Practice Questions

All the programming questions use **MNIST** dataset. Store all the samples in X and labels in y

# Common imports

import numpy as np

import mkdtemp from shutil…..

Do other common imports for feature scaling, pipeline utility and classifiers.

Fetch openml MNIST dataset

Split the dataset in the following ratio.

1.Training : Take the first 70% of samples from X and store them in x\_train

2.Testing: Take the remaining 30% of samples from X and store them in x\_test

3.Store the respective labels in y\_train, y\_test respectively.

The last training sample is of digit\_\_\_\_?

* Build a classifier that differentiate digit 6 from digit 9.
  1. Steps to be followed
  2. Collect all digit-6 (Positive class) and digit-9 (Negative class) images from x\_trainand stack them properly as a single datamatri x\_train\_69.
  3. Keep all digit-6 images from index 0 to i followed by digit-9 images from index i+1 ton (i denotes the end index of digit-6 images)
  4. Similarly, collect the respective labels and store it in a variable y\_train\_69
  5. Set the label values to 1 for positive classes and 0 for negative classes.
  6. Load from sklearn.utils import shuffle
  7. Shuffle the data matrix and labels. (Set random\_state value to 1729).
  8. Create x\_test\_69 and y\_test\_69 by repeating the steps from 1 to 6 with required modifications
  9. What is the sum of all the labels in the vector y\_train\_69?

1. What is the sum of all the labels in the vector y\_test\_69?

* Apply StandardScaler to all the training samples in x\_train\_69 and store the result in another variable (say, x\_train\_69Tf).

What is the mean of the zeroth sample?\* What is the mean of zeroth feature?\* What is the standard deviation of the zeroth sample?\* What is the standard deviation of the zeroth feature?

1. Pack the answers (in order) in a tuple.

* Train the LogisticRegression model using SGDClassifier() with the following commonsettings.

1. No Regularization
2. random\_state : 10
3. Iteration : 10
4. Capture the loss for each iteration and plot the iteration vs loss curve. For which of the following settings, the iteration vs loss curve decreased monotonically?
5. A. Set Learning rate : 0.01 and plot the curve and fit the model with `x\_train\_69 `B.Set learning rate to 0.000001 and fit the model with `x\_train\_69 ` .