STAT 154: Homework 2

Release date: **Thursday, February 7**

Due by: **11 PM, Wednesday, February 20**

Submission instructions

It is a good idea to revisit your notes, slides and reading; and synthesize their main points BEFORE doing the homework.

A .Rnw file corresponding to the homework is also uploaded for you. You may use that to write-up your solutions. Alternately, you can typeset your solutions in latex or submit neatly handwritten/scanned solutions. However for the parts that ask you to implement/run some R code, your answer should look something like this (code followed by result):

```
myfun<- function(){
show('this is a dummy function')
}
myfun()

## [1] "this is a dummy function"
```

Note that this is automatically generated if you use the R sweave environment.

You need to submit the following:

1. A pdf of your write-up to "HW2 write-up".
2. A Rmd or Rnw file, that has all your code, to "HW2 code".

Ensure a proper submission to gradescope, otherwise it will not be graded. Make use of the first lab to clear all your doubts regarding the submission/gradescope.

The honor code

(a) Please state the names of people who you worked with for this homework. You can also provide your comments about the homework here.

(b) Please type/write the following sentences yourself and sign at the end. We want to make it *extra* clear that nobody cheats even unintentionally.

I hereby state that all of my solutions were entirely in my words and were written by me. I have not looked at another students solutions and I have fairly credited all external sources in this write up.

This homework revisits principal component analysis (PCA) and related computations.

1 A few basics of SVD (10*3 = 30 points)

Singular value: Given a matrix $M \in \mathbb{R}^{(m \times n)}$ (assume $m \geq n$). The non-zero singular values of M correspond to the square roots of the non-zero eigenvalues of either $M^\top M$ or $MM^\top$.

Singular value decomposition: For real matrix in finite dimension, it is always to write M in the following decomposed form

$$M = UDV^\top, \tag{1}$$

where

- U is an $m \times n$ matrix of left singular vectors.
- D is a $n \times n$ diagonal matrix of singular values.
- V is a $n \times n$ matrix of right singular vectors.

(a) Show that $M = \sum_{i=1}^{n} d_i u_i v_i^\top$ where $d_i = D_{ii}$ is the i-th singular value, u_i and v_i are the i-th left and right singular vectors (column vectors) respectively.

(b) For $1 \leq i \leq n$, show that the i-th eigenvalue of $M^\top M$ is given by d_i^2 with the corresponding eigenvector v_i. And show that the i-th eigenvalue of $MM^\top$ is given by d_i^2 with the corresponding eigenvector u_i.

(c) Generate a random matrix M of size $n \times n$ for $n \in \{2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048\}$. And then plot the time taken to (i) to generate this matrix, and (ii) to compute the svd of the matrix as n increases. Note that you need to plot two figures. You may want to use *Sys.time* for the same. DO NOT report the SVD values, just plot the time as a function of n. Do you see some scaling with n? Justify your observations. It may be useful to plot the values on a log-log scale to get a better idea.

2 Power Method

The Power Method is an iterative procedure for approximating eigenvalues. First assume that the matrix A has a dominant eigenvalue with corresponding dominant eigenvector. Then choose an initial approximation w_0 (must be a non-zero vector) of one of the dominant eigenvectors. This choice is arbitrary (and in theory should work with almost any vector). Then, form the sequence w_1, w_2, . . . , w_k, given by:

$$\begin{aligned} w_1 &= Aw_0 \\ w_2 &= Aw_1 \\ &\vdots \\ w_k &= Aw_{k-1} = A^{k-1}w_0 \end{aligned}$$

For large powers of k, and by **properly scaling** this sequence, you will see that you obtain a good approximation w_k of the dominant eigenvector of A.

2.1 First eigenvector and eigenvector

Here is the full procedure of the Power Method to find the largest eigenvalue and its corresponding eigenvector. Write code in R to implement such method (Implementation code required).

(a) Start with an arbitrary vector w_0

(b) Iteration for a series of steps $k = 0, 1, 2, ..., n$ to form the series of w_k vectors: $w_{k+1} = \frac{Aw_k}{s_{k+1}}$. where s_{k+1} is the entry of Aw_k which has the largest absolute value (this is actually a scaling operation dividing by the L_∞-norm).

