# Lab 2

Lab 2 will consist of 3 tasks:

# Task 1 Artificial Neural Network (40 Points)

1.  Load the dataset (CIFAR – 10) from the given link:
    https://www.cs.toronto.edu/~kriz/cifar.html (2 points)

    This dataset consists of 60,000 images of 10 different classes, 6,000 images per class and 50,000 (5,000 per class in training) and 10,000 (1,000 per class in testing)

Here are the classes in the dataset, as well as 10 random images from each:
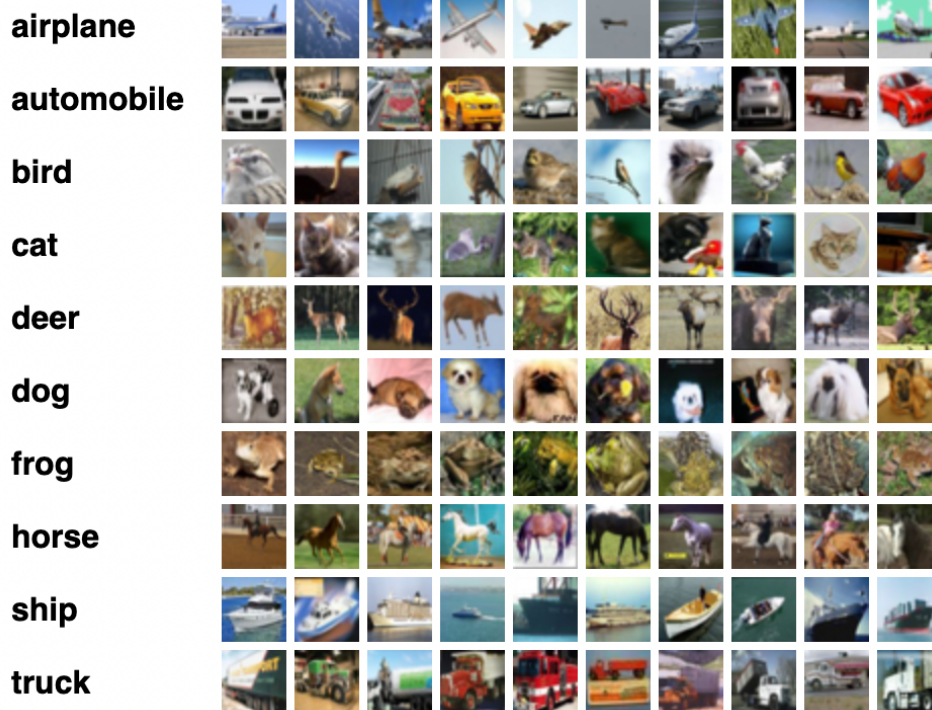


2.  Explain in your own words: (7 points)
    a.  Gradient Descent (1 point)
    b.  Drop out (1 point)
    c.  Activation functions (1 point)
    d.  Back Propagation (1 point)
    e.  Epochs, Iterations and Batch size (3 points)

3. Visualize / summarize the data (10 points)
    a. Number of entities in training and testing set and number of classes in target variable (1 point)
    b. Number of pixels in the image (Height and width individually) (2 points)
    c. Number of images per class (1 point)
    d. Display at least 2 images of each class (6 points)
4. Train a neural network (21 points)
    a. Decide number of layers and neurons in each layer (2 points)
    b. Try out different activation functions (explain each one you used) (4 points)
    c. Try with and without using drop out (4 points)
    d. Try different regularizations apart from drop out (4 points)
    e. Try different optimization algorithms (such as Gradient Descent, Adam etc.) (4 points)
    f. Create a graph of f1 score vs epochs for training and validation set. (2 points)
    g. Calculate number of trainable parameters in your final model. (1 point)

# Task 2 Natural Language Processing (20 Points)

1. Load the movie reviews sentiment analysis dataset and split into 80:20 ratio for training and test data (2 points)
2. What is lexical vs semantic text analysis (3 points)
3. Perform required preprocessing on the dataset (removing stop words, vectorization etc.) (10 points)
4. Build a model to classify the rows (5 points).

# Task 3 Recommender System (20 Points)

1. Download m1.zip file from the link (https://grouplens.org/datasets/movielens/1m/) (0 points)
2. Load the movies and ratings data (1 points)
3. What do you mean by Singular Value Decomposition (2 points)?
4. What do you mean by Principal Component Analysis (2 points)?
5. Explain content-based vs collaborative recommendation (2 points)
6. Create m x u matrix with movies as row and users as column. Normalize the matrix. (2 points)
7. Perform SVD to get U, S and V (2 points)
8. Select top 50 components from V.T (1 point)
9. Calculate the covariance matrix for the entire dataset (from step 6) (1 point)
10. Get the eigen vectors from the covariance matrix (1 point)
11. Get the top 50 eigen vectors using eigen values (1 point)

12. Using cosine similarity find 10 closest movies using the 50 components from SVD (step 8) (2 points)
13. Using cosine similarity find 10 closest movies using the 50 components from PCA (step 11) (2 points)
14. Compare the results of above two methods (1 point)

# Task 4 Random Forest – Self implementation (10 Points)

1. Complete the functions in the given Random Forest ipynb file.

**Create a detailed report of the lab – 10 Points.**