Evaluation Questions

1. You are required to create a Python dictionary with two keys: data and target. The data key should store a 28x28 matrix filled with random integer values between 0 and 255 (both values included), representing pixel intensities in a grayscale image. The target key should store a binary label (either 0 or 1, chosen randomly). Then print the shape and type of the dictionary elements.

(**25 mins**)

Steps:

    i.   Use the randint function from Python's random module to generate the 28x28 matrix of random integer values ranging between 0 and 255, both ends included. You can use internet to look at documentation of randint, and not for any other purpose.

    ii.   Convert the data generated in step 1 to a numpy array.

    iii.   Assign a binary label (either 0 or 1, chosen randomly) as the target.

    iv.   Construct the dictionary with the following structure:

        data: The 28x28 matrix generated in step 1.

        target: label obtained in step 2.

    v.   Print the shape and type of dictionary elements.

2. You can go through the logistic regression docs. Then do the following. (**1 hr**)

    [Note: For this exercise, you can continue from the end in the same notebook you submitted; you can reuse the code you have already written wherever required.]

    i.   Implement a logistic regression model for 5 versus non-5 mnist digit classification wherein the data imbalance must be handled automatically. Keep a hold out test set of 200 samples wherein 100 samples are from class 5 and 100 samples are from non-5 classes. The choice of choosing the 100 samples from each class is left to you.

    ii.   Record the average train and test set accuracy.

    iii.   Incorporate L1 regularization to the above model. Choose the regularization parameter based on the results in step (ii).

    iv.   Record the average train and test set accuracy.

    v.   Incorporate L2 regularization to the above model. The regularization parameter chosen in step (iii) can be used here.

    vi.   Record the average train and test set accuracy.

    vii.   Print the number of parameters of the model strictly less than 1e-5 in both L1 and L2 regularization implementations.

3. Speak for 2 minutes on your attempt in question 2. (**2 mins**)