REPORT

ON

"CIFAR 10 - Machine Learning"

Submitted By

Pyramid Group

Under the Guidance of

Pabitra Mitra

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITIUE OF TECHNOLOGY KHARAGPUR

KHARAGPUR - 721 302, WEST BENGAL

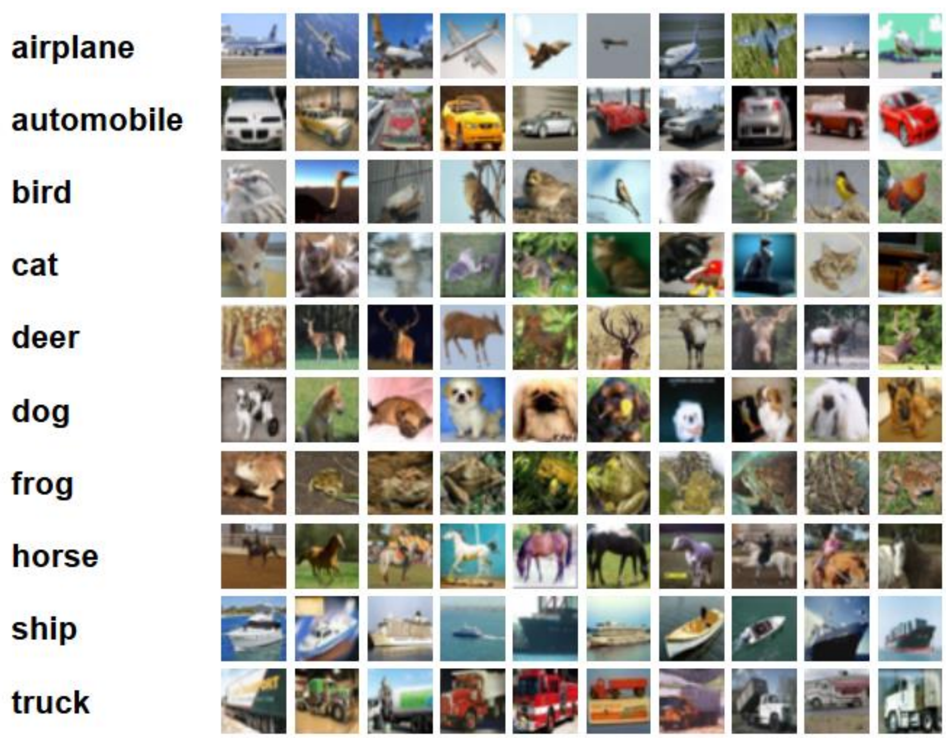
ACADEMIC YEAR

2016-2017

1. INTRODUCTION

This report will basically explore various machine learning techniques and will apply them on the open source dataset called CIFAR-10. The CIFAR-10 dataset consists of 60000 32x32 color images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images.

The dataset is divided into five training batches and one test batch, each with 10000 images. The test batch contains exactly 1000 randomly-selected images from each class. The training batches contain the remaining images in random order, but some training batches may contain more images from one class than another. Between them, the training batches contain exactly 5000 images from each class.



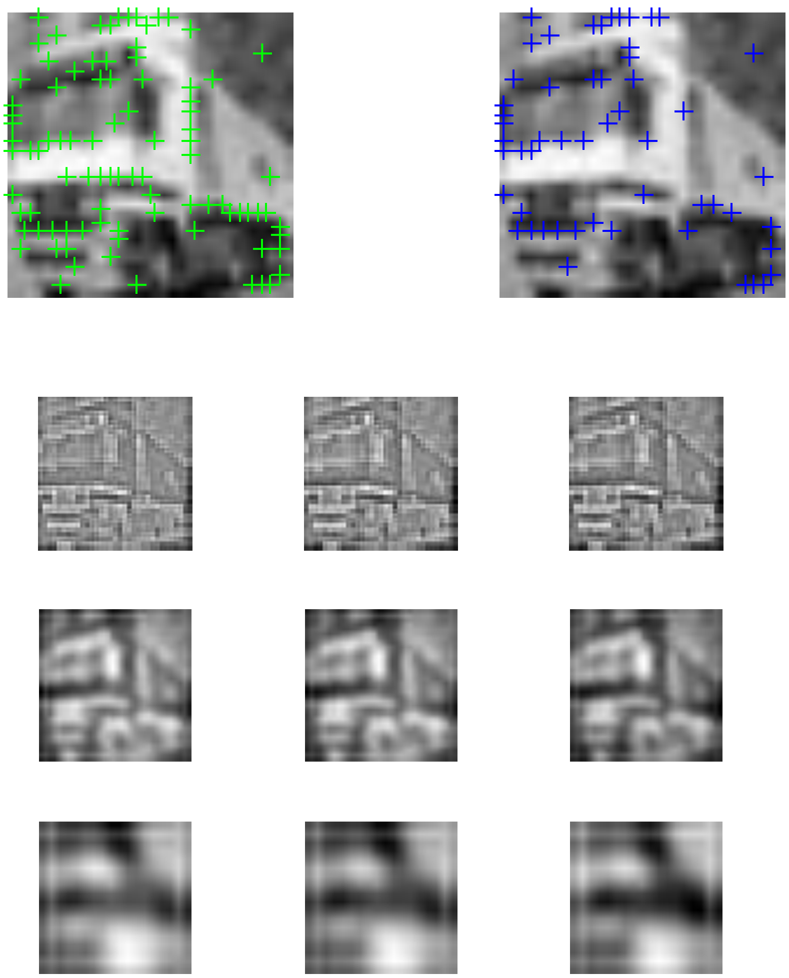
2. **TASK PERFORMED**

- Feature Extraction
    - SIFT
    - SURF
- Classification Techniques
    - Logistic regression
    - K Nearest Neighbor
    - Support Vector Machine
    - 2-Layer Fully Connected Neural Network
    - Convolutional Neural Network
- Clustering
    - Gaussian Mixture Model (GMM)
    - K-Means
- Dimension Reduction using PCA

