STAT 154: Homework 1

Release date: **Thursday, January 24**

Due by: **6 PM, Wednesday, February 6**

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 "HW1 write-up".
2. A Rmd or Rnw file, that has all your code, to "HW1 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 several concepts from matrix algebra, besides doing some EDA with R, and, some concepts from the lectures.

1 A few basics (10*9 = 90 points)

Define the following quantities:

(a) Linear space spanned by a given set of vectors $S = \{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$

(b) Column space of a matrix

(c) Rank of a matrix.

Answer the following:

(d) When is a square matrix called a singular matrix?

(e) When is a square matrix called an orthogonal matrix?

(f) When is a matrix said to have full column rank?

Recall that a symmetric matrix $\mathbf{A} \in \mathbb{R}^{d\times d}$ is *positive semi-definite* (PSD) if $\mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0$ for any $\mathbf{x} \in \mathbb{R}^d$. Now answer the following:

(g) Show that the matrix $\mathbf{A} = \mathbf{a}\mathbf{a}^\top$, where $\mathbf{a} \in \mathbb{R}^d$ is nonzero, is PSD.

(h) What is the rank of $\mathbf{A} = \mathbf{a}\mathbf{a}^\top$?

(i) For any matrix $\mathbf{C} \in \mathbb{R}^{d\times d}$, is the matrix $\mathbf{C}\mathbf{C}^\top$ a PSD matrix?

2 Eigendecomposition with R (10*6 = 60 points)

Consider the following matrices

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{Y} = \begin{bmatrix} -1/3 & 2/3 \\ 2/3 & -1/3 \end{bmatrix}$$

(a) Create these matrices in R and show them.

(b) Which of the two matrices are invertible and why? Find the inverse if it exists. How does one compute the inverse in R?

(c) Compute the eigenvalues and eigenvectors of the matrix $\mathbf{X}$. We expect you to be able to do this by hand but you may still use R for this part.

(d) Use the results of previous part to directly compute the eigenvalues and eigenvectors of the matrix $\mathbf{Y}$.

(e) What are the eigenvalues and eigenvectors of $\mathbf{X}^2$? You may use direct computations or R to answer this question.

(f) Prove that if (λ, v) denotes an eigenvalue-eigenvector pair for the matrix $\mathbf{X}$, then the pair (λ^2, v) is a valid eigenvalue-eigenvector pair for the matrix $\mathbf{X}^2$.

3 Understanding orthogonal projection (10*12+40 = 160 points)

In this problem, we will understand orthogonal projection in quite some detail. The length of the problem should not bother you as several parts are relativey easy.

(a) For any vector $\mathbf{x} = [x_1, x_2, x_3]^\top \in \mathbb{R}^3$, what is its projection on the vector $[1, 0, 0]^\top$?

(b) What is the orthogonal projection of $\mathbf{x} = [x_1, x_2, x_3]^\top \in \mathbb{R}^3$ on the subspace spanned by the vectors $[1, 0, 0]^\top$ and $[0, 1, 0]^\top$?

(c) Write the projection matrices in the previous two cases.

(d) Create the previous projection matrices in R and compute the projections of two random vectors whose entries are drawn independently from the uniform distribution on $[0, 1]$. Do not forget to set the seed.

(e) Let's make the problem more general. Write the expression for the orthogonal projection of a vector $\mathbf{x} \in \mathbb{R}^d$ along any given vector $\mathbf{a} \in \mathbb{R}^d$. What is the projection matrix in this case?

(f) Create a function in R that takes in input two vectors $\mathbf{x}, \mathbf{a}$ and computes the projection from previous part. Call the function with $\mathbf{x} = [3, 2, -1]^\top$ and $\mathbf{a} = [1, 0, 1]^\top$.

(g) Given two orthogonal vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ in $\mathbb{R}^d$, what is the orthognal projection of a generic vector $\mathbf{x}$ on to the subspace spanned by these two vectors?

(h) Now let's make the problem a bit more challenging. Suppose that the two vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are not orthogonal. How will you compute the orthogonal projection of $\mathbf{x}$ in this case? It may be useful to revise Gram Schmidt Orthogonalization.

(i) Implement your method in the previous part in R for $\mathbf{x} = [3, 2, -1]^\top$ and $\mathbf{a}_1 = [1, 0, 1]^\top$ and $\mathbf{a}_2 = [1, -1, 0]^\top$.

(j) Can you generalize the answer from previous part to the case to compute the orthogonal projection along a k-dimensional subspace spanned by the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ which need not be orthogonal?

(k) **Challenging (40 points):** Define the matrix

$$\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_k] \in \mathbb{R}^{d \times k}$$

such that the columns are independent. Then the orthogonal projection of any vector $\mathbf{x} \in \mathbb{R}^d$ onto the k-dimensional subspace spanned by the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ is given by

$$\mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{x}. \tag{1}$$

Hint: Consider the least squares problem

$$\min_{y} \|\mathbf{A}\mathbf{y} - \mathbf{x}\|_2^2.$$

(l) Compute the projection for $\mathbf{x} = [3, 2, -1]^\top$ on the space spanned by the vectors $\mathbf{a}_1 = [1, 0, 1]^\top$ and $\mathbf{a}_2 = [1, -1, 0]^\top$ using the expression in previous part using R. Compare it with your answer from part (i).

(m) How does the expression in the equation (1) simplify if the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are orthogonal?

4 Exploring a dataset with R (10*8 + 2*20 = 120 points)

ISL (with applications in R): Chapter 2, Problem 10 (Boston dataset) After parts (a)-(h), please answer the following parts that try to emphasize a few aspects of the PQRS framework:

(i) What do you think defines the population for this dataset?

(j) Can you define a prediction problem that you may use this dataset for? Describe it in the framing of the three-circle representation discussed in the class.

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

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

(a) Cross validation is a powerful tool to select hyper-parameters in several machine learning tasks.

(b) Cross validation error is always a good proxy for the prediction error.

(c) Vanilla cross validation is a good idea for time-series data.

(d) For a machine learning problem, exploratory data analysis by itself is generally sufficient to determine the complete relevance of the dataset for the problem.

(e) Data collection process usually has no-to-little influence on the outcome of a prediction problem.

(f) For a model to make meaningful predictions on the future data, we need some similarity between the representative data that was used to build the model and the future data.

(g) Prediction is often the end goal of a machine learning task.