1. We discussed the restaurant problem in the last class. Recall that every day, $1-p$ fraction of students moved from Chinese restaurant to the Mexican restaurant and $1-q$ fraction of students moved from Mexican to Chinese restaurant. Put formally the transition matrix is given by
$$\begin{bmatrix} p & 1-q \\ 1-p & q \end{bmatrix}$$
Now consider two different scenarios given by the following transition matrices.
$$S1 = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} \quad S2 = \begin{bmatrix} 0.6 & 0.7 \\ 0.4 & 0.3 \end{bmatrix}$$
Which of the two scenarios converge to a steady state faster ? Give a formal proof justifying your answer.

   Solution:

2. Prove that the largest eigen value of a stochastic matrix is 1.

   Solution:

3. Prove the following statement: If a set of vectors $\{d_1, d_2, \ldots d_n\}$ are pairwise H-orthogonal, then they are linearly independent. Recall that for pairwise H-orthogonal vectors, $d_i^T H d_j = 0 \; \forall i, j \in 1, 2, \ldots n; i \neq j$

   Solution:

4. We studied PCA in last class. Recall that $X \in \mathbb{R}^{m \times n}$ was the matrix of m points in n dimensional space. We said that if the matrix had zero mean and unit variance along individual dimensions then $X^T X$ was the covariance matrix. Turns out that is not correct.
Let x_{ij} denote the j-th dimension of point i. ($j \in \{1, \ldots n\}$ and $i \in \{1, \ldots m\}$)
We build a new matrix $\widehat{X}$ with
$$\hat{x}_{ij} = \frac{1}{a}\Big(x_{ij} - \frac{1}{b}\sum_{k=0}^{m} x_{ik}\Big)$$
What are the values of a and b you should use to ensure that $\widehat{X}^T \widehat{X}$ is the covariance matrix ?

Solution:

5. Gradient based optimization algorithms have the following template : (i) **Initialize parameters randomly to $\boldsymbol{\theta_0}$** (ii) Iteratively update the parameters using a certain rule until convergence. The point to be noted here is that the random initialization can affect the number of steps needed to converge. Now consider the following optimization problem. minimize $x^2 + y^2 + z^2 - 8$. Say we choose to use Newton's method for updating the parameters and we run the optimization algorithm separately thrice with the following initializations (x_0, y_0, z_0) : (i) (-3, 0, -1) (ii) (1, 1, 1) (iii) (1, 0, 1). What are the number of steps needed for convergence with each of the above initializations ?

Solution:

6. Compute the Frobenius norm of the following matrix

$$S1 = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 1 \\ 1 & 9 & 0 \end{bmatrix}$$

Solution: