Northeastern University

DS5020 - Introduction to Linear Algebra and Probability for Data Science

Assignment 5

1. {20 pts} Suppose

$$\mathbf{A} = \begin{bmatrix} 3 & -3 \\ 4 & -4 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -1 \\ 7 \\ 2 \end{bmatrix}$$

a. Find the least squares solution to $\mathbf{Ax} = \mathbf{b}$. Recall that least squares solution is the solution to $\mathbf{A}^{\mathrm{T}}\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^{\mathrm{T}}\mathbf{b}$. Find the error vector $\mathbf{e} = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$ and show that it is orthogonal to $\mathbf{A}\hat{\mathbf{x}}$.

b. Using Gram-Schmidt process on columns of $\mathbf{A}$, find two orthonormal vectors $\mathbf{q}_1$ and $\mathbf{q}_2$ to form orthonormal matrix $\mathbf{Q} = [\mathbf{q}_1 \quad \mathbf{q}_2]$. Double-check that the matrix you found is correct by showing $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{I}$.

c. Find QR-decomposition of $\mathbf{A}$, $\mathbf{A} = \mathbf{QR}$. Recall the general formula for QR-decomposition:

$$\underbrace{\begin{bmatrix} | & | & & | \\ | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ | & | & & | \\ | & | & & | \end{bmatrix}}_{\mathbf{A}} = \underbrace{\begin{bmatrix} | & | & & | \\ | & | & & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & & | \\ | & | & & | \end{bmatrix}}_{\mathbf{Q}} \underbrace{\begin{bmatrix} (\mathbf{q}_1^{\mathrm{T}}\mathbf{c}_1) & (\mathbf{q}_1^{\mathrm{T}}\mathbf{c}_2) & \cdots & (\mathbf{q}_1^{\mathrm{T}}\mathbf{c}_n) \\ 0 & (\mathbf{q}_2^{\mathrm{T}}\mathbf{c}_2) & \cdots & (\mathbf{q}_2^{\mathrm{T}}\mathbf{c}_n) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\mathbf{q}_n^{\mathrm{T}}\mathbf{c}_n) \end{bmatrix}}_{\mathbf{R}}$$

Hint: Note that you already found $\mathbf{Q}$ *in part* **b**. *You are now asked to find the terms* $\mathbf{q}_i^{\mathrm{T}}\mathbf{c}_i$ *to form the upper triangular matrix* $\mathbf{R}$.

d. Plug $\mathbf{A} = \mathbf{QR}$ in least squares system of equations $\mathbf{A}^{\mathrm{T}}\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^{\mathrm{T}}\mathbf{b}$, and multiply both sides from left with $(\mathbf{R}^{\mathrm{T}})^{-1}$ to get equivalent system of equations $\mathbf{R}\hat{\mathbf{x}} = \mathbf{Q}^{\mathrm{T}}\mathbf{b}$. Use generic letters $\mathbf{Q}$ for orthonormal matrix and $\mathbf{R}$ for upper triangular matrix instead of plugging in numbers from this particular example.

e. Solve $\mathbf{R}\hat{\mathbf{x}} = \mathbf{Q}^{\mathrm{T}}\mathbf{b}$ to find least squares solution $\hat{\mathbf{x}}$ using QR-decomposition of $\mathbf{A}$ you found in part **c**. Verify that this solution is the same as the solution in part **a**.

2. {20 pts} Suppose that an F1 pit crew measures the speed of the car at four times $\mathbf{t} = 1, 2, 3, 4$ and get the following readings: $\mathbf{s} = 163, 186, 195, 198$ km/h, respectively.

a. Suppose that the crew models the speed as a linear function of time first and tries to fit a line to the data points they read: $s_i = a + bt_i$. Using the least squares find the constants a and b that minimize the length of error vector $\mathbf{e} = \mathbf{s} - \hat{\mathbf{s}}$, where $\hat{\mathbf{s}}$ is the least squares estimate of the speed ($\hat{s}_i = a + bt_i$).

b. Find the error vector $\mathbf{e}$ for the line fit in part **a**.

c. Realizing that the error they found in part **b** is too high, the pit engineers go back and try to better model the speed as a function of time. This time, they try a logarithmic function fit : $s_i = a + b(1 - e^{-t_i})$. The corresponding system of equations they need to solve then becomes:

$$\begin{bmatrix} 1 & 0.63 \\ 1 & 0.86 \\ 1 & 0.95 \\ 1 & 0.98 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 163 \\ 186 \\ 195 \\ 198 \end{bmatrix} \tag{1}$$

Solve above system of equations using least squares to find a and b.

d. Find the error vector $\mathbf{e}$ for the logarithmic fit in part **c**. If you find that this error vector is $zero - vector$, what does this mean in terms of existence of a solution to the system of equations shown in (1)?

3. {20 pts} Suppose the transition matrix $\mathbf{T}$ below describes the percentage of people moving from cities to suburbs and visa versa.

$$\mathbf{T} = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix}$$

In words, 20% of city population moves to the suburbs and 10% of population in suburbs move to the city in unit time. Suppose that 1 million people live in the cities and 1 million people live in the suburbs at time $t = 0$, that is population distribution vector $\mathbf{p}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

a. Find the eigenvalues and eigenvectors of $\mathbf{T}$.

b. Find the population distribution vector at time $t = 1$, $\mathbf{p}_1$.

c. Find the population distribution vector at time $t = 100$, $\mathbf{p}_{100}$.

4. {20 pts}

a. Prove that the two eigenvalues of a general 2x2 transition matrix $\mathbf{T} = \begin{bmatrix} \alpha & 1-\beta \\ 1-\alpha & \beta \end{bmatrix}$ are 1 and $\alpha + \beta - 1$.

b. Prove that vectors $\mathbf{v}_1 = \begin{bmatrix} \frac{1-\beta}{1-\alpha} \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ are eigenvectors for transition matrix $\mathbf{T}$ in part **d**.

c. Given that $0 < \alpha < 1$ and $0 < \beta < 1$, show that vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent. This means that we can uniquely represent <u>any</u> population distribution vector $\mathbf{p}$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$: $\mathbf{p} = c\mathbf{v}_1 + d\mathbf{v}_2$.

d. For a starting population distribution vector $\mathbf{p}_0 = c\mathbf{v}_1 + d\mathbf{v}_2$, find the population distribution vector as time goes to infinity, $\mathbf{p}_\infty$.

5. {20 pts} Suppose

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

a. Find the eigenvalues of $\mathbf{A}$.

b. Find the eigenvectors for each eigenvalue of $\mathbf{A}$.

c. Is $\mathbf{A}$ diagonalizable, why?

d. If your answer to part **c** is *yes* (and it should be), diagonalize $\mathbf{A}$. That is find matrices $\mathbf{X}$ and $\mathbf{\Lambda}$ such that $\mathbf{\Lambda} = \mathbf{X}^{-1}\mathbf{A}\mathbf{X}$.

e. Find $\mathbf{A}^{100}$ in terms of matrices $\mathbf{X}$ and $\mathbf{\Lambda}$.