3. FEATURE EXTRACTION

3.1 SIFT (Scale Invariant Feature Transform)

Scale-invariant feature transform (or SIFT) is an algorithm in computer vision to detect and describe local features in images. The algorithm was published by David Lowe in 1999. Applications include object recognition, robotic mapping and navigation, image stitching, 3D modeling, gesture recognition, video tracking, individual identification of wildlife and match moving.

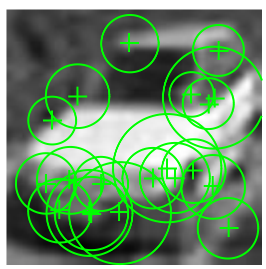


Sample run of the SIFT code on CIFAR-10

3.2 SURF (Speeded Up Robust Features)

In computer vision, Speeded Up Robust Features (SURF) is a local feature detector and descriptor. It can be used for tasks such as object recognition, image registration, classification or 3D reconstruction. It is partly inspired by the scale-invariant feature transform (SIFT) descriptor. The standard version of SURF is several times faster than SIFT and claimed by its authors to be more robust against different image transformations than SIFT.

To detect interest points, SURF uses an integer approximation of the determinant of Hessian blob detector, which can be computed with 3 integer operations using a precomputed integral image. Its feature descriptor is based on the sum of the Haar wavelet response around the point of interest. These can also be computed with the aid of the integral image.