(c) When the scaling factors s_k are not changing much, s_{k+1} will be close to the largest eigenvalue of A, and w_{k+1} will be close to the eigenvector associated to s_{k+1}.

(d) You can also veirfy that s_{k+1} will be very close to the eigenvalue given by the Rayleigh quotient:

$$\lambda \sim \frac{w_{k+1}^\top A w_{k+1}}{w_{k+1}^\top w_{k+1}}$$

(e) Typically, once youve obtained w_{k+1}, it is re-scaled in such a way that its euclidean norm is 1, that is: $\|w_{k+1}\|_2 = 1$.

Use your code to find the largest eigenvalue of the following matrix.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -1 & 4 \\ 3 & 4 & -5 \end{bmatrix}$$

Compare your results with those provided by **eigen()**. Keep in mind that the eigenvectors of **eigen()** have unit Euclidean norm (i.e. L2-norm). Likewise, recall that the eigenvectors are only defined up to a constant: even when the length is specified they are still only defined up to a scalar.

2.2 Deflation and more eigenvectors

When a matrix A is symmetric, you can use the power method to get more eigenvectors and eigenvalues. How? You need to apply the Power Method on the residual matrix obtained by deflating A with respect to the first eigenvector. This deflation operation is:

$$A_1 = A - \lambda_1 v_1 v_1^\top,$$

where λ_1 is the first eigenvalue and v_1 is the corresponding eigenvalue.

Consider the matrix

$$B = \begin{bmatrix} 5 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(a) Apply the your Power Method on B to get an approximation of the first eigenvector and eigenvalue.

(b) Deflate the matrix B and apply the power method on the residual matrix B_1 to obtain the second eigenvalue and eigenvector.

(c) Deflate the matrix B_1 again and apply the power method to obtain the third eigenvalue and eigenvector.

3 Principal Component Analysis

Recall that the principal components correspond to the eigenvectors of the covariance matrix of the data.

We will use the PCA on the **USArrests** dataset. Take a look at the R documentation with **?USArrests**. For each of the 50 states in the United States (50 rows), the data set contains the number of arrests per 100, 000 residents for each of three crimes: Assault, Murder, and Rape. We also record UrbanPop (the percent of the population in each state living in urban areas).

(a) Use **apply()** function to compute mean and variance of all the four columns

(b) Plot a histogram for each of the four columns

(c) Do you see any correlations between the four columns? Plot and comment

(d) Use **prcomp()** function to perform principal component analysis. Make sure you standarlized the data matrix. Print a **summary** at the end.

(e) Obtain the principal vectors and store them in a matrix, include row and column names. Display the first three loadings.

(f) Obtain the principal components (or scores) and store them in a matrix, include row and column names. Display the first three PCs.

(g) Obtain the eigenvalues and store them in a vector. Display the entire vector, and compute their sum.

(h) Create a scree-plot (with axis labels) of the eigenvalues. What do you see? How do you read/interpret this chart?

(i) Create a scatter plot based on the 1st and 2nd PCs. Which state stands out? Provide some explanations. In this plot you should annotate the points with state names.

(j) Create the same scatter plot but color the states according to the variable UrbanPop.

(k) Create a scatter plot based on the 1st and 3rd PCs. Comment on the difference between this plot and the previous one.

4 K-means and PCA

ISL book: Problems 1 (on K-means) and 10 (PCA and K-means) from Exercises 10.7 (Chapter 10).

5 True or false (10*7 = 70 points)

Examine whether the following statements are true or false and *provide one line justification.*

(a) Eigenvalues obtained from the principal component analysis is always nonnegative.

(b) The first principal vector and the second principal vector are always orthogonal.

(c) Singular values of a square matrix M are the same as the eigenvalues of M.

(d) Principal components analysis can be used to create a low dimensional projection of the data.

(e) Eigenvalue of a matrix are always nonnegative.

(f) The y-axis of a Scree plot is always from 0 to 1.

(g) The maximum number of principal components is always less or equal to the feature dimension.