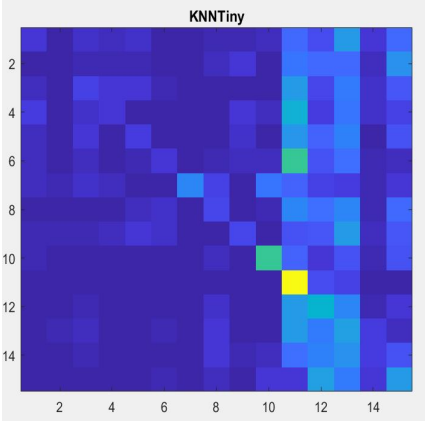
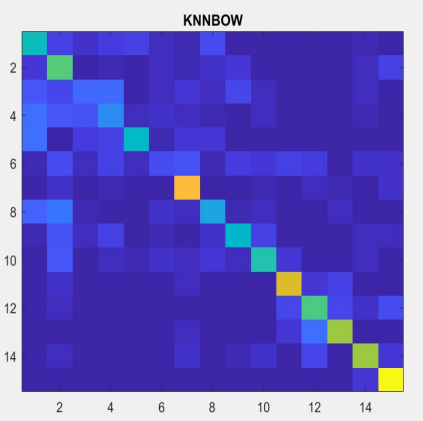
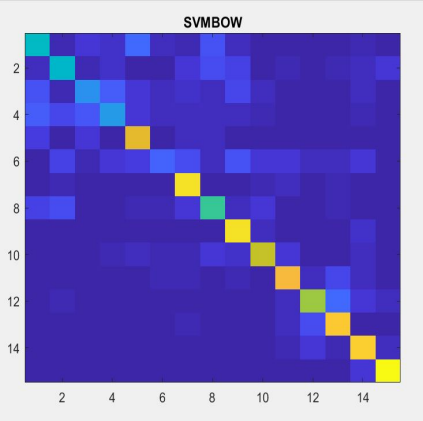


(a) Confusion Matrix for Tiny with KNN



(b) Confusion Matrix for BoW with KNN



(c) Confusion Matrix for BoW with SVM

Summary-Homework:3

The problem is to build a set of visual recognition systems that classify the scene categories. The dataset comprises of 15 categories of scenes. The system computes image representations and classifies the scene image using classifiers.Two kinds image representations have been done. One being Tiny Image and the other using Bag of Words Model. The classifiers developed are, namely, K-Nearest Neighbour(KNN) and Support Vector Machines(SVM).

To this end, I wrote the following functions in MATLAB, GetTinyImage.m, PredictKNN.m, ClassifyKNNTiny.m, BuildVisualDictionary.m,ComputeBoW.m,ClassifyKNNBoW.m,PredictSVM.m, ClassifySVMBoW.m.

The first image representation-classifier pair are Tiny Image combined with a KNN classifier. Three functions have been used to develop this module. GetTinyImage.m takes in an input image and outputs a tiny image representation of the input, i.e it resizes the image to a smaller dimension((16X16) has been used here) and then vectorizes the pixel intensity in a column major order. Post vectorization, the image vector is normalized by having zero mean and unit length. The Classifier used is KNN with the value of K set as 5. K is the number of neighbours used for label prediction. The accuracy obtained is 19.4667% and the confusion matrix for the same has been shown in fig(a). The low accuracy obtained in this case is justified because of the poor representation of the image features.

The second image representation-classifier pair are Bag of Words Model combined with a KNN classifier. Three functions have been used to develop this module. BuildVisualDictionary takes as input all the training images from 15 different categories and the dictionary size(set as 50 here) and builds a list of quantized visual words. Each visual word is described using dense SIFT features of dimension 128. This has been done using vldsift function from VLFeat library. The next step is to compute the Bag of Word feature for each image, given the vocab computed in the previous step and the dense SIFT feture for every image.The function outputs a histogram which is constructed by counting the SIFT features that fall into each cluster of the vocabulary.Nearest Neighbour Search has been used to identify the closest cluster centre. The histogram is normalized to have unit length. The classifier used is KNN with K set as 7. The accuracy obtained is 51.4% and the confusion matrix for the same has been shown in fig(b). The accuracy in this case increases because of the better representation of the image features using Bag of Words model.

The last image representation-classifier pair are Bag of Words Model combined with a SVM classifier. Three functions have been used to develop this module. BuildVisualDictionary takes as input all the training images from 15 different categories and the dictionary size(set as 50 here) and builds a list of quantized visual words. Each visual word is described using dense SIFT features of dimension 128. This has been done using vldsift function from VLFeat library. The next step is to compute the Bag of Word feature for each image, given the vocab computed in the previous step and the dense SIFT feture for every image.The function outputs a histogram which is constructed by counting the SIFT features that fall into each cluster of the vocabulary.Nearest Neighbour Search has been used to identify the closest cluster centre. The histogram is normalized to have unit length. The classifier used is SVM. In this case, 15 binary classifiers have been

trained by marking a particular label as 1 and all the other labels as -1. All the classifiers are evluated on each test sample and the classifier which is most confidently positive is selected as the winner.The accuracy obtained is 62% by using MATLAB's built-in function fitcsvm and predict and the confusion matrix for the same has been shown in fig(c). The accuracy in this case increases because of the better representation of both the image features using Bag of Words model and a better classifier than KNN. Note that vl_svmtrain produces unstable accuracies and hence has not been used for prediction.