1. For this problem, you are asked to use the data in the `Pizza.csv` file (the same file used for HW #4). The feature values are contained in columns 3 through 9. Construct an autoencoder with 3 layers (i.e., one hidden layer for the code). Let $\mathbf{X}$ be the matrix that contains the samples as its rows. Similarly, the rows of $\hat{\mathbf{X}}$ are the reconstructed inputs from the autoencoder. You will evaluate the performance of autoencoders using the mean squared error (MSE) given below as you vary the dimension of the code:

$$MSE(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{n}\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2, \tag{1}$$

   where n is the number of samples, and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix.

   (a) Design an optimal autoencoder with the linear encoder and decoder without nonlinear activation functions, which minimizes the MSE. Compute the MSE as a function of the dimension of the code $h \in \{1, 2, \ldots, 6\}$.

   (b) Train an autoencoder with the ReLU activation function using the data with varying size of the code h. Plot the MSE as a function of h.

2. Consider the optimal autoencoder with linear encoder and decoder in problem 1(a) above. For a fixed code dimension h, explain how the MSE in (1) of the optimal autoencoder can be computed from the singular values of $\mathbf{X}$.