flooding'    'for'    'from'    'half'    'has'    'have'    'he'
    'in'

  Columns 56 through 62
```

```
    'is'    'it'    'last'    'like'    'likely'    'meaning'
  'months'

  Columns 63 through 69

    'more'    'northern'    'not'    'number'    'of'    'on'    'out'

  Columns 70 through 76

    'people'    'point'    'rain'    'rainfall'    're'    'really'
  'record'

  Columns 77 through 83

    'region'    's'    'said'    'saw'    'seen'    'should'
  'since'

  Columns 84 through 90

    'snow'    'so'    'some'    'southern'    'state'    'storms'
  'sure'

  Columns 91 through 98

    't'    'than'    'that'    'the'    'think'    'third'    'this'
   'time'

  Columns 99 through 105

    'to'    'unusual'    'warm'    'was'    'water'    'we'
  'weather'

  Columns 106 through 112

    'week'    'when'    'which'    'white'    'will'    'with'
  'year'

  Columns 113 through 114

    'years'    'â##'

Jaccard Similarity=0.174312
v.) Jaccard Similarity

Jaccards =

    1.0000    0.4000    0.6000    0.6000    0.8000    1.0000
    0.4000    1.0000    0.4000    0.7500    0.6000    0.4000
    0.6000    0.4000    1.0000    0.6000    0.8000    0.6000
    0.6000    0.7500    0.6000    1.0000    0.8000    0.6000
    0.8000    0.6000    0.8000    0.8000    1.0000    0.8000
    1.0000    0.4000    0.6000    0.6000    0.8000    1.0000
```

```
Question 3b
crc values for each shingle:
   1.0e+09 *

   0.858285995000000
   1.615091124000000
   3.766622571000000
   1.959884327000000
   3.753301509000000
   2.167986080000000
   1.248644152000000
   3.111360042000000

hashvalues for each shingle:
       67019
       45476
       63332
        6516
       78709
       83684
        5719
        9903
```

Published with MATLAB® R2016b