The American University in Cairo

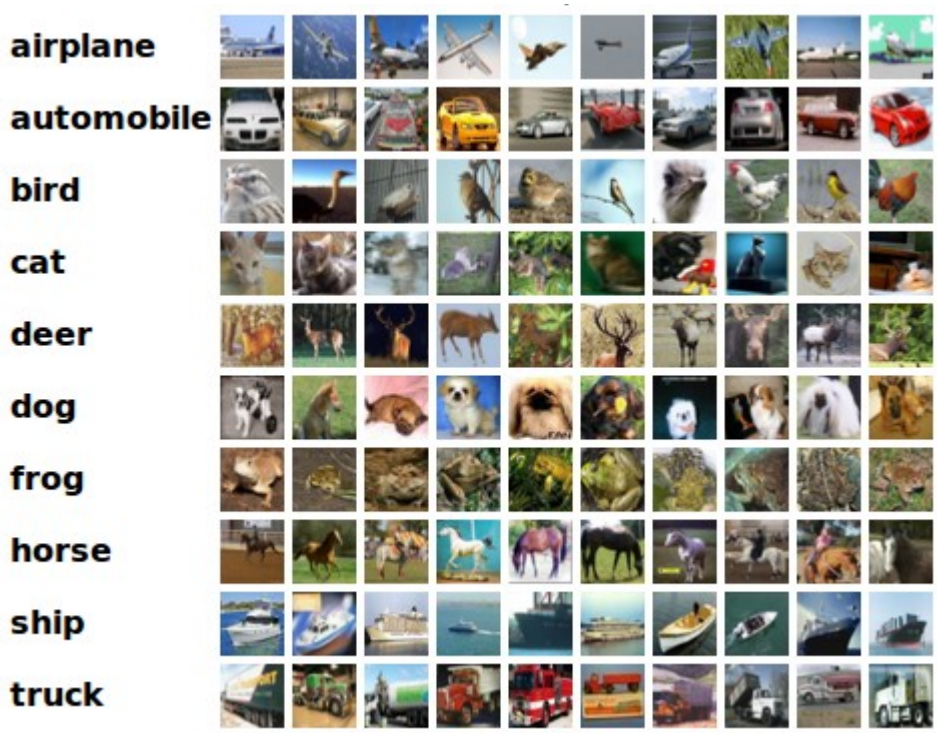
Department of Computer Science and Engineering

CSCE 4930 – Practical Machine Deep Learning

Dr. Mohamed Moustafa	Assignment 1 [10%]	Fall 2017

Part 1

Implement a k nearest neighbor (k-NN) classifier that can ***best*** recognize the 10 different classes in the CIFAR-10 dataset.



Details

a) Download the dataset from http://www.cs.toronto.edu/~kriz/cifar.html The CIFAR-10 dataset consists of 60000 32x32 colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images.

b) Build your k-NN classifier using 10000 images selected from the 50000 training set. Make sure that all classes are equally represented.

c) Using 3-crossfold validation technique, find the best “k” for this dataset.

Deliverable

1. Source code of your implementation. You can use any available language/tools/libraries for data/matrix manipulation and plotting. However the K-NN code has to be your own. **[2%]**
2. Plot of cross-validation accuracy (Y axis) versus number of neighborhood 'k' (X axis) showing the best chosen 'k'. **[2%]**
3. Using the 10000 testing set, report your Correct Classification Rate of each of the 10 classes separately. CCRn is the ratio of the number of correctly classified images in class n divided by the total number of test images in class n (which is 1,000). **[0.5%]**
4. Using the 10000 testing set, report your Average Correct Classification Rate (ACCR). ACCR is the number of correctly classified images divided by the total number of testing images (which is 10,000). **[0.5%]**

Part 2

Implement a Least Square (LS) regressor that can ***best*** predict the net hourly electrical energy output (P) of the power plant in the CCPP dataset. You can find the dataset at
http://archive.ics.uci.edu/ml/datasets/Combined+Cycle+Power+Plant

Deliverable

1. Source code of your implementation. You can use any available language/tools/libraries for data/matrix manipulation and plotting. However the LS regression code has to be your own. **[2%]**
2. Using 5-crossfold validation, find the best order for your LS regression model (linear, quadratic, etc...) for only the Exhaust Vacuum (V) as the input feature. Plot both training and validation E_{RMS} (Y axis) versus order of the regression polynomial model (X axis), showing the best chosen polynomial order. **[1.5%]**
3. Using 5-crossfold validation, find the best subset of the available 4 features: Temperature (AT), Pressure (AP), Relative Humidity (RH) and Exhaust Vacuum (V) for only the Linear LS model. Plot both training and validation E_{RMS} (Y axis) versus different combinations of the 4 features, showing the best combination. **[1.5%]**
4. Repeat 2 above for all features combined as input rather than only the Exhaust Vacum (V) **[Bonus 1%]**