Work out the following problems without the use of a computer.

1) After calculating the covariance matrix of a feature matrix $X$, we get the eigenvalues as listed below:

$$[12.43, 10.56, 5.50, 4.21, 0.53, 0.10, 0.01].$$

How many principal components would you use so that at least 99% of the variance is retained? 95%? 90%? (15 points)

2) For the graphs below, do the following:
   a. Draw the line that results when we perform PCA that minimizes the projection error. (3 points)
   b. Draw the approximate transformed values ($z$) that we get by transforming the original data values with PCA. (3 points)
   c. Draw the reconstructed data set. (4 points)
3) Suppose we develop a linear stock model of the ABC toy company to predict tomorrow's future price of a stock:

\[ w^T x = 30 + 2x_1 - 3x_2. \]

a. If the current price of the stock is $60, specify a perceptron model which returns a 1 when the stock is predicted to go up or remain unchanged and returns a 0 when the stock is predicted to go down. (5 points)

b. What value will the perceptron model predict when \(x_1 = 20, x_2 = 2\)? When \(x_1 = 15, x_2 = 3\)? (5 points)

c. With the perceptron model, update the weights when we have the training observation: \(y = 1, x_1 = 20, x_2 = 2\) with \(\eta = 0.1\) and the perceptron learning rule: (5 points)

\[ w_j := w_j + \Delta w_j \]

\[ \Delta w_j = \eta(y^{(i)} - \hat{y}^{(i)})x_j^{(i)}. \]

d. Repeat part c with the training observation: \(y = 1, x_1 = 15, x_2 = 3\). (5 points)
4) Solve the following minimization problem using stochastic gradient descent with 2 epochs

\[ J(w) = \frac{1}{2}w_0^2 + \frac{1}{2}(w_0 + w_1)^2 \]

starting from \( w^{(0)} = (0.1, 0.1) \) with learning rate/step size 0.001. (15 points)
5) For this problem, we will be working with the MNIST data set and making submissions to Kaggle. You can view the competition and download the data from “kaggle” at https://www.kaggle.com/c/digit-recognizer. For the competition, you will be applying KNN with \( K = 3 \) to the data set and submitting your predictions to kaggle. However, you will first need to perform PCA on the data set given the amount of variance to be retained. Also, you should normalize the data using the standard scaler before performing PCA on the data set. For each problem, hand in your code (10 points total) and specify the number of Principal components (5 points each) and specify the classification accuracy (5 points each).

a. Retain 90% of the variance and specify:

   number of principal components used: ____________

   classification accuracy: ____________

b. Retain 95% of the variance and specify:

   number of principal components used: ____________

   classification accuracy: ____________

c. Retain 99% of the variance and specify:

   number of principal components used: ____________

   classification accuracy: ____________