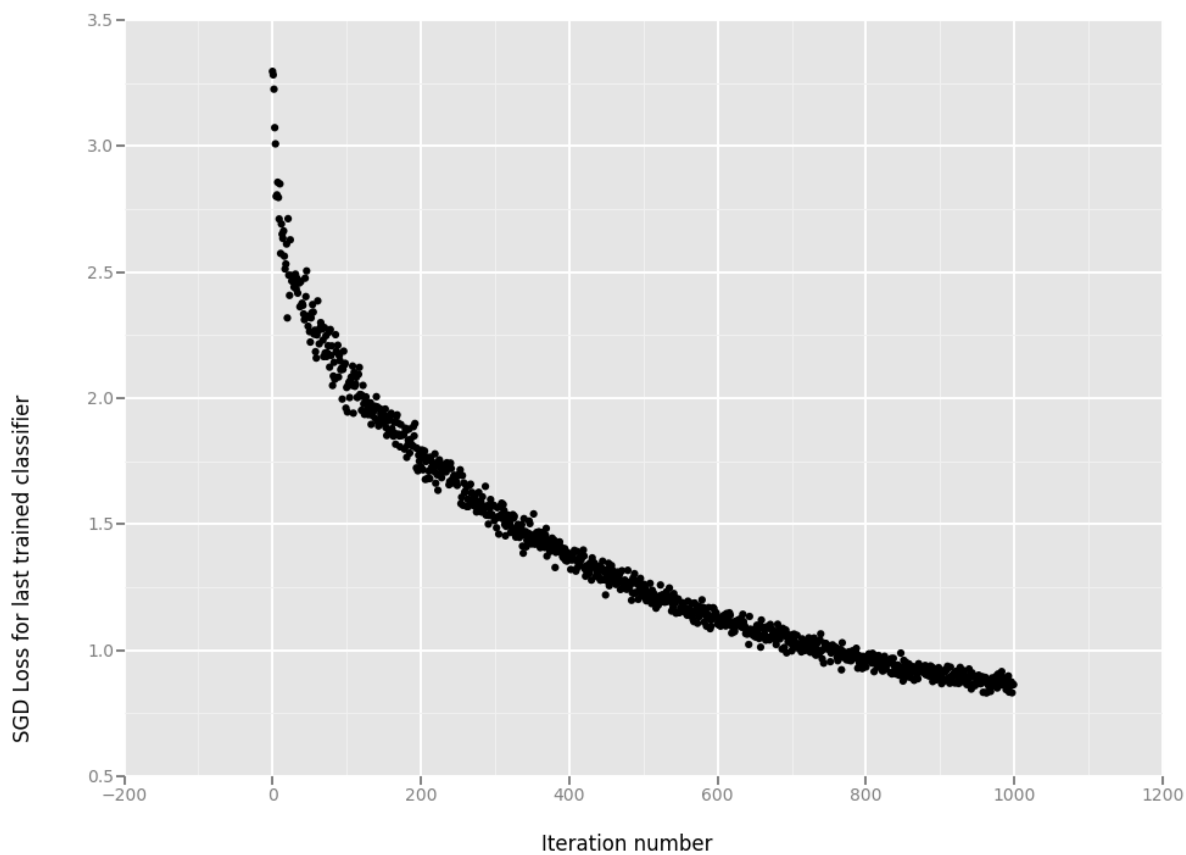


Sample run of SURF code on CIFAR-10

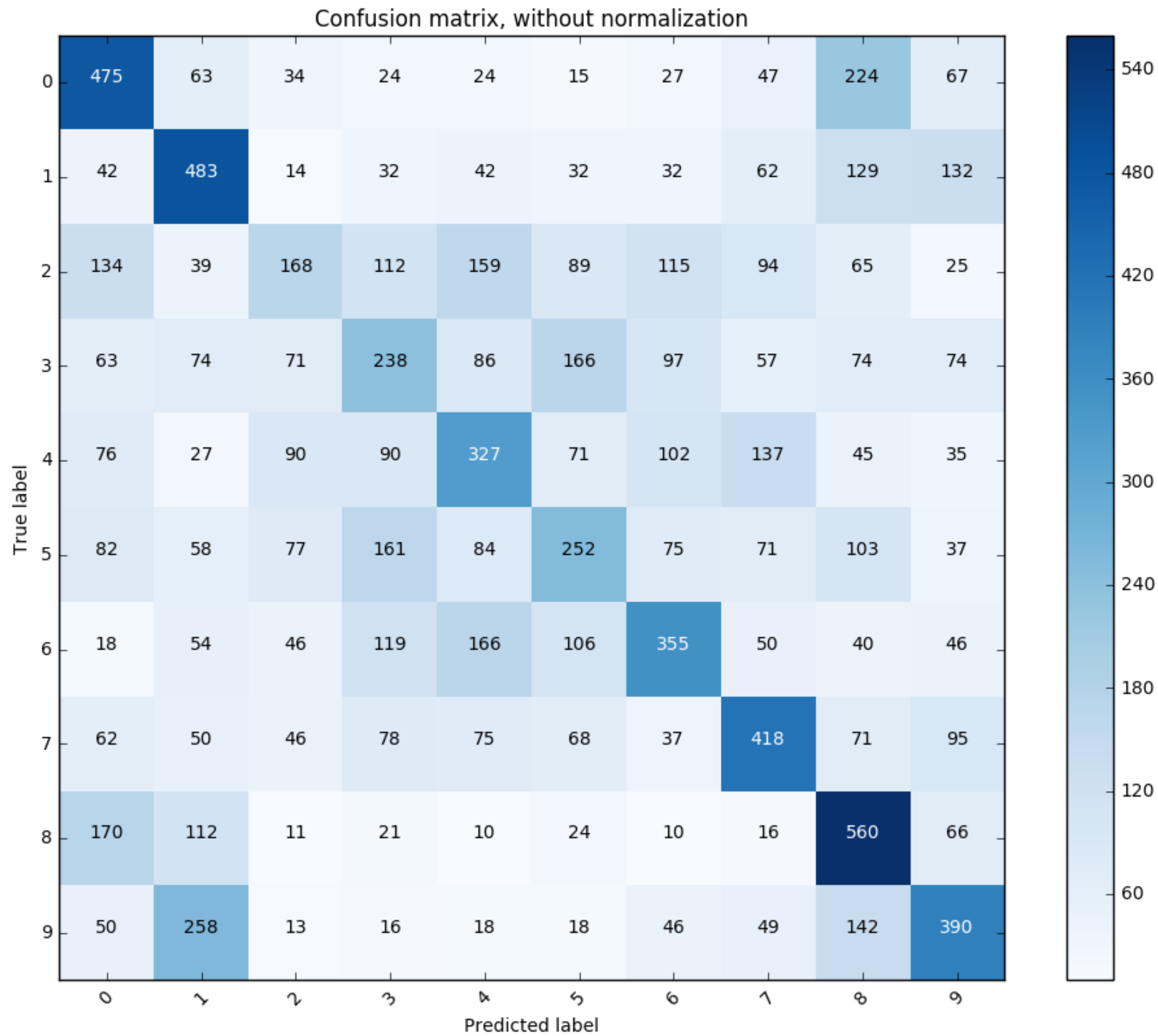
4. CLASSIFICATION TECHNIQUES

4.1 Logistic regression

In statistics, logistic regression, or logit regression, or logit model is a regression model where the dependent variable (DV) is categorical. This article covers the case of binary dependent variables—that is, where it can take only two values, such as pass/fail, win/lose, alive/dead or healthy/sick. Cases with more than two categories are referred to as multinomial logistic regression, or, if the multiple categories are ordered, as ordinal logistic regression.



Plot of the loss of the trained classifier



Confusion matrix for Logistic Regression on CIFAR - 10

4.1.1 Results

Training dataset accuracy: 0.384041

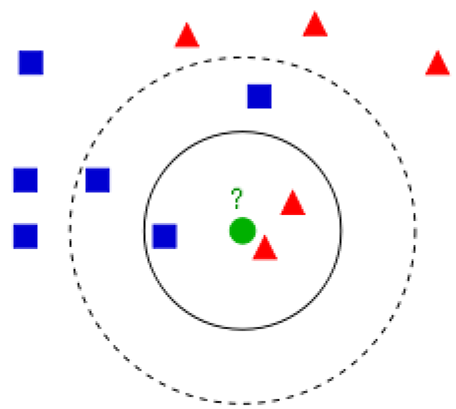
Validation dataset accuracy: 0.376000

Test dataset accuracy: 0.366600

4.2 K Nearest Neighbor

In pattern recognition, the k-Nearest Neighbors algorithm (or k-NN for short) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression:

- In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.
- In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbors.
- Example of k-NN classification. The test sample (green circle) should be classified either to the first class of blue squares or to the second class of red triangles. If k = 3 (solid line circle) it is assigned to the second class because there are 2 triangles and only 1 square inside the inner circle. If k = 5 (dashed line circle) it is assigned to the first class.

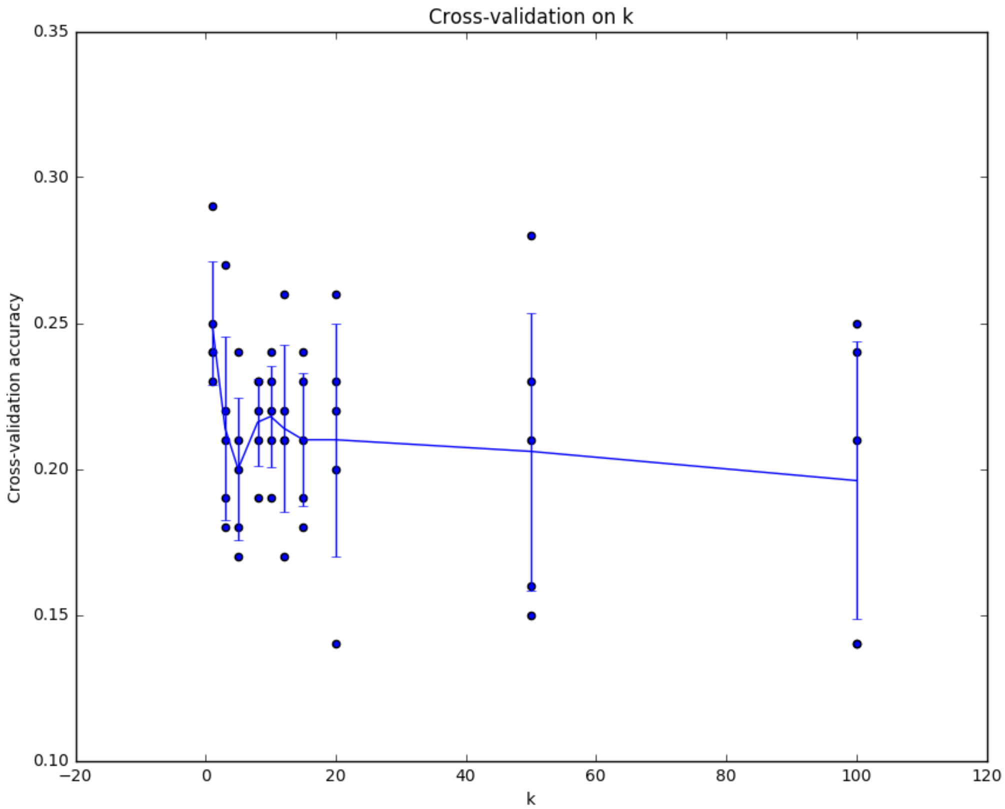


4.2.1 K Nearest Neighbor Accuracy on CIFAR-10

We got maximum accuracy for k value 50 –

k = 50, accuracy = 0.280000

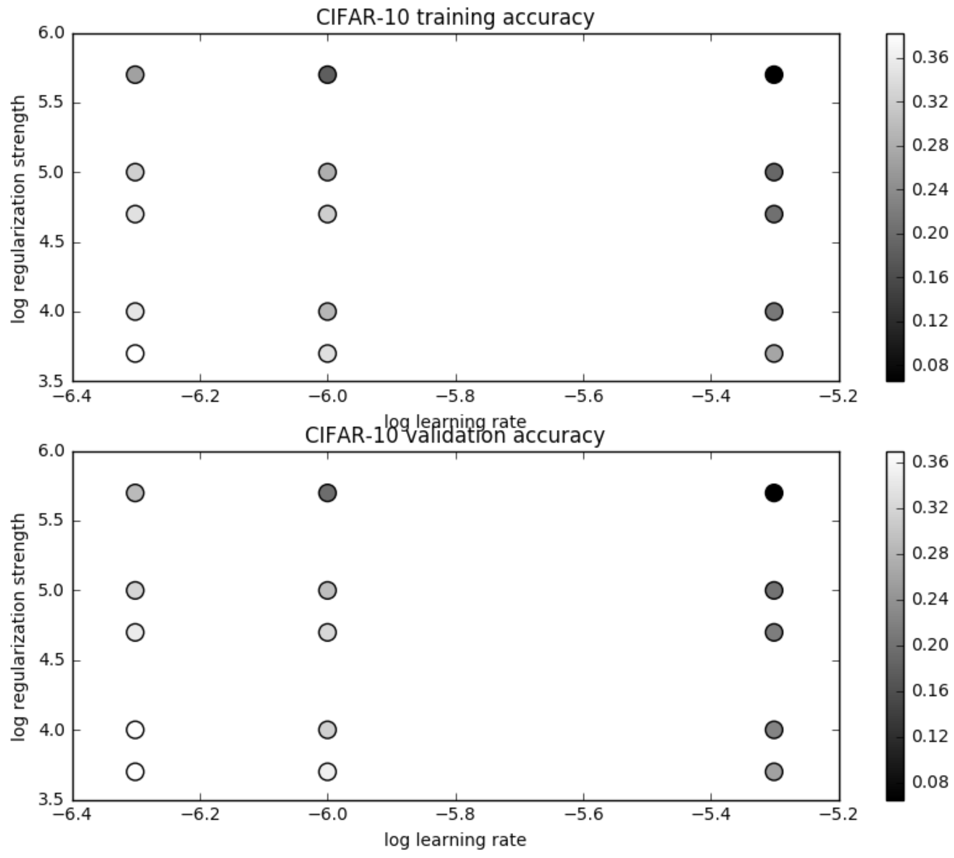
4.2.2 KNN Cross Validation Accuracy



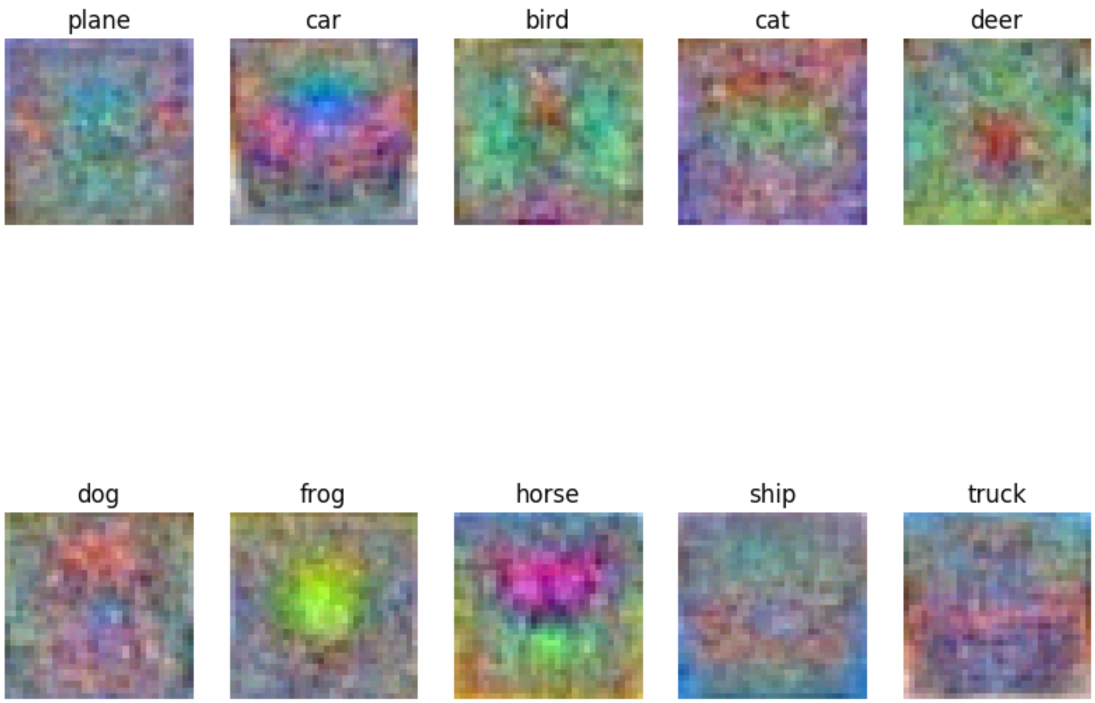
Cross validation accuracy

4.3 Support Vector Machine

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.



SVM cross Validation accuracy on CIFAR-10



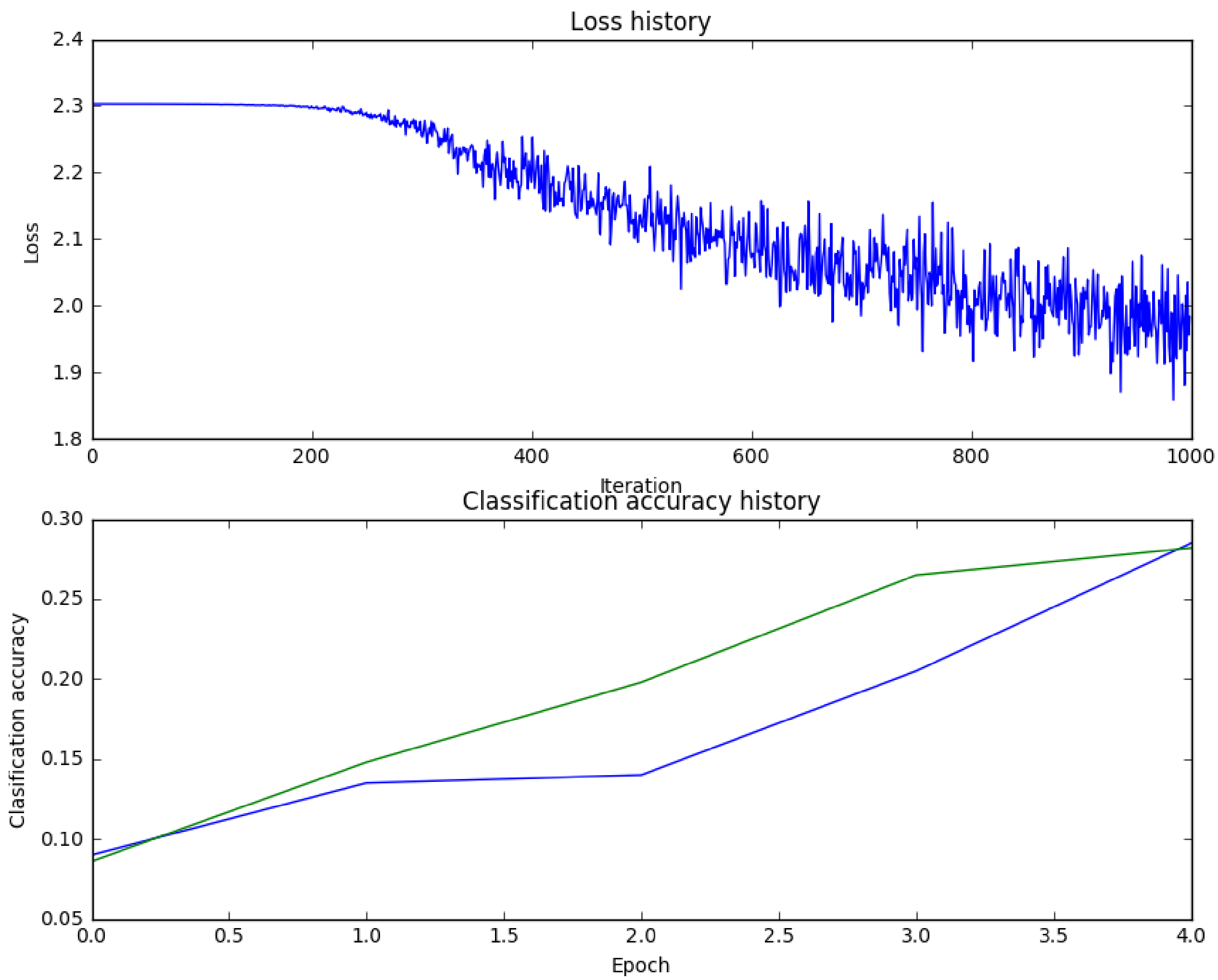
SVM trained weight vector

4.3.1 SVM Accuracy

36.9 % accuracy was observed on SVM

4.4 Two Layer Fully Connected Neural Network

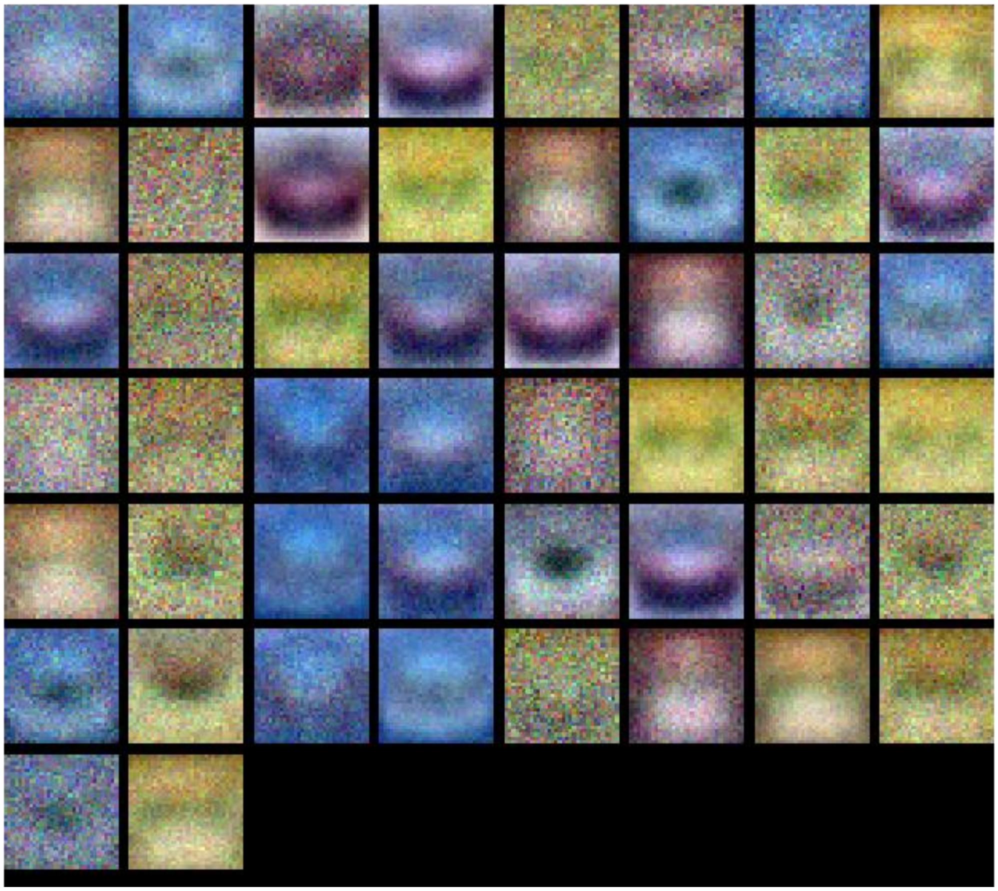
Here we built 2 layer fully connected neural network and performed classification on CIFAR-10 dataset. Neural Networks (also referred to as connectionist systems) are a computational approach which is based on a large collection of neural units loosely modeling the way the brain solves problems with large clusters of biological neurons connected by axons. Each neural unit is connected with many others, and links can be enforcing or inhibitory in their effect on the activation state of connected neural units. Each individual neural unit may have a summation function which combines the values of all its inputs together.



Cross validation accuracy on CIFAR-10

4.4.1 Neural Network Accuracy

43.0 % accuracy was observed on CIFAR-10



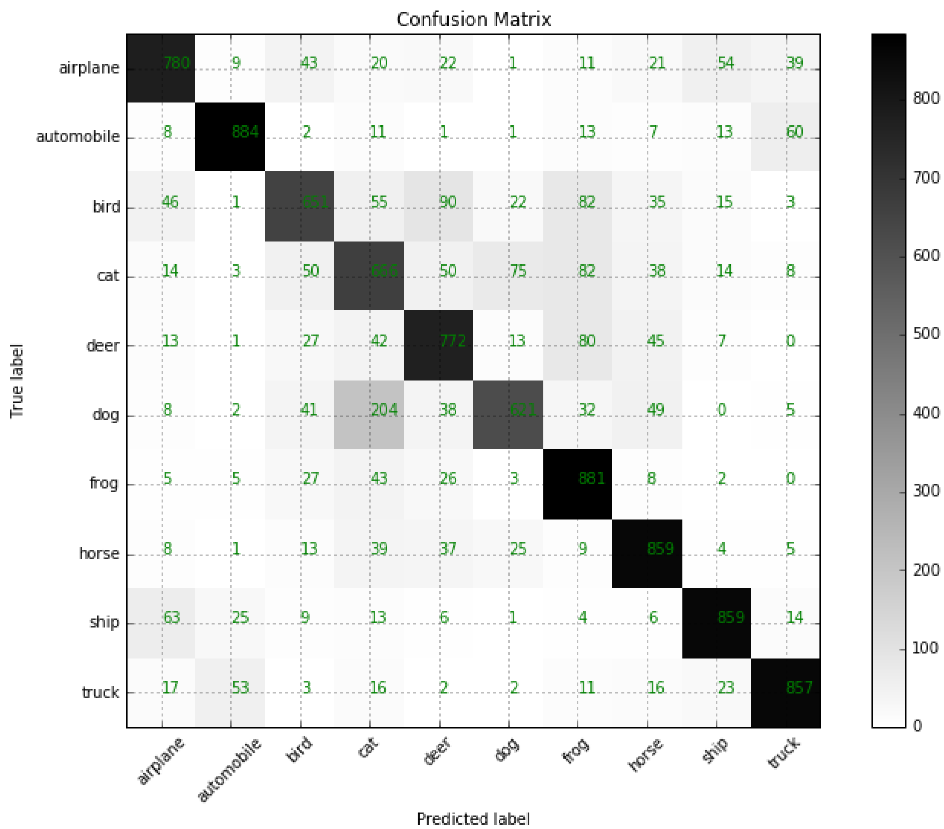
Learnt weights of 2 Layer NN for CIFAR-10

4.5 Convolutional Neural Network

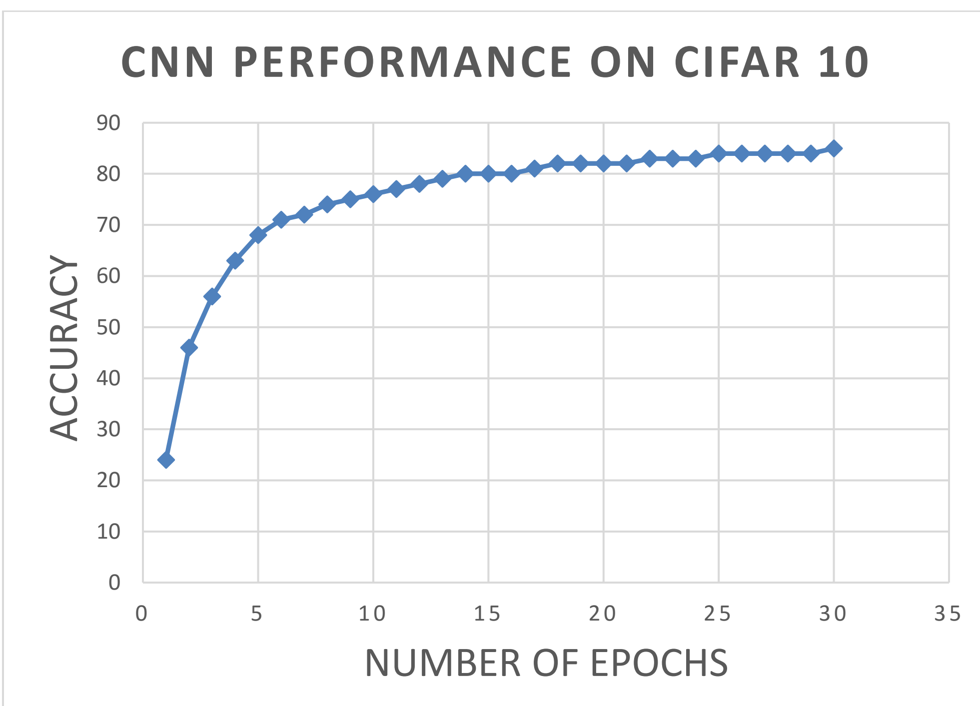
In machine learning, a convolutional neural network (CNN, or ConvNet) is a type of feed-forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex. Individual cortical neurons respond to stimuli in a restricted region of space known as the receptive field.

4.5.1 CNN Accuracy

85.0 % accuracy was observed on CIFAR-10



CNN Confusion Matrix on CIFAR-10



5. CLUSTERING TECHNIQUES

5.1 Gaussian Mixture Model

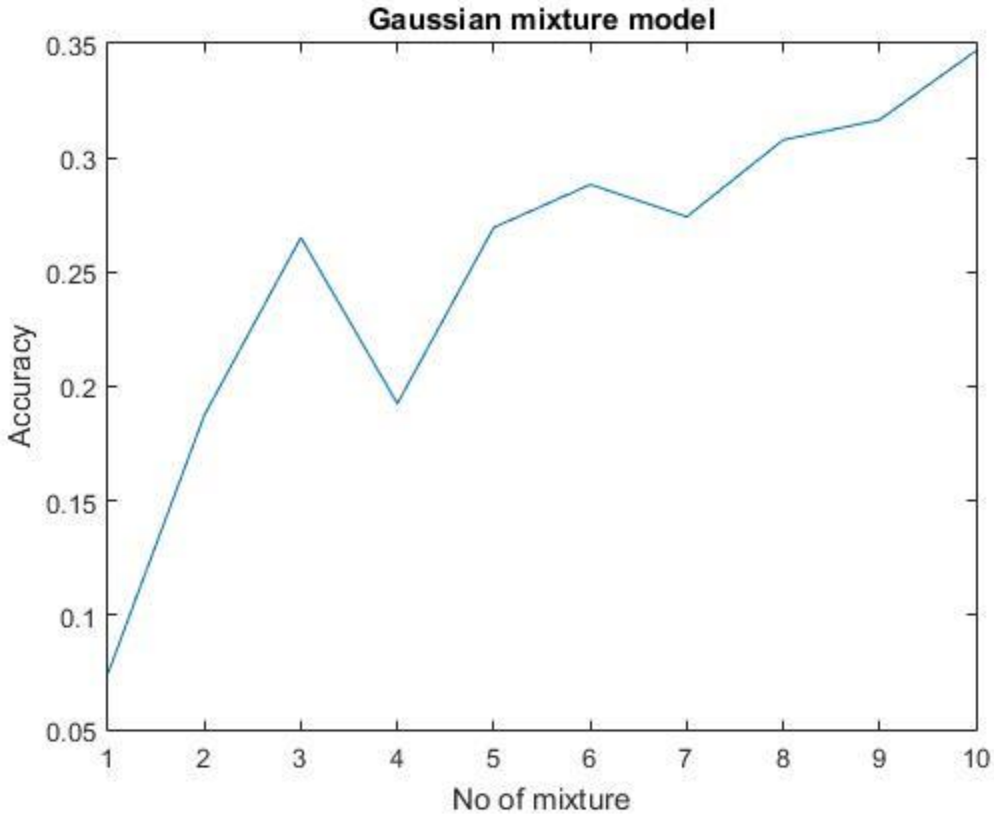
In statistics, a mixture model is a probabilistic model for representing the presence of subpopulations within an overall population, without requiring that an observed data set should identify the sub-population to which an individual observation belongs. Formally a mixture model corresponds to the mixture distribution that represents the probability distribution of observations in the overall population. However, while problems associated with "mixture distributions" relate to deriving the properties of the overall population from those of the sub-populations, "mixture models" are used to make statistical inferences about the properties of the sub-populations given only observations on the pooled population, without sub-population identity information.

We have also implemented ***Dimensionality Reduction using PCA*** in this GMM model.

A Gaussian mixture model is a weighted sum of M component Gaussian densities as given by the equation,

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^{M} w_i \; g(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i),$$

$$g(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)' \; \Sigma_i^{-1} \; (\mathbf{x}-\boldsymbol{\mu}_i)\right\},$$



GMM model accuracy for CIFAR-10

5.1.1 GMM Accuracy

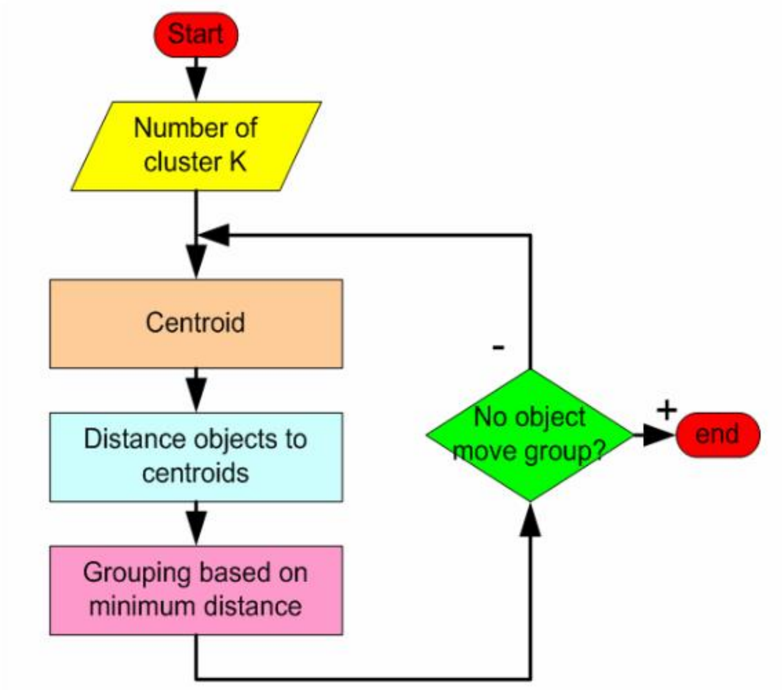
GMM has approximately 34.6% accuracy on CIFAR-10

5.2 K-Means

k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells.

5.3 K-Means Algorithm

1. For each data point, the closest cluster (in Euclidean distance) is identified;
2. Each cluster center is replaced by the coordinate-wise average of all data points that are closest to it.
3. Steps 1 and 2 are repeated until convergence. Algorithm converges to a local minimum of the within-cluster.



5.2.1 K-Means Accuracy

38.7 % accuracy was observed on CIFAR-10