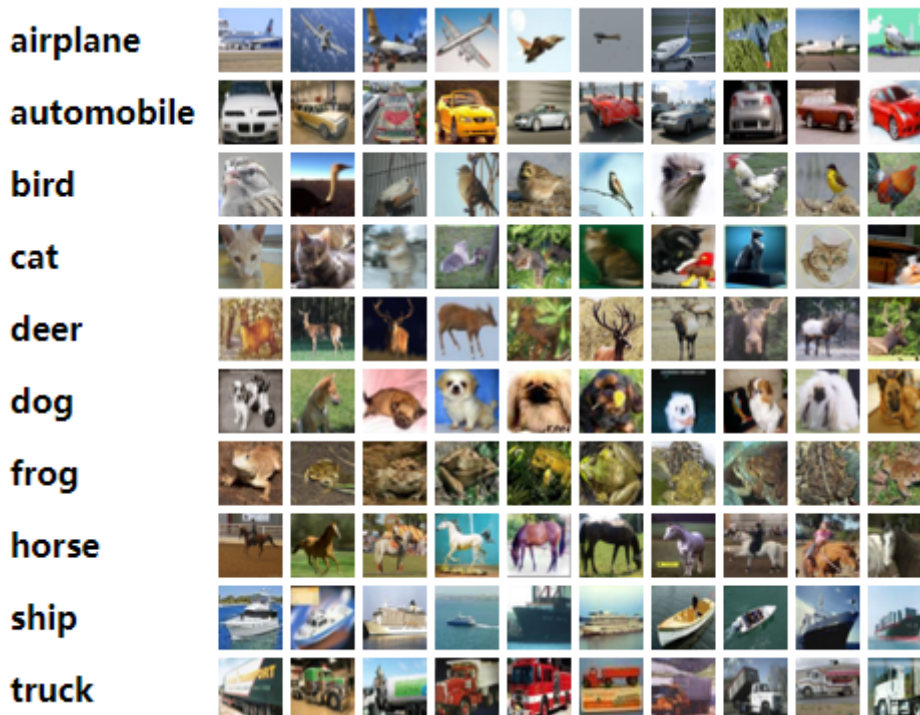


The CIFAR-10 dataset

CIFAR-10数据集由60000张32x32的彩色图片组成，包含10类图片，每一类图片均包含6000张图片。采用随机划分的方式，将60000个样本划分成了50000个训练样本和10000个测试样本。

图片为RGB彩色图片，即图片数据有三个颜色通道，所以每一张图片由32x32x3（长x宽x通道）个像素点组成，在数据集中是将多维矩阵展开存放，即 $60000 \times 32 \times 32 \times 3 \rightarrow 60000 \times 3072$ ，需要同学们自行reshape成想要的形状。

下面是数据的十个类别：



数据存放在Assignment2/material/data文件夹中，总共为6个文件，可以使用一下代码进行读取：

```
def load_data(dir):
    import pickle
    import numpy as np
    X_train = []
    Y_train = []
    for i in range(1, 6):
        with open(dir + r'/data_batch_' + str(i), 'rb') as fo:
            dict = pickle.load(fo, encoding='bytes')
            X_train.append(dict[b'data'])
            Y_train += dict[b'labels']
    X_train = np.concatenate(X_train, axis=0)

    with open(dir + r'/test_batch', 'rb') as fo:
        dict = pickle.load(fo, encoding='bytes')
    X_test = dict[b'data']
    Y_test = dict[b'labels']

    return X_train, Y_train, X_test, Y_test
```

需要传入本地数据所在的目录，例如D:/Assignment2/material/data

你需要做的：

- 将数据读取进内存，根据所要训练的模型（Softmax 分类器，多层感知机MLP和卷积神经网络CNN）对数据形状进行调整。
- 在训练集（X_train, Y_train）上对模型进行训练。
- 在测试集（X_test, Y_test）上对模型性能进行评